

The Structure of Risk Adjustment for Private Plans in Medicare

Joseph P. Newhouse, PhD; Jie Huang, PhD; Richard J. Brand, PhD; Vicki Fung, PhD;
and John Hsu, MD, MBA, MSCE

For a quarter of a century, Medicare has risk adjusted its payments to private plans that accept at-risk contracts and participate in Part C of the Medicare program, now known as Medicare Advantage (MA). Risk adjustment means Medicare pays plans more for enrollees who are expected to use more services and less for enrollees who are expected to use fewer services, thereby better matching enrollee reimbursement to expected use and, most important, minimizing plan incentives to select against the sick. For example, Medicare pays more for a beneficiary with cancer than for a beneficiary with no chronic disease.

Medicare's initial risk adjustment system, introduced in 1985, accounted for only the enrollee's age, sex, county of residence, institutional status, Medicaid eligibility (for the noninstitutionalized), and whether a working beneficiary had employment-based insurance that was primary. Although Medicare paid plans more for enrolling older beneficiaries than younger beneficiaries, reflecting their greater medical spending, the risk adjustment system at that time included no direct measures of health status.

Without adequate risk adjustment, a plan could make money by attracting low-cost healthy beneficiaries with a tightly managed low-premium product, whereas a plan enrolling high-cost beneficiaries would not be paid commensurately more and could lose money.¹ In particular, if expected profitability varies with the health status of the beneficiary, plans have incentives to manage care for different services more or less aggressively. Services that are predictable by the beneficiary (and can be used as a device for selection), that are predictive of total medical costs, and that are less profitable are the services a plan would cut back on or manage aggressively to deter enrollment by enrollees who want those services, and conversely.² Ellis and McGuire² provide evidence that service-level selection is operative in Medicare.

Because plan reimbursement was set by Congress in 1985 at 95% (later 100%) of predicted spending in traditional Medicare (TM) in the beneficiary's county, the enrollment of those whose utilization was less than this amount resulted in increased overall spending by Medicare.³⁻⁷ For example, the Congressional Budget Office⁴ estimated in 1994 that

Medicare was paying 8% more to private plans than it would have paid had the same beneficiaries instead been enrolled in TM. In 2006, Medicare changed from

Medicare bases its risk adjustment method for Medicare Advantage plan payment on the relative costs of treating various diagnoses in traditional Medicare. However, there are many reasons to doubt that the relative cost of treating different diagnoses is similar between Medicare Advantage plans and traditional Medicare, including the varying applicability of care management methods to different diagnoses and the varying degrees of market power among suppliers of services to plans. We use internal cost data from a large health plan to compare its cost of treating various diagnoses with Medicare's reimbursement. We find substantial variability across diagnoses, implying that the current risk adjustment system creates incentives for Medicare Advantage plans to favor beneficiaries with certain diagnoses, but find no consistent relationship between the costliness of the diagnosis and the difference between reimbursement and cost.

(Am J Manag Care. 2011;17(6):e231-e240)

In this article

Take-Away Points / e232

Published as a Web exclusive

www.ajmc.com

For author information and disclosures,
see end of text.

Take-Away Points

The Centers for Medicare and Medicaid Services–Hierarchical Condition Categories (CMS-HCC) system that Medicare uses to reimburse health plans establishes relative prices for different diagnoses based on fee-for-service system data. This makes the implicit assumption that health plans reduce costs equiproportionately across diagnoses. This article tests that assumption.

- We overwhelmingly reject that the relative cost of diagnoses in the health plans in our sample is the same as that in the CMS-HCC.
- The magnitude of the errors is large, exceeding 100% for some HCCs.
- This is particularly true for uncommon conditions for which certain providers have substantial market power but Medicare uses lower administered prices.

paying plans a take-it-or-leave-it price to a bidding system, but the bids were for an average risk mix, and the actual monies paid to the plans continued to be adjusted for the risk mix of the plan's actual enrollees. Therefore, to the degree the risk adjustment system remains inadequate, the MA program continues to be vulnerable to selection problems.

In response to recommendations of the Prospective Payment Assessment Commission and the Physician Payment Review Commission (the predecessors of the Medicare Payment Advisory Commission), the Balanced Budget Act⁸ mandated that Medicare introduce a risk adjustment method that would account for the beneficiary's health status in addition to demographic variables, such as age and sex, that it already used. In response, the Centers for Medicare and Medicaid Services (CMS [née Health Care Financing Administration]) contracted for the development of a risk adjustment method that became known as CMS–Hierarchical Condition Categories (CMS-HCCs). In addition to most of the demographic factors that the prior risk adjustment system had used, this method also adjusted plan reimbursement on the basis of diagnoses recorded on claims from the prior year.

In 2000, an initial version of this method, based solely on diagnoses from inpatient claims, was implemented.⁹ This initial version was used through 2003, but to avoid an incentive to hospitalize a beneficiary simply to record a diagnosis and obtain higher reimbursement, the method initially applied only to 10% of the reimbursement to plans; the remaining 90% was based on the prior system that did not use diagnostic data. Starting in 2004, diagnostic data from outpatient claims were considered reliable enough to use to reimburse, and a transition began such that the new system of CMS-HCCs, which used diagnoses recorded in both the inpatient and outpatient settings, by 2007 applied to 100% of the Part C payment.¹⁰ Although the developers initially distinguished 189 categories of conditions, the final set of CMS-HCCs that CMS implemented had only 70 categories, which balanced concerns about coding feasibility, adequate sample size, and predictive accuracy.

The principal task of the new risk adjustment method was to distinguish how much more or less expensive beneficiaries with a given diagnosis or diagnoses were to treat (in terms of total Medicare-covered services) relative to beneficiaries with other diagnoses. A key problem that CMS and its contractor faced in answering this question was that the data to determine the relative costliness of those with various diagno-

ses were available only from TM. Using data on relative costs in TM to pay for treatment in MA presumes that the relative cost of diagnoses in the managed system has the same structure as TM (ie, any reduction in cost from care management is equiproportionate across all diagnoses or CMS-HCCs).

The equiproportionality assumption is unlikely to hold for several reasons. We have known for many years that between-area variation in rates of procedures is related to the degree of physician discretion in treating the condition.^{11–15} For diagnoses with less discretion in choice of treatment, there is less variation across areas. This finding seems likely to carry over into managed care as well, meaning that diagnoses with little discretion will be treated similarly to TM and those with more discretion may be treated in a less costly fashion. For example, plans could selectively contract with those physicians and hospitals that are more conservative in their management of conditions with more physician discretion or could place them in a favorable tier such that patients would pay less to use them.

In addition, health plan interventions in the care process, such as integration of care, utilization management, and prior authorization, almost certainly affect different conditions differently. For example, in the case of chronic problems, long-term compliance with a treatment regimen is often important, and health plans can influence compliance with prescribed medication for a chronic disease, such as diabetes mellitus, through benefit design and disease management.¹⁶ Such interventions are unlikely to be appropriate or used for emergency treatment, such as that of an aortic dissection. In general, acute problems tend to be less amenable to care management than chronic problems because rapid action by the physician on the scene may be important in treating an acute problem and no intervention by a third party may be feasible.

Furthermore, the number and effectiveness of potential treatment options vary across medical conditions, limiting the ability of any plan to achieve gains through more efficient care or better contracting. At one extreme, if there is only a single treatment option available nationally from a single manufacturer, the cost of this option may not vary

substantially. At the other extreme, when there are multiple treatment options with comparable effectiveness (eg, diuretic prescription drugs), plans could obtain favorable prices through negotiation and encourage greater use of the highest-value option through care management programs. Even if there is only 1 standard treatment option, its cost may vary among local providers (eg, procedures performed at teaching vs nonteaching hospitals). Furthermore, if the degree of improvement in compliance that managed care can achieve and the resulting reduction in cost varies across chronic conditions, as is surely the case, the equiproportionate assumption would also fail.

One dramatic, but dated, demonstration of differences in treatment patterns between managed and “unmanaged” care comes from the RAND Health Insurance Experiment,¹⁷ which randomized families of beneficiaries younger than 65 years to a staff model health maintenance organization (HMO) in the 1970s and others to an indemnity insurance plan like TM that had no utilization management features (but had cost sharing similar to the staff model HMO). Those assigned to the HMO had similar ambulatory visit rates but 40% lower hospitalization rates and, using the same set of price weights for specific services to impute spending, 28% less spending.^{17,18} Therefore, for diagnoses that do not customarily result in hospitalization, the CMS-HCC weights that one would have estimated using data on the cost incurred by the staff model HMO probably would have differed little from the weights estimated using data from those with indemnity insurance, whereas for many of the diagnoses with nontrivial proportions hospitalized, the weights would have differed substantially simply from the lower frequency of hospitalization.

Finally, health plans contract with providers who have varying degrees of market power. Particularly for rare or unusual diagnoses, there may be only 1 local or even only 1 regional provider. That provider can probably obtain higher reimbursement from managed care plans relative to Medicare’s administratively set reimbursement compared with other diagnoses for which there are more numerous local providers, thus giving plans a stronger negotiating position.

In short, the relative price structure that Medicare uses to reimburse at-risk MA plans almost certainly contains errors, because the cost structure of health plans across persons with various diagnoses likely differs substantially from that of TM. The aim of this study was to determine how different the structure of the risk adjustment scheme would be if it were based on health plan costs rather than data from TM and whether the scheme creates incentives to select the healthy and avoid the sick. We do not examine actual selection by plans or beneficiaries.

METHODS

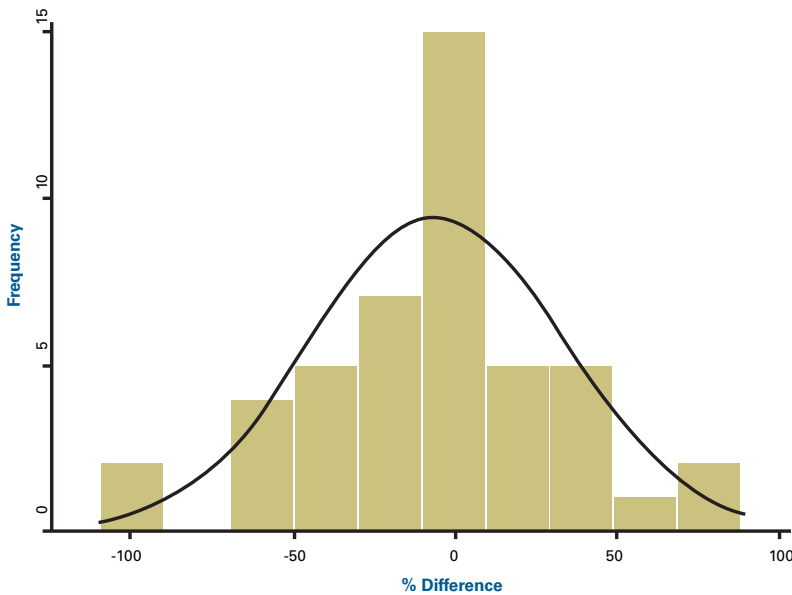
We studied more than 300,000 Medicare beneficiaries enrolled in risk contracts in a large MA-HMO plan during 2006 and 2007. This MA-HMO insurer offered multiple types of benefit arrangements to its beneficiaries. The MA risk adjustment system does not adjust for differences in benefit arrangements between MA and TM, and neither do we. Data from 2005 provided prior year diagnoses for 2006. We included beneficiaries who had aged into Medicare eligibility by the beginning of the study year; we excluded the institutionalized because of small numbers. The 2006 cohort consisted of 322,237 persons, and the 2007 cohort comprised 336,507 persons, as the plan gained enrollment over this period.

Per our protocol, we have masked the identity of the plan. As is generally the case, the hospitals and physicians in the MA-HMO we studied treat both Medicare and commercially insured enrollees; therefore, the mix of providers and capital equipment are configured to treat both types of enrollees. In other words, Medicare reimbursement is not the only influence on the choices of inputs and in turn on the cost structure by disease.

We included all Medicare beneficiaries in the MA-HMO who were enrolled throughout the year before the study year and throughout the study year (2006 or 2007), as well as those who died during a study year. We allowed up to a 2-month gap in membership in a calendar year because we conjecture that such gaps represent data processing errors rather than true changes in membership status. However, we believe that any medical claims from this gap are included in our utilization data. For the decedents, we replicated the methods by Pope et al¹⁰; we annualized the decedent’s spending by multiplying by 12, divided by the number of months eligible, and then weighted the observation by the reciprocal of that ratio. We excluded the small number (3%-4%) who did not have coverage for both Parts A and B or who left the plan in the middle of the year. Excluding the nondecedents with less than a full year of eligibility avoided the issue of how best to annualize spending for beneficiaries who disenrolled because they left the service area or changed plans in the middle of a calendar year and had missing spending data.

We obtained internal cost accounting data for the services the plan provided to beneficiaries; these data approximate (or are proportional to) total allowed charges. We then replicated the methods that Pope et al¹⁰ used to derive the CMS-HCC weights, substituting the plan’s cost accounting data at the beneficiary level for the TM spending data that Pope et al used. We compared the resulting coefficients of the HCCs with the coefficients of the CMS-HCC from the model by Pope et al that CMS used to reimburse plans. To simplify

■ **Figure 1.** Histogram of Percentage Differences in Traditional Medicare vs Medicare Advantage Coefficients, 2006^a



^aThe differences shown are for the 46 hierarchical condition categories whose standard error was less than 20% of the absolute difference between the weights from traditional Medicare and from the Medicare Advantage plan, divided by the mean of these 2 values. The bin width is 20 percentage points.

comparison of the distribution of MA-HMO values with the values in the study by Pope et al, we rescaled the distribution of MA-HMO values to have the same mean as the distribution of the values estimated from TM. Rescaling (ie, normalizing the weights) corrects for any factor that affects all HCCs proportionately.

Because the MA-HMOs' geographic distribution of beneficiaries differs from the national TM distribution that Pope et al used in their calculations, absolute spending levels (and hence absolute values of weights) could differ for reasons that are incidental to our purposes herein, such as variation in nominal wage levels. Moreover, it is well known that there are geographic differences in the treatment of various conditions,¹⁹ an additional reason why there could be differences in relative costs by geography. We make no effort to reweight our data to match the TM geographic distribution but note that, from the point of view of Medicare policy, any geographic differences in relative costs simply add to any distortions in the current structure because the structure of the CMS-HCC is the same across all regions.

We proceeded by following the specification by Pope et al.¹⁰ We regressed the annual accounting cost for each beneficiary in the MA-HMO sample on the following dummy variables: age and sex (24 categories), Medicaid- \times -sex- \times -disability (4 categories), HCC (70 categories), HCC- \times -disability (5 categories), and HCC interactions (6 categories). This gave a total of 81 HCC-related coefficients that we compared with those

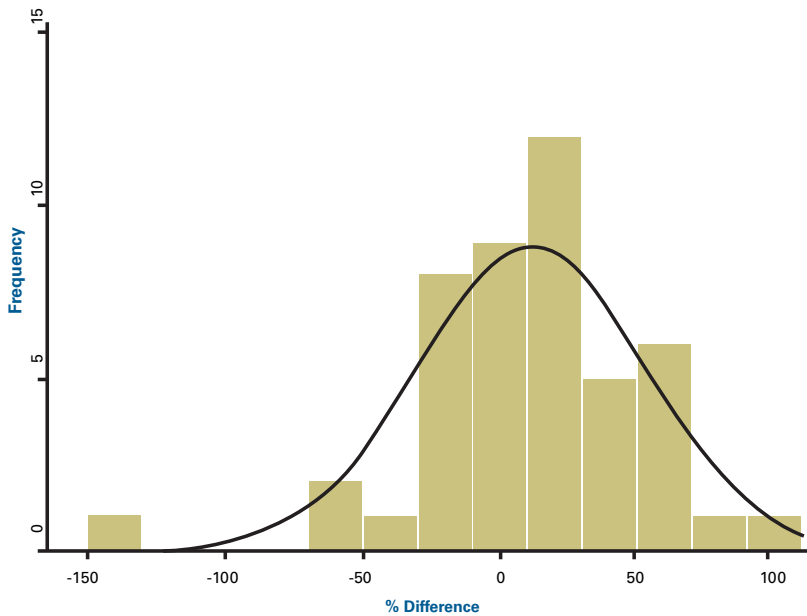
in the study by Pope et al. After initial estimation, we constrained several of the MA-HMO coefficients such that categories with a higher ranking in the disease hierarchy would have at least as high predicted costs. Specifically, we constrained the coefficients of the following HCCs to be equal: hcc008 = hcc009, hcc067 = hcc068, hcc081 = hcc082, hcc107 = hcc108, hcc075 = hcc154, and hcc161 = hcc177. These coefficients should be monotonically ordered but were not in our sample, presumably because of small numbers. Our specification differed from that by Pope et al in a minor aspect: because we were unable to determine which of the older beneficiaries may have been eligible for Medicare before age 65 years because of disability, we did not estimate intercept terms corresponding to older beneficiaries who were disabled before

age 65 years and those who were not. Specifically, Pope et al included interaction terms for originally disabled and sex (originally_disabled- \times -female and originally_disabled- \times -male). Instead, our estimates effectively are the average intercept over all older persons. Because most older beneficiaries were not eligible before age 65 years for reasons of disability (about 15% of all Medicare beneficiaries are younger than 65 years at any point in time) and because those who were eligible because of disability rather than by becoming 65 years old are distributed throughout the HCCs, any bias from this difference should not materially affect our results.

We compared our results for 2006 with the relative risk scores from the 2004 version of the CMS-HCC software, which CMS used for payment in 2006 and are based on the values by Pope et al.¹⁰ The values that CMS used in 2005 and 2006 were based on a 5% sample of 1999 and 2000 data from TM. We similarly compared our 2007 results with the results from the 2007 CMS-HCC software.

As a descriptive comparison of the risk adjustment structure derived from the MA-HMO data with that derived from the TM data, we computed the percentage differences between the values derived from the MA-HMO data and from the TM data for those HCCs whose values in the MA-HMO data were estimated with a sufficient degree of precision (specifically, whose standard error was <20% of the absolute difference between the weights from TM and from the MA plans, divided by the mean of these 2 values). For this purpose, we

■ **Figure 2.** Histogram of Percentage Differences in Traditional Medicare vs Medicare Advantage Coefficients, 2007^a



^aThe differences shown are for the 45 hierarchical condition categories whose standard error was less than 20% of the absolute difference between the weights from traditional Medicare and from the Medicare Advantage plan, divided by the mean of these 2 values. The bin width is 20 percentage points.

rescaled the MA-HMO values to have the same mean as the TM values over all 81 coefficients.

We adopted the one-fifth cutoff for the descriptive statistic to represent a compromise between having enough HCCs to obtain a meaningful distribution of differences between MA-HMO and TM values but also to be sufficiently precise that we did not show large deviations simply from sampling error, especially in the MA-HMO estimates; 46 HCCs in 2006 and 45 HCCs in 2007 satisfied this criterion.

If the equiproportionate assumption held, these differences would bunch around zero. Some differences could arise from differences in relative input prices that led to a different mix of inputs in MA-HMO treatment; other differences could arise from varying distributions of the demographic variables or illness severity within HCCs between TM and the MA-HMO sample. There is also a negligible discrepancy because the means of the TM and MA-HMO distributions were adjusted to be equal across the 81 coefficients rather than across the 46 (in 2006) and 45 (in 2007) used in the descriptive analysis. Nonetheless, it seems unlikely that differences from these causes would be large. However, differences in treatment patterns owing to utilization management could be large. Some of those differences could arise from input price or demographic differences, but most of them probably arise from differences between the incentives

that full capitation offers and those of an unmanaged fee-for-service reimbursement environment, as well as differences in market power in contracting with providers.

As already noted, some differences between the 2 sets of weights would be expected from sampling error; even with more than 300,000 observations, some of the HCCs have only a few hundred observations in the MA-HMO data, and the values by Pope et al¹⁰ have a sample only about 6 times as large (Pope et al used a 5% sample of 2 years of data), so that sampling error is also relevant to the values by Pope et al. We describe herein a formal test across all the coefficient values of whether the relative price structures are similar.

RESULTS

Figure 1 and **Figure 2** show the results of the percentage differences

between the weights derived from TM and from the MA-HMO sample for those HCCs in which the standard error of the MA-HMO values is less than 20% of the difference between the TM and MA-HMO coefficients (46 HCCs in 2006 and 45 HCCs in 2007). **Table 1** and **Table 2** give values for these HCCs.

Ignoring sampling error for the moment, if there were good agreement between the weights, almost all of the mass in the 2 histograms would be concentrated near zero because each distribution has approximately the same overall mean. As is readily apparent, this is not the case; there are substantial deviations between the MA-HMO cost accounting data and Medicare reimbursement in both years.

We want to formally test the hypothesis that the vector of the CMS-HCC coefficients equals the corresponding vector of values for the same HCCs estimated on MA-HMO data, and we must account for the sampling error to do so. Unfortunately, we only have the published CMS-HCC weights and their published standard errors from the study by Pope et al¹⁰; we do not have the covariance terms with the demographic variables for the CMS-HCC weights. (The HCCs themselves are orthogonal.) Because we do not have raw TM claims data, to carry out a formal test, we must make the following 3 assumptions that are known not to hold exactly but may hold approximately: (1) The MA-HMO and TM coefficients are

■ **Table 1.** Differences Between Traditional Medicare and Medicare Advantage Coefficients in 2006^a

HCC	Description	No.	TM	MA	TM – MA	% Difference
HCC002	Septicemia/shock	2246	0.890	0.664	0.226	29.09
HCC007	Metastatic cancer and acute leukemia	3069	1.464	1.037	0.427	34.14
HCC008	Lung/upper digestive/other severe cancer	2234	1.464	1.037	0.427	34.14
HCC009	Lymphatic/head/neck/brain/major cancer	3555	0.690	0.631	0.059	8.99
HCC010	Breast/prostate/colorectal/other cancer	21,914	0.233	0.228	0.005	2.11
HCC015	Diabetes with renal manifestation	16,874	0.764	0.359	0.405	72.09
HCC016	Diabetes with neurological/peripheral circulatory manifestation	11,820	0.552	0.248	0.304	76.02
HCC017	Diabetes with acute complications	616	0.391	0.767	-0.376	-64.97
HCC025	End-stage liver disease	371	0.900	1.514	-0.614	-50.86
HCC026	Cirrhosis of liver	1055	0.516	0.601	-0.085	-15.30
HCC031	Intestinal obstruction/perforation	2378	0.408	0.496	-0.088	-19.56
HCC032	Pancreatic disease	1962	0.445	0.479	-0.034	-7.42
HCC038	Rheumatoid arthritis/inflammatory connective tissue	9371	0.322	0.430	-0.108	-28.72
HCC044	Severe hematological disorders	1367	1.011	0.710	0.301	34.99
HCC051	Drug/alcohol psychosis	1108	0.353	0.672	-0.319	-62.23
HCC054	Schizophrenia	1136	0.543	0.662	-0.119	-19.82
HCC055	Major depressive, bipolar, paranoid	9533	0.431	0.476	-0.045	-9.90
HCC067	Quadriplegia, other extensive paralysis	333	1.181	1.208	-0.027	-2.29
HCC068	Paraplegia	331	1.181	1.208	-0.027	-2.29
HCC071	Polyneuropathy	19,814	0.268	0.257	0.011	4.02
HCC073	Parkinson and Huntington disease	3718	0.475	0.565	-0.090	-17.26
HCC074	Seizure disorders and convulsions	4186	0.269	0.405	-0.136	-40.29
HCC077	Respirator dependence/tracheostomy status	174	2.102	3.300	-1.198	-44.35
HCC079	Cardiorespiratory failure and shock	3678	0.692	0.698	-0.006	-0.82
HCC080	Congestive heart failure	25,814	0.417	0.434	-0.017	-3.90
HCC081	Acute myocardial infarction	3731	0.348	0.318	0.030	8.99
HCC082	Unstable angina/other acute ischemic heart disease	3046	0.348	0.318	0.030	8.99
HCC083	Angina pectoris/old myocardial infection	27,112	0.235	0.153	0.082	42.23
HCC092	Specified heart arrhythmias	29,081	0.266	0.323	-0.057	-19.46
HCC096	Ischemic or unspecified stroke	6096	0.306	0.361	-0.055	-16.62
HCC100	Hemiplegia/hemiparesis	6702	0.437	0.253	0.184	53.40
HCC104	Peripheral vascular disease with compensation	3681	0.677	1.058	-0.381	-43.89
HCC105	Peripheral vascular disease	16,873	0.357	0.360	-0.003	-0.92
HCC107	Cystic fibrosis	4	0.376	0.345	0.031	8.46
HCC108	Chronic obstructive pulmonary disease	29,854	0.376	0.345	0.031	8.46
HCC130	Dialysis status	2048	3.076	8.391	-5.315	-92.70
HCC131	Renal failure	10,407	0.576	0.514	0.062	11.45
HCC148	Decubitus ulcer of skin	2294	1.030	1.445	-0.415	-33.56
HCC149	Chronic ulcer of skin, except decubitus	3359	0.484	0.899	-0.415	-60.01
HCC157	Vertebral fracture without spinal cord injury	4855	0.490	0.311	0.179	44.56
HCC158	Hip fracture/dislocation	3504	0.392	0.324	0.068	19.05
HCC161	Traumatic amputation	17	0.843	0.638	0.205	27.73
HCC164	Major complication of medical care/trauma	5657	0.262	0.390	-0.128	-39.32
HCC174	Major organ transplant status	265	0.722	2.341	-1.619	-105.72
HCC176	Artificial openings for feeding/elimination	1478	0.790	0.844	-0.054	-6.59
HCC177	Amputation status/lower limb/amputation complete	1304	0.843	0.638	0.205	27.73

HCC indicates hierarchical condition category; MA, Medicare Advantage; TM, traditional Medicare.

^aThe 46 HCCs shown are those HCCs whose MA–Health Maintenance Organization standard errors are less than 20% of the difference between the TM and MA-HMO coefficients.

Risk Adjustment for Private Plans in Medicare

■ Table 2. Differences Between Traditional Medicare and Medicare Advantage Coefficients in 2007^a

HCC	Description	No.	TM	MA	TM – MA	% Difference
HCC002	Septicemia/shock	2247	0.887	0.517	0.370	52.66
HCC007	Metastatic cancer and acute leukemia	3414	1.648	1.027	0.621	46.45
HCC008	Lung/upper digestive/other severe cancer	2306	1.648	1.027	0.621	46.45
HCC009	Lymphatic/head/neck/brain/major cancer	3838	0.771	0.596	0.175	25.69
HCC010	Breast/prostate/colorectal/other cancer	24,367	0.258	0.222	0.036	15.04
HCC015	Diabetes with renal manifestation	20,154	0.608	0.320	0.288	62.06
HCC016	Diabetes with neurological/peripheral circulatory manifestation	11,478	0.452	0.223	0.229	67.78
HCC025	End-stage liver disease	456	0.996	1.695	-0.699	-51.94
HCC026	Cirrhosis of liver	1178	0.519	0.545	-0.026	-4.97
HCC032	Pancreatic disease	2255	0.383	0.391	-0.008	-2.00
HCC037	Bone/joint/muscle infection/necrosis	1381	0.550	0.602	-0.052	-9.01
HCC038	Rheumatoid arthritis/inflammatory connective tissue	10,591	0.363	0.329	0.034	9.93
HCC044	Severe hematological disorders	1604	1.136	0.763	0.373	39.22
HCC045	Disorders of immunity	729	0.841	1.064	-0.223	-23.45
HCC052	Drug/alcohol dependence	3263	0.250	0.337	-0.087	-29.73
HCC054	Schizophrenia	1263	0.515	0.517	-0.002	-0.35
HCC055	Major depressive, bipolar, paranoid	11,460	0.370	0.396	-0.026	-6.74
HCC067	Quadriplegia, other extensive paralysis	361	0.961	1.235	-0.274	-24.96
HCC068	Paraplegia	404	0.961	1.235	-0.274	-24.96
HCC071	Polyneuropathy	22,218	0.324	0.254	0.070	24.35
HCC073	Parkinson and Huntington disease	4061	0.547	0.481	0.066	12.85
HCC074	Seizure disorders and convulsions	4922	0.280	0.261	0.019	7.01
HCC079	Cardiorespiratory failure and shock	4406	0.629	0.512	0.117	20.43
HCC080	Congestive heart failure	28,187	0.395	0.307	0.088	25.07
HCC081	Acute myocardial infarction	3531	0.349	0.295	0.054	16.73
HCC082	Unstable angina/other acute ischemic heart disease	3753	0.332	0.295	0.037	11.76
HCC083	Angina pectoris/old myocardial infection	31,384	0.231	0.099	0.132	79.62
HCC092	Specified heart arrhythmias	32,771	0.295	0.290	0.005	1.60
HCC100	Hemiplegia/hemiparesis	7503	0.410	0.215	0.195	62.27
HCC104	Peripheral vascular disease with compensation	3967	0.645	0.794	-0.149	-20.74
HCC105	Peripheral vascular disease	20,465	0.324	0.371	-0.047	-13.39
HCC107	Cystic fibrosis	7	0.398	0.304	0.094	26.70
HCC108	Chronic obstructive pulmonary disease	33,575	0.398	0.304	0.094	26.70
HCC111	Aspiration/specific bacterial pneumonias	1068	0.761	0.547	0.214	32.80
HCC119	Proliferative diabetic retinopathy/vitreous hemorrhage	1519	0.278	0.336	-0.058	-18.96
HCC130	Dialysis status	2114	1.432	8.064	-6.632	-139.68
HCC131	Renal failure	20,153	0.389	0.332	0.057	15.67
HCC148	Decubitus ulcer of skin	2617	1.167	0.764	0.403	41.72
HCC149	Chronic ulcer of skin, except decubitus	3940	0.463	0.751	-0.288	-47.51
HCC157	Vertebral fracture without spinal cord injury	5527	0.501	0.247	0.254	67.81
HCC161	Traumatic amputation	91	0.736	0.619	0.117	17.23
HCC164	Major complication of medical care/trauma	6566	0.299	0.339	-0.040	-12.65
HCC174	Major organ transplant status	318	1.073	2.129	-1.056	-65.97
HCC176	Artificial openings for feeding/elimination	1709	0.758	0.691	0.067	9.27
HCC177	Amputation status/lower limb/amputation complete	1486	0.653	0.619	0.034	5.30

HCC indicates hierarchical condition category; MA, Medicare Advantage; TM, traditional Medicare.

^aThe 45 HCCs shown are those HCCs whose MA–Health Maintenance Organization standard errors are less than 20% of the difference between the TM and MA-HMO coefficients.

independent. (2) The estimated variances have no sampling error; that is, we use the published standard errors as if they were true. (3) The covariances between the demographic variables and the HCC dummy variables are ignorable.

Each CMS-HCC coefficient estimate (and corresponding MA-HMO coefficient estimate) is proportional to a conditional mean of the cost for a group of patients with a given diagnosis configuration and, by the central limit theorem, is approximately normally distributed. After adjusting the means of the 2 distributions of estimated coefficients to be equal, we calculate the difference between each of the corresponding HCC coefficients (eg, the coefficient for HCC1 from the study by Pope et al¹⁰ and the coefficient for HCC1 from the MA-HMO data). The distribution of the difference between the 2 vectors of means is also normally distributed, and each element has a mean of zero under the null. Assuming independence, the variance of the difference in each element of the HCC and MA-HMO means is simply the sum of the variance of each mean. Therefore, dividing each estimated difference in the 2 means by the square root of the sum of the variances gives a standardized $N(0,1)$ variable. The sum of those variables squared over the 81 coefficients for 2006 (and similarly for 2007) is then distributed as χ^2_{81} .

In symbols, let the estimated coefficient for HCC_{*i*} in the MA-HMO sample be a_i and the corresponding estimated coefficient in TM be b_i ($i = 1 - 81$). Rescale the distribution of the a_i to have the same mean as the distribution of the b_i . The distribution of each $(a_i - b_i) / [\text{sqrt}(\sigma_{a_i}^2 + \sigma_{b_i}^2)]$ is approximately $N(0,1)$, so $\sum_{i=1-81} [(a_i - b_i) / \text{sqrt}(\sigma_{a_i}^2 + \sigma_{b_i}^2)]^2 = \chi^2_{81}$.

The resulting test statistics for the 2006 and 2007 distributions have χ^2 statistics of 2546 and 5299, respectively, both of which have $P < .005$. In short, subject to the approximations we have made in assuming that the statistics we calculated are distributed as χ^2 , we can overwhelmingly reject the null that the relative price structures of the MA-HMO and TM are the same. Eighty-one is the number of estimated coefficients, including the interaction terms that use HCC dummy variables. The critical value for a χ^2_{81} distribution at $P = .005$ is 126, more than an order of magnitude less than the test statistics we obtained.

Although there are some appreciable differences between the TM relative prices and the MA-HMO relative prices, it should not be surprising that there is nonetheless a strong relationship between the absolute levels of the 2 prices across the HCCs. Diagnoses that are expensive to treat in TM are also generally expensive to treat in MA-HMOs. The correlation coefficient between the 2006 TM and MA-HMO relative prices is 0.87 and between the 2007 TM and MA-HMO relative prices is 0.55. However, despite the seemingly different appearance of the 2 figures, the differences shown in Figures 1 and 2 are stable between the 2 years; the correlation coef-

ficient (r) of the differences using the 41 HCCs that are in common between the 2 years is 0.98.

The foregoing is strictly statistical, but there is also economic content in the results. One HCC that appears substantially more expensive in the MA-HMO sample is renal dialysis, HCC130. Patients who require renal dialysis often receive their dialysis in centers specializing in dialysis and from physicians in these centers who manage a range of the patients' conditions. Most important for our purposes, the renal dialysis industry is concentrated into a small number of companies, which have market power in local markets. The MA-HMO pays the market rate for these patients. We see a similar phenomenon with major organ transplants (HCC174), which also tend to be concentrated in a few major centers in the United States. As with renal dialysis, major organ transplant centers have market power. Therefore, these diagnoses seem expensive in the MA-HMO compared with the TM, where market power is not relevant.

We also tested whether the old adage that entities paid by capitation had incentives to select against sicker people still held with the new risk adjustment system. To do so, we computed the correlation between the difference of the TM and MA-HMO coefficients and the level of the TM coefficients. This correlation was -0.668 ($P < .01$) in 2006 and -0.281 ($P = .05$) in 2007 for the 46 and 45 HCCs listed in Tables 1 and 2. In other words, CMS-HCC with higher values seem unprofitable, appearing to suggest that the old adage continues to hold. However, closer inspection revealed that the values are driven by the outlier value for renal dialysis. If renal dialysis is omitted, the correlation for 2006 changes from -0.668 to -0.155 (not significant) and for 2007 from -0.281 to 0.171 (not significant). In other words, while the risk adjustment scheme has substantial errors for individual HCCs, there is no detectable pattern with respect to the size of the HCC weight, implying that the old adage likely no longer holds. In that sense, the CMS-HCC must be regarded as a success.

DISCUSSION

Medicare bases its reimbursement to MA health plans on the relative costliness of treating various maladies, but it computes that relative costliness using data from TM. Such data are likely to be in error when applied to health plans for many reasons, including that the treatment of some conditions is more amenable to the medical management techniques used by the plan than the treatment of other conditions and that plans face providers and suppliers with varying degrees of market power across diagnoses. Our results suggest not only that there are likely errors in the pricing structure Medicare uses

Risk Adjustment for Private Plans in Medicare

but also that those errors may well be large. Most important, these errors do not seem to be correlated with predicted risk, suggesting that simple strategies aimed at selecting lower-risk patients would not result in favorable selection.

Particularly notable are HCC130 (renal dialysis) and HCC174 (major organ transplants) because they are expensive in the MA-HMOs in our sample compared with TM. Whereas TM pays for renal dialysis and major organ transplants with administratively set take-it-or-leave-it prices, MA-HMOs pay market prices for these services. We believe that these diagnoses are unprofitable for the HMO because Medicare can exploit its ability to set prices to a greater degree for these diagnoses than for other diagnoses for which plans face a more competitive supplier market.

Crude selection strategies based simply on predicted risk (ie, skim the healthy) are unlikely to yield favorable selection not only because of the poor correlation between predicted risk and the payment distortions that we found but also because older persons often have multiple concurrent conditions. In other words, a plan enrolls the whole person, and for a beneficiary with multiple conditions, the reimbursement may be too generous for one condition but too skimpy for another.

Despite the mitigation that comes from having to enroll an entire patient, our results imply that the risk adjustment structure used by Medicare makes beneficiaries with certain conditions more or less profitable for MA plans than beneficiaries with other conditions. These distortions create incentives for plans to specialize in certain conditions, that is, to select for and against certain conditions in how they structure their networks and formularies with the objective of attracting or not attracting beneficiaries with specific types of conditions.

One can ask whether specialization by MA plans in certain conditions could be desirable in the sense that it might minimize social cost. There are at least 3 problems with this reasoning. One, as just mentioned, many Medicare beneficiaries have multiple conditions, and it does not seem optimal to have relatively good service or access for one condition but not for another. Second, a beneficiary may develop a new condition, in which case the beneficiary may want to change physicians or drugs and would need to change plans in the extreme case. This also does not seem optimal. Third, social cost is not the same as budget cost; in other words, to the degree that Medicare uses the same real resources to treat a condition but simply pays less because of its monopsony power, there is no necessary desirability in having TM specialize in that diagnosis. Inferences about economic efficiency in health care markets, however, are problematic for many reasons.²⁰

Our work is subject to the obvious limitation that it comes from a single MA insurer, and we do not know to what de-

gree these results would generalize to other insurers. However, even if on average MA plans replicated TM, any heterogeneity among insurers would leave problematic incentives for individual plans to engage in selection.

A second limitation is that we cannot exclude the possibility that some of the differences we found between MA costs and TM costs are attributable to diagnostic coding because coding has an element of endogeneity. Song et al²¹ recently showed that TM beneficiaries who moved to regions with greater intensity of services than their region of origin had substantially greater increases in CMS-HCC risk scores (ie, more coded diagnoses or higher-weighted diagnoses) than beneficiaries who moved to regions with the same or lower intensity and suggested that more diagnostic testing in regions of higher intensity led to additional diagnoses being recorded. Plans have an incentive to code diagnoses more completely than physicians treating TM beneficiaries (although less incentive to use a large number of tests to do so) because their reimbursement depends on the CMS-HCC score, whereas physician reimbursement in TM does not turn on a beneficiary's diagnoses. However, hospital reimbursement in TM, like health plan reimbursement, does depend on the coding of diagnoses, and this should serve to lessen any coding differences between TM and MA.

A third limitation is that the hypothesis test of the null that the MA coefficients computed on MA data are the same as the TM coefficients had to make some assumptions that are known not to hold; these assumptions can be relaxed, and one can compute an exact test with TM claims data. Nonetheless, the violations of the assumptions seem sufficiently weak (given the degree to which the test statistics reject the null) that it seems unlikely an exact test would overturn these results. Finally, CMS required that the risk adjustment through 2006 remain budget neutral (ie, the total MA pie did not decrease because of the phase-in of the CMS-HCC risk adjustment approach), whereas starting in 2007, more of the MA pie reflected the actual risk of enrolled beneficiaries. This budget neutrality phase of the risk adjustment implementation could have mitigated some of the incentives to the extent that beneficiaries enrolled in MA differed from those enrolled in TM.

Part C of Medicare has the advantage relative to TM that CMS does not have to set thousands of prices for individual services that inevitably will diverge from cost, which is burdensome and subject to the political process. Rather, private plans negotiate prices with providers. Such prices will, of course, reflect the market power of the providers that treat patients with various HCCs in the plan's local market (and the plan's market power if it has a large market share), as indeed they appear to for renal dialysis and major organ transplants. Nonetheless, on balance, the negotiated prices between plans

and providers may be closer to cost than the administratively set prices in TM are, a potential strength of Part C relative to TM. However, this work emphasizes that there is also an element of administered pricing in Part C because risk adjustment is a form of administratively set prices. Our results also suggest that the current methods for risk adjustment share a common problem of administratively set prices in that they can substantially depart from cost.

Acknowledgment

We thank James H. Ware for helpful discussions.

Author Affiliations: From Harvard University (JPN, JTH), Boston, MA; Division of Research (JH, VF, JTH), Kaiser Permanente, Oakland, CA; and University of California, San Francisco (RJB).

Funding Source: This study was funded by grant P01 032952 from the National Institute on Aging.

Author Disclosures: Dr Newhouse reports that he is a director of and holds equity in Aetna, which sells Medicare Advantage plans. Drs Huang and Fung are employed by Kaiser Permanente, which also sells Medicare Advantage plans. Dr Brand reports serving as a paid consultant for Kaiser Permanente Division of Research. Dr Hsu reports receiving grants from the NIH, whose research affects CMS.

Authorship Information: Concept and design (JPN, Dr Hsu); acquisition of data (JPN, Dr Hsu); analysis and interpretation of data (JPN, JH, RJB, VF, JH); drafting of the manuscript (JPN, Dr Hsu); critical revision of the manuscript for important intellectual content (JPN, JH, VF, JH); statistical analysis (JPN, RJB); obtaining funding (JPN); administrative, technical, or logistic support (JH, VF, JH); and supervision (Dr Hsu).

Address correspondence to: Joseph P. Newhouse, PhD, Harvard University, 180 Longwood Ave, Boston, MA 02115. E-mail: joseph_newhouse@harvard.edu.

REFERENCES

- Rothschild M, Stiglitz JE.** Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Q J Econ.* 1976;90(4):629-650.
- Ellis RP, McGuire TG.** Predictability and predictiveness in health care spending. *J Health Econ.* 2007;26(1):25-48.
- Brown RS, Clement DG, Hill JW, Retchin SM, Bergeron JW.** Do health maintenance organizations work for Medicare? *Health Care Financ Rev.* 1993;15(1):7-23.
- Congressional Budget Office.** *Effects of Managed Care: An Update.* Washington, DC: Congressional Budget Office; 1994.
- Physician Payment Review Commission and Prospective Payment Assessment Commission.** *Joint Report to the Congress on Medicare Managed Care.* Washington, DC: Physician Payment Review Commission; 1995.
- Physician Payment Review Commission.** *Annual Report to Congress, 1996.* Washington, DC: Physician Payment Review Commission; 1996.
- Medicare Payment Advisory Commission.** *Report to the Congress: Improving Risk Adjustment in Medicare.* Washington, DC: Medicare Payment Advisory Commission; 2000.
- Balanced Budget Act of 1997.** Pub L No. 105-33.
- Pope GC, Ellis RP, Ash AS, et al.** Principal inpatient diagnostic cost group model for Medicare risk adjustment. *Health Care Financ Rev.* 2000;21(3):93-118.
- Pope GC, Kautter J, Ellis RP, et al.** Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev.* 2004;25(4):119-141.
- McPherson K, Wennberg JE, Hovind OB, Clifford P.** Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. *N Engl J Med.* 1982;307(21):1310-1314.
- Phelps CE.** Information diffusion and best practice adoption. In: Culyer AJ, Newhouse JP, eds. *Handbook of Health Economics.* Vol 1. Amsterdam, the Netherlands: Elsevier; 2000:223-264.
- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL.** The implications of regional variations in Medicare spending, part 1: the content, quality, and accessibility of care. *Ann Intern Med.* 2003;138(4):273-287.
- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL.** The implications of regional variations in Medicare spending, part 2: health outcomes and satisfaction with care. *Ann Intern Med.* 2003;138(4):288-298.
- Wennberg JE, Fisher ES, Stukel TA, Sharp SM.** Use of Medicare claims data to monitor provider-specific performance among patients with severe chronic illness. *Health Aff (Millwood).* 2004;suppl variation: VAR5-VAR18.
- Wennberg DE, Marr A, Lang L, O'Malley S, Bennett G.** A randomized trial of a telephone care-management strategy. *N Engl J Med.* 2010;363(13):1245-1255.
- Newhouse JP.** *Free for All: Lessons From the Health Insurance Experiment.* Boston, MA: Harvard University Press; 1993.
- Manning WG, Leibowitz A, Goldberg GA, Rogers WH, Newhouse JP.** A controlled trial of the effect of a prepaid group practice on use of services. *N Engl J Med.* 1984;310(23):1505-1510.
- Dartmouth Medical School.** *The Dartmouth Atlas of Health Care.* Chicago, IL: AHA Press; 1999.
- Arrow KJ.** Uncertainty and the welfare economics of medical care. *Am Econ Rev.* 1963;53(5):941-973.
- Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES.** Regional variations in diagnostic practices [published correction appears in *N Engl J Med.* 2010;363(2):198]. *N Engl J Med.* 2010;363(1):45-53. ■