Risk Adjustment And Medicare: Taking A Closer Look

The time for Medicare to experiment with new payment methods to managed care plans is now.

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

PROLOGUE: Almost from the first day that Medicare began contracting with health maintenance organizations (HMOs) for the purpose of enrolling its elderly beneficiaries, policy analysts have criticized the payment methodology used by the Health Care Financing Administration (HCFA), but no ready solutions to correct the problem were available. The problem is that, without adequate payment adjustments for the medical risk posed by individual beneficiaries, health plans have a financial incentive to enroll the healthy and spurn the sick. Congress, as a part of its recent efforts to address Medicare’s financial problems, did little to resolve this issue, but it did instruct HCFA to develop recommendations on how to reform the payment methodology. In this paper, Joe Newhouse and two of his colleagues propose that Medicare experiment with paying HMOs different amounts based on a patient’s diagnosis. Newhouse is the John D. MacArthur Professor of Health Policy and Management at Harvard University and one of the nation’s most respected health economists. He was the founding editor and continues to edit the Journal of Health Economics. He also currently chairs the Prospective Payment Assessment Commission. Newhouse’s impeccable reputation as a health economist evolved from his early work as the principal investigator for the RAND Health Insurance Experiment, the most far-reaching health services research project ever conducted in the United States. Melinda Beeuwkes Buntin, a doctoral student in the Harvard health policy program, holds a bachelor’s degree from Princeton University, where she graduated magna cum laude in 1993. John Chapman, who earned a doctorate in public policy from the Kennedy School of Government at Harvard last June, works as a private health care consultant.
ABSTRACT: Medicare’s method for reimbursing at-risk managed care plans causes potential problems with selection (when beneficiaries with higher-than-expected costs stay in traditional plans) and stinting (the tendency to underprovide health services). Adjusting payment by diagnosis offers substantial improvement. We favor large-scale demonstrations of diagnosis-based reimbursement. Reducing payment, a Clinton administration proposal, would recoup excess payments in the short run but not address the selection problem, which could reemerge. Selection makes current payments vulnerable to upward spirals. We propose not using traditional Medicare to update reimbursement. Basing some payment on enrollees’ actual use addresses selection and stinting. Rather than reinsurance, we propose blending traditional Medicare and risk-adjusted capitation. Ceding some cases to traditional Medicare in advance appears to be useful for terminally ill patients.

Many observers now agree that the Medicare managed care (at-risk) program needs better risk adjusters than the current adjusted average per capita cost (AAPCC) payment methodology.\(^1\) The agreement stems in part from concerns about selection—that those with higher expected costs have chosen to remain in traditional Medicare. Others have worried about the possibility of underprovision of services, what we term stinting.

In this paper we review the incentives for selection and stinting that the current managed care payment methodology creates. The selection incentives can be ameliorated, although probably not eliminated, by taking account of diagnosis in payment. Methods to do so now exist, and we propose that they be introduced into the program on a large scale. To reduce selection to minimal levels, it may prove necessary to base payment partially on an adjusted capitation and partially on actual use. Such a blend also would address stinting. Selection also artificially increases Medicare costs because reimbursement for at-risk health plans is now tied to the average fee-for-service payment, and selection increases that payment. We therefore propose that the mechanical link between payment in the traditional program and payment in the at-risk program be broken and that a new update methodology for the at-risk program be introduced. Finally, we propose that plans be asked to name some modest portion of enrollees in advance (enrollees with predictably high costs) for whom plans would be reimbursed partially or wholly on the basis of actual use; this should further mitigate selection problems.

Lack Of Adequate Risk Adjustment And Selection

The lack of adequate risk adjustment causes three problems related to selection.
“The lack of adequate risk adjustment means that the government overpays Medicare managed care plans.”

- **Access problems.** Bad risks, especially the chronically ill, face a potential access problem. The problem arises because health plans are reimbursed the same amount for persons with very different expected spending. Without adequate risk adjustment, all health plans have an incentive to induce those with high expected spending to enroll in other plans, which creates a possible access problem for such persons. In particular, plans have incentives to configure themselves such that they will not appeal to bad risks; under current reimbursement arrangements it is not healthy for a plan’s bottom line to have the reputation of having the best cancer service in town. Further, even if no plan in a local market is now discriminating against bad risks, a new entrant may attempt to do so. If many good risks were to switch to the new entrant, existing plans would either have to discriminate against bad risks as well, or be left losing money with a disproportionate number of bad risks as enrollees.

- **Efforts to attract good risks.** Plans have an incentive to spend resources to attract good risks. This is the flip side of the potential access problem for the chronically ill. Because the AAPCC makes enrolling good risks profitable, it is worth spending money to attract them. In particular, plans have an incentive to spend up to the amount of profit on good risks to attract them, but such marketing expense may be an inefficient use of funds. For example, plans in high-AAPCC areas now pass some of their profits back to enrollees in the form of additional benefits. Offering enrollees a subsidized health club membership, for example, is not only an added benefit but one that disproportionately attracts good risks. Even the good risks, however, may not value the benefit at what it costs the health plan to provide it. (Similarly, it is worth plans' spending money to avoid bad risks, and those monies are also not likely to be socially productive.) For the same reason that plans may not want a reputation for having the best cancer service in town, they may want a reputation for having the best sports medicine service in town.

- **Risk-adjustment deficits.** The lack of adequate risk adjustment means that the government overpays Medicare managed care plans. The AAPCC formula pays plans 95 percent of the average spending in traditional Medicare within age, sex, welfare, institutional, and employment classes by basis of eligibility: elderly, disabled, or those with end-stage renal disease. Thus, as just described, plans have an incentive to attract persons within each class whose
expected spending is less than the reimbursement and to dump others. An incentive, of course, may be acted upon to a greater or lesser degree. Although the claim is controversial, our reading of the evidence is that plans as a group, not necessarily every plan, have in fact disproportionately attracted beneficiaries whose expected spending is less than 95 percent of AAPCC; thus, there has been favorable selection into plans.

Strong evidence of favorable selection comes from analyses by the Physician Payment Review Commission (PPRC) and others. The large spending and mortality differences found by the PPRC for at-risk versus traditional Medicare plan enrollees do tend to shrink over time. Although the evidence is somewhat conflicting on whether the differences are ever eliminated, our reading is that they are not.

Because of favorable selection, the Congressional Budget Office (CBO) has in the past scored increased enrollment in at-risk plans as costing the Medicare program money. To address the overpayment, the Clinton administration proposes to reduce reimbursement to at-risk plans from 95 percent of AAPCC to 90 percent. A better approach would be to improve risk adjustment. As we describe below, better risk adjusters are now available, but to judge them and other proposed adjusters, we first take up, in as nontechnical a way as possible, the issue of how one assesses a risk adjuster.

**How Would We Know An Adequate Risk Adjuster If We Saw One?**

An ideal risk adjuster should make the payment to a plan match an enrollee’s expected spending, that is, what a given person who has average “luck” would spend while enrolled. Expected rather than actual spending is relevant for selection because it is in principle what a plan could predict and thereby determine in advance who might and might not be financially profitable. If a set of risk adjusters caused payment to match expected individual spending exactly, they would explain all of the variance in expected spending across individuals; hence, the proportion of variance explained by a set of risk adjusters has become a principal measure of the adequacy of risk adjustment.

For reasons that will become clear momentarily, the literature calculates the proportion of variance in actual rather than in expected annual spending that different risk adjusters explain. The variance in expected spending is less than the variance in actual spending by the amount of luck. A person with obstructed coronary arteries is at higher risk of a heart attack than is a similar person with no obstructions and so has higher expected spending. The
person’s actual spending, however, will be importantly influenced by whether a heart attack actually occurs (luck). To avoid selection, a risk adjuster does not need to explain the variance attributable to luck; purely random effects on spending that could happen to anyone—for example, being hit by a drunk driver—should not cause plans to either want or shun any particular enrollee.5

The adjusters in the current AAPCC formula explain about 1 percent of the variance in actual spending. On the face of it, 1 percent does not seem like much of an improvement over no risk adjustment at all, and, although it is better than nothing, in fact it is not much better. A seventy-year-old noninstitutionalized woman with breast cancer has very different expected spending from a seventy-year-old noninstitutionalized woman with no chronic illness, yet the present formula pays a plan the same amount for each woman. But a 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. First, the variance in expected spending cannot be precisely estimated; in fact, one can at best give a lower bound on the relevant figure. As a result, what could be ideally estimated is not precisely known. Second, it is unclear how close to the ideal an adequate adjuster must come. The difference in variance that an ideal and actual adjusters explain only indicates the magnitude of the financial incentives plans face for enrolling good risks. It is unclear to what degree plans can or will act on changes in those incentives. Third, the adequacy of any adjuster presupposes agreement on the appropriate treatment and relative expense of various diseases. In many cases, agreement is lacking.

A lower bound. The first problem is purely statistical. What we wish to know—the variance in expected spending—we do not directly observe, which is why the literature analyzes actual spending. One can, however, try to estimate the variance in expected spending from observations on actual spending. Usually the literature has done this by dividing the variance in actual spending into three components, the first two of which make up 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 have 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
“Regardless of how good risk adjusters become, we will never have a perfect formula, only better approximations to it.”

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 calculating the average spending of a person over a period of years and using that as an estimate of the person’s expected spending for the following year. There is a reasonable consensus in the literature that the proportion of variance in actual annual spending that could be attributed to a fixed effect is on the order of 15–20 percent.

In addition, if one knows that 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 also will be above the person’s average (and conversely). This is part of the second component. Again, there is something of a consensus in the literature that if one knows last year’s spending, and to a lesser degree spending in earlier years, one can explain another three to five percentage points of variance above the 15–20 percent that the fixed effect could predict. Thus, it appears that anyone observing the past spending of a given person could explain around 20–25 percent of the variance in actual annual spending.

This value, however, is a lower bound on the ability to predict future spending because other factors may be observed that are not reflected in past spending but will predict future spending—that is, that are part of the second component. For example, a woman may be pregnant; or a person could have just been diagnosed with cancer. In other words, plans could potentially predict more than the 20–25 percent of the actual variance, but 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 around 1 percent leaves substantial incentives for cream skimming and dumping. As adjusters improve, however, the ambiguity begins to matter more.

• Plans’ incentives to act. Regardless of how good risk adjusters become, we will never have a perfect formula, only better approximations to it. Thus, the question arises of how close to perfect the formula must be to make plans’ incentive and ability to seek favorable risks a de minimus problem.

Actions to enroll better risks, such as subsidizing a health plan
membership, can be costly. As risk-adjustment formulas improve and the profitability of cream skimming and dumping correspondingly diminishes, plans may not find it worthwhile to spend resources to take such actions. Moreover, better risk-adjustment formulas also are likely to change the profile of “good” and “bad” risks. Good risks will not necessarily be healthier persons but rather healthier within each class in the formula. Changing the classes may materially alter the cost or difficulty of achieving a given level of favorable selection. In short, a workable adjuster need not achieve the ideal. But how far from perfection will be sufficient is unknown.

**Variation in treatment methods.** Typically, estimates of use at the individual level have not been available for enrollees in the Medicare at-risk program. As a result, most risk-adjustment methods have been calibrated using data from traditional Medicare or from nonelderly persons using fee-for-service providers. But weights based on indemnity plans’ patterns of treatment may not be 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 pose no problem, but such proportionate reductions are unlikely. Some conditions may be treated much less intensively (for example, perhaps little individual psychotherapy in at-risk plans), whereas other conditions may be treated just as they would have been with traditional indemnity insurance.

One can draw two conclusions from the disproportionate reductions in use across classes of persons. First, most existing risk adjusters, whether based on age, diagnosis, or health status, are distorted because they have been calibrated on fee-for-service patterns of treatment. That is, even if 95 percent of AAPCC were the right aggregate payment level, at-risk plans using their preferred modes of treatment would face profits and losses on particular subgroups with resulting incentives to favor or disfavor those groups. Second, and in some ways more importantly, risk adjustment sets payments for different classes of persons—for example, persons with and without certain diagnoses if adjusters based on diagnoses are used. As such, risk adjustment embodies judgments about how persons with various diagnoses should be treated. Most, if not all, 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 that is 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.
In sum, an ideal risk adjuster that does not use any information from the current year to set rates (that is, is prospective) should explain at least 20–25 percent of the variance in spending across persons. But a plan could observe something that would predict future spending but would not have affected past spending or affect it minimally; as a result, the actual proportion of variance that might be predictable by a plan could be higher than 20–25 percent—perhaps much higher—but we do not know how much higher. How much higher did not really matter for the purpose of documenting the inadequacies of the AAPCC, because it is plainly inadequate even with a 20 percent criterion. Neither did questions about the magnitude of plans’ incentives or ability to act or about appropriate treatment patterns matter much in judging the AAPCC. The newer risk adjusters, however, are much better than the AAPCC, so the lower-bound qualification, uncertainty about plans’ incentives and ability to act, and concern about what is appropriate treatment all become more important in judging what is good enough.

■ Explaining monthly spending: a technical caveat. Current Medicare regulations allow beneficiaries to enroll and disenroll in at-risk plans monthly. As a result, the criterion for the adequacy of any risk adjuster should be the ability to explain the variance in monthly rather than annual spending. We have discussed annual spending, however, because that is the time period the literature has used. In general, one’s ability to explain monthly spending should be less than one’s ability to explain annual spending because random or unpredictable events will play a proportionately larger role in a shorter period. In other words, if the analysis had been done on monthly data, the 20–25 percent lower-bound figure given above would be smaller. Although we believe that the policy conclusions we and others draw about various risk adjusters from analyses of annual data are likely to hold if monthly data had been used instead, demonstrating this empirically remains for future work.

The State Of The Art

Three years ago Joseph Newhouse assessed the state of the art in risk adjustment as primitive. Since that time the Health Care Financing Administration (HCFA) and others have continued to support the further development of risk-adjustment methods. In our view, the two leading methods at the moment are diagnostic cost groups (DCGs), which have evolved into hierarchical coexisting conditions (HCCs), and ambulatory diagnostic groups (ADGs). Both methods improve on the AAPCC by exploiting diagnostic information to predict expected spending. Thus, to continue the example used above, both DCGs and ADGs would predict that a
seventy-year-old, noninstitutionalized woman with breast cancer would spend more than an otherwise similar woman without breast cancer and would correspondingly pay more to a plan that enrolled such a woman.

Prospective versus retrospective adjusters. How much more variance one can explain using diagnostic information depends upon whether the methods are implemented prospectively or retrospectively. Prospectively means that only diagnostic information from a past period or periods, usually the previous year, is used; retrospectively means that diagnostic information from the current period is used. Thus, if a person is diagnosed with diabetes mellitus in 1997, a prospective adjuster would not use that information to set reimbursement in 1997, whereas a retrospective adjuster would. A prospective adjuster would, however, use the information for setting 1998 reimbursement.

A retrospective adjuster will explain more variance in actual spending than a prospective adjuster will because it explains some of the random variation or luck. Suppose, for example, that I have no known diseases but have a heart attack in the current year, something that will not be known in advance (that is, is random). The additional current-year spending associated with the heart attack will be partly explained by a retrospective adjuster; it will place me in a higher category that reflects the diagnosis of acute myocardial infarction. But, as pointed out earlier, explaining random variation is not important for purposes of mitigating selection behavior.\(^\text{17}\)

We focus on the variance that can be explained by prospective rather than retrospective adjusters for two reasons. More importantly, we do not have a standard for how much variance a good retrospective adjuster should explain. Second, as just described, explaining random variation is not important for 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. Randall Ellis and colleagues and Daniel Dunn and colleagues found 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.\(^\text{18}\) Thus, their results imply that much of the gain in explained variation achieved by retrospective adjusters comes from explaining random spending in the current year. By contrast, John Chapman in unpublished work and using data from the population under age sixty-five found that retrospective adjusters do pick up some spending that appears predictable, in which case the use of a retrospective adjuster could reduce the incentives to select.\(^\text{19}\) 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 persons with identified heart disease; with a retrospective adjuster the plan can recoup some of the expense of the heart attack if one occurs.

**Prospective adjusters in traditional Medicare.** In terms of ability to explain variance prospectively using fee-for-service Medicare data, the HCC model appears to do better, explaining about 9 percent of the variance, whereas ADGs explain about 6 percent.\(^\text{20}\) Although these figures are a fair ways from the 20–25 percent lower bound, they are clearly a substantial improvement on the AAPCC. In addition to explained variance, both Ellis and colleagues and Jonathan Weiner and colleagues present data on how these risk adjusters affect profitability, thereby giving one a feel for both the improvement that they can make and the magnitude of the remaining problem.\(^\text{21}\) For example, Ellis and colleagues calculated the average profit in 1992 from enrolling a beneficiary with traditional Medicare who was in the lowest quintile of 1991 expenditure, assuming that 1992 spending in the at-risk plan would be exactly what was spent in traditional Medicare. If the plan is paid the AAPCC, the average profit for enrolling this group of persons is $2,134; using HCCs to adjust payment, that figure drops substantially, to $424. At the other end of the spectrum, the average loss from enrolling someone in the highest quintile of 1991 expenditure is $4,425 under the AAPCC but only $1,311 under HCCs.\(^\text{22}\) In sum, both of these risk-adjustment methods yield a noteworthy improvement in incentives to cream and dump, but nontrivial profits and losses remain, especially in the case of losses.\(^\text{23}\) Whether the changes in profitability brought about by HCCs and ADGs would reduce selection behavior to negligible levels is an empirical question.

**Use of self-reported measures.** Another strand of risk-adjustment literature uses self-reported health 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 the diagnostic-based measures do. When used in conjunction with diagnostic measures, self-reported health status appears to add around one percentage point of explained variation.\(^\text{24}\) Both because of the modest increment in explained variance and because we have concerns that using self-reported health status measures for reimbursement is susceptible to gaming, we have not further explored the use of self-reported measures.

**Behavioral Response And Fighting The Last War:**

**Incentives To Stint**

The designers of HCCs have taken care to limit higher payments to
diagnoses for which treatment is relatively nondiscretionary. But more 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 the incentives for discretionary treatment at the possible expense of increasing incentives to select—the choice the creators of 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 makes the implicit assumption that if there is a choice of treatment, the lower-cost method should be preferred. But there are important objections to this view.

First, and most importantly, reimbursement that is independent of use means by definition that the additional revenue from doing something more is zero. But if excess services are observed with overly generous fees, one may observe underservice with no additional payment. In short, even apart from issues of selection, no payment for additional services raises the issue of possible stinting. Although stinting has not been thought to be a major problem in the under-age-sixty-five market, methods for detecting underuse are at a less advanced stage than is true for detecting overuse. For example, quality-assurance mechanisms in capitated physician groups seem directed more toward overuse than underuse.

Unlike pure capitation, Medicare’s prospective payment system (PPS) does in many cases pay more if discretionary services are rendered. For example, it pays a hospital more for patients with coronary artery disease who have a bypass operation than for similar patients who do not. The designers of PPS presumably deemed paying nothing more for bypass operations as likely to lead to too few operations. Although excessive fees can induce unwanted or unnecessary services, to conclude that marginal revenue should always be zero seems to be an overreaction to the excesses of reimbursement in the indemnity insurance system—fighting the last war. Second, a substantial amount of medicine involves learning by doing. When open-heart surgery was first introduced, it took much more time and yielded much poorer results than it does today, because cardiac surgeons improved with practice. If today’s lowest-cost option is always chosen, the possibility that a more expensive option today will become cheaper tomorrow is lost. Third, even if there is no...
learning by doing, there may be great uncertainty about which course of treatment is better; without some variation in treatment, we will not learn.\textsuperscript{28}

Thus, as described in the next section, we believe that a portion of reimbursement to a health plan should be based on actual use. This could reduce the likelihood of selection and also could reduce stinting, although if the payments become too high, the issue of overservice emerges.

A different proposal is to use a much finer classification of diagnoses to reimburse plans than HCCs now use, especially for high-cost diagnoses. Doing so addresses the issue of selection in the context of diagnoses that have a higher mean cost than their present HCC category, but insofar as the payment is still fully prospective it does not address the issue of possible stinting. Moreover, obtaining stable, much less appropriate, relative weights for more numerous categories requires massive databases.

Where Does Policy Go From Here?

Based on the foregoing analysis, we have one negative and four positive suggestions.

- **Proposed reduction in payment to 90 percent of AAPCC.** Although cutting payments to at-risk plans will recoup some monies for the government in the short run, it is not likely to end overpayment. In effect, it moves the line with respect to who is and is not a profitable enrollee; persons who at 95 percent were not profitable become even less profitable, and a few enrollees who before were marginally profitable now become marginally unprofitable. Thus, plans still have an incentive to cream and dump. Furthermore, under this proposal payments to plans would fall, and one thus would expect plans to take back some of the additional benefits they now offer enrollees. These could well be those benefits that disproportionately attract bad risks, such as a drug benefit. In sum, for selection to diminish as an issue, it is not sufficient to reduce plan payments; better risk adjusters are needed.

- **A demonstration of better risk adjusters.** We believe that a demonstration of risk adjusters that uses diagnostic information is in order. Indeed, in our view, the evidence is sufficiently strong that these methods greatly improve upon the AAPCC that we would make the demonstration a large-scale, multistate one rather than the small-scale ones that seem to be going forward.\textsuperscript{29}

- **Breaking the link between payments in the traditional program and the AAPCC.** Any favorable selection into the at-risk program by definition raises the average payment in the traditional program, which further increases the AAPCC and Medicare spend-
ing. This potentially vicious circle can be eliminated by specifying an update factor for the AAPCC that is not mechanically linked to the traditional program. Of course, the rate of increase in the traditional program could be taken into account in setting an update factor; for example, one could specify that as an upper bound. Nevertheless, judgment would be used about the degree to which increases in the AAPCC should track increases in the traditional program. This recommendation also applies if diagnostic-based risk-adjustment schemes are implemented or if partial capitation is implemented, as described next.

**What to do in the absence of adequate risk adjustment?**

Suppose a demonstration shows that diagnostic-based risk adjustment does not reduce selection to negligible levels, an outcome we judge as reasonably likely. Then both for that reason and to address possible stinting, one of us has elsewhere proposed moving to partial capitation, or a blend of adjusted capitated payment and payment under the traditional program. Plans thus would be paid more for enrolling those who used more services and conversely.

Paying plans more for additional services could be done in several ways. The most straightforward conceptually is to use a literal blend for each person, such as half the payment being (risk-adjusted) capitation and half being based on what the traditional program would have paid (for example, the PPS payment for hospitalized patients, the Medicare fee schedule for physician services). What the right blend is (50/50 or 75/25 or something else) would have to be determined by experience. This is an area in which a formal experiment could be useful.

Moving to a blended payment would require collecting and coding encounter data for each person. A seemingly less burdensome scheme is an outlier or reinsurance scheme; payments for additional use would be made only after some threshold level had been exceeded, analogous to PPS outlier payments. Moreover, such a system appears to target payment toward persons where the selection incentives are the worst (the high spenders).

Relative to a blend, however, a reinsurance or outlier proposal has three important drawbacks. First, reinsurance for large spenders does little 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 nonoutlier cases, there is an incentive to stint.

Moreover, the apparent advantages of an outlier scheme may not be real. The additional payments do not necessarily go to the cases
with the largest predictable losses; to the degree that the outlier spending arises from random events, there is not a selection problem. As for administrative burden, even with an outlier scheme plans presumably would have to keep track of all spending because they would need to establish that a person went over a threshold and will want to collect all monies owing for such cases. Thus, whether additional payment should be limited only to high-cost cases or should apply to all cases is an empirical question of how serious selection is under each method and how much an actual as opposed to an ideal fee-for-service payment might degrade efficient production. We suspect some support for an outlier arrangement rather than a blend arises from trying to minimize additional payments for additional services, which, as explained above, is too facile.

A different reinsurance type method, proposed by Erik van Barneveld and colleagues, would improve targeting by giving additional payment to those cases that the plan predicted in advance would be high cost.\(^3\) Their proposal, made in the Dutch context, would have plans cede a certain proportion of their cases—specifically, those they predict will be the highest cost—to a pool. Although the percentage to be ceded is a policy choice, they illustrate using 2 percent and 4 percent of the cases in each plan.

In the Medicare context, there is, of course, no common pool to which to cede cases, but one could adapt the idea by stipulating that a plan could continue to treat an enrollee, but instead of receiving a capitated payment the plan would be reimbursed in whole or in part under traditional Medicare for some specified fraction of its enrollees whom it named in advance. This could be kept invisible to the enrollee and be done solely as a transaction between Medicare and the plan. Van Barneveld and colleagues estimated 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.\(^3\) In short, plans could effectively place their highest-cost cases in traditional Medicare, thereby receiving additional reimbursement and mitigating the financial hit from those cases.

Because plans might only obtain information about a person’s expected use of services over time, this proposal may be useful mainly for persons who had been enrolled in the plan for a period of time. Nevertheless, this proposal may be especially useful for dealing with the terminally ill, as we now discuss.

**Medicare and the terminally ill.** As is well known, those who die in any year incur disproportionately high costs.\(^3\) Although the figures vary slightly from year to year, the 5 percent of Medicare beneficiaries who die each year account for about 30 percent of the
dollars. There thus are strong incentives for at-risk plans to avoid the terminally ill and leave this population in traditional Medicare. And 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 that in the traditional program, after adjusting for age and sex. To the degree that the terminally ill are disproportionately in traditional Medicare, the opportunity for managed care to cut waste in their care is not being fully exploited.

Insofar as better risk adjusters would pay additional amounts for those with a diagnosis that has a high risk of mortality, this problem would be mitigated. But for reasons of appearance, one would not want to pay more simply because an enrollee died, just as PPS does not pay more for cases of persons who die in the hospital. If a plan can cede certain persons back to traditional Medicare, however, these might well be those it identifies as terminally ill, which would avoid the appearances problem. If the plan remained partially at risk for patients it ceded, Medicare expenses could be reduced. Even without the blend, however, the managed care plan might be less expensive for terminally ill patients because it might use similar treatment methods for its Medicare and commercial business. Ceding a certain fraction of patients, of course, is not mutually exclusive with introducing diagnostic-based risk adjusters or partial capitation for those patients not ceded.

**Concluding Comment**

As the Medicare at-risk program grows, lack of satisfactory risk adjustment is causing serious problems. Better risk adjusters now exist and should be introduced into the program, in our view more rapidly than they are being introduced. Our judgment, however, is that these risk adjusters are not likely to reduce the current problems to negligible levels and that one could do even better by combining these methods with partial capitation; that is, making some payments based on actual use of services.

Furthermore, the better risk adjusters are available to health plans. In principle, plans could use them to improve their ability to select by altering marketing, service intensity, or location strategies. Thus, unless Medicare improves its payment methods, it may overpay health plans to an even greater extent in the future.

The authors are grateful to The Commonwealth Fund for research support. No endorsement of the findings by the fund is implied. The authors thank Randy Ellis and three anonymous referees for helpful comments on a preliminary draft.
NOTES


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

3. Airlines before deregulation were in a similar position to the one that health plans are in today; their fares were regulated such that long-distance routes, typically between large cities, were profitable, and short-haul routes, typically to small towns, were unprofitable. Similar to open enrollment requirements for health plans, 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; most notably, they scheduled frequent flights, with the result that load factors on long-distance routes were relatively low. C. Winston, “Economic Deregulation: Days of Reckoning for Microeconomists,” *Journal of Economic Literature* (September 1993): 1263–1289. 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 for passengers paying their own fares.

4. Assuming they could buy the membership at the same rate as the health plan.


7. Some literature explains the variance in spending across groups rather than individuals. Of course, a higher fraction of the variation in group spending can be explained because random events are averaged within the group. But explaining group spending is only relevant if the group as a whole is enrolling in a plan (for example, an employer enrolls all employees in a single group). If individuals are free to choose plans, as they are in Medicare, it is the variance across individuals that is relevant because the group is not a fixed set of persons.

8. Plans must, of course, be reimbursed at least the average amount of such random expense; otherwise, they will not want to participate.


10. Ibid.

11. We do know that as explained variation improves, incentives to select do not diminish proportionately. See Newhouse et al., “Adjusting Capitation Rates;” and Newhouse, “Reimbursing Health Plans and Health Providers.”
12. D. Cutler, M. McClellan, and J. Newhouse, “Prices, Output, and Productivity in Managed Care Insurance” (Mimeo, National Bureau of Economic Research, Cambridge, Massachusetts, 1997), for example, found that nonelderly persons enrolled in Massachusetts HMOs who have a heart attack receive approximately the same proportion of catheterization and revascularization as do those enrolled in an indemnity plan. (Cost, however, is much lower in the HMOs because prices paid for treatment are much lower.) That is, use of these invasive procedures seems little affected by HMO enrollment. By contrast, 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. J. Newhouse and the Insurance Experiment Group, Free for All? Lessons from the RAND Health Insurance Experiment (Cambridge: Harvard University Press, 1993). Together, these findings indicate that not only do at-risk plans treat conditions differently than indemnity plans do, but that any reduction in actual services will not be uniform across conditions.

13. The analog in the prospective payment system (PPS) is getting the weights across diagnosis-related groups (DRGs) right.

14. For example, one is not likely to have a serious auto accident or a heart attack every month.


17. 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.


20. Dunn and colleagues found ADGs better than DCGs using data from the population under age sixty-five but did not estimate results for HCCs. Ellis et al., “Diagnosis-Based Risk Adjustment for Medicare Capitation Payments,” and Weiner et al., “Risk-Adjusted Medicare Capitation Rates.”

21. Ibid.

22. To convert these figures to profit or loss rates, divide by the mean spending of about $3,800.

23. The asymmetry between the profit from cream skimming and the loss from not dumping reflects the skewness of the distribution of spending, which is not mirrored in the reimbursement formula.


27. The ideal payment would be at the marginal cost that would induce the optimal quantity, which is positive. See Pauly, *Doctors and Their Workshops*, chap. 4. If combined with ideal capitation, this would address both concerns of selection and stinting. In practice, we do not know marginal cost and thus will not have an ideal fee structure. The argument about the trade-off between selection and efficient production thus reduces to the system, which in practice is likely to be closer to the ideal. A different argument is that if a manager keeps the entire dollar from any savings, he or she puts the optimal amount of effort into efficient production. See Newhouse, “Reimbursing Health Plans and Health Providers.” This, however, is mostly an argument against cost-based reimbursement; it does not argue strongly against paying plans a set rate for additional procedures. Related to this point, patients may prefer the higher-cost discretionary option and may have been willing to pay the extra premium for it. Willingness to pay is a standard economic test for efficiency.

28. 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 medicine there can be too little incentive to find cost-reducing technologies.

29. Although in principle such a demonstration could be introduced in a budget-neutral fashion, experience suggests that this may not happen because of improved coding after implementation, analogous to DRG creep when PPS was introduced. G. Carter, J. Newhouse, and D. Relles, “How Much Change in the Case Mix Index Is DRG Creep?” *Journal of Health Economics* 9, no. 4 (1990): 411–428. These costs from upcoding would, however, be one-time and are likely to be incurred whenever improved risk adjustment is introduced. Partially offsetting the expenditure effects of improved coding, some plans that now benefit from favorable selection may withdraw from the at-risk program; because they were being paid an average amount for a below-average risk mix, the government would save money, at least in the short run, if enrollees were to revert to the traditional program.


31. If the reinsurancence were financed by reducing the overall capitation payment (analogous to the outlier payments under PPS), profitability on each case would be reduced by the size of the reduction in the overall payment.

32. As pointed out in note 27, the ideal blend pays a plan a prospectively set marginal cost and a capitation amount such that the plan breaks even on that case. See Pauly, *Doctors and Their Workshops*; and A. Shleifer, “Yardstick Competition,” *The RAND Journal of Economics* (Autumn 1985): 319–327.


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