Do Financial Incentives in a Globally Budgeted Healthcare Payment System Produce Changes in the Way Patients Are Categorized? A Five-year Study

Laura A. Petersen, MD, MPH; Tracy H. Urech, MPH; Margaret M. Byrne, PhD; and Kenneth Pietz, PhD

Objective: To assess the responses to financial incentives after a change in the payment system in a capitation-style healthcare payment system over a 5-year period.

Study Design: Cross-sectional and longitudinal examination of cost, utilization, and diagnostic data.

Methods: Using Veterans Health Administration (VHA) administrative data on healthcare users between fiscal years 1998 and 2002, we calculated the proportion of new patients entering each of the payment classes, the illness burden of patients entering the payment classes, and the profitability index (a ratio of payment to costs) for each class suspected of gaming and each control class. Our main dependent variables of interest were the differences in the measures between each utilization-based class and each diagnosis-based class. We used 2 different analytic approaches to assess whether these differences increased or decreased over time.

Results: No clear evidence of gaming behavior was present in our results. A few comparisons were significant, but they did not show a consistent pattern of responses to incentives. For example, 6 of 16 comparisons of profitability index were significant, but (contrary to the hypothesis) 4 of these had a negative value for the time parameter, indicating decreasing profitability in the utilization-based classes versus the diagnosis-based classes.

Conclusions: Although the payment system could be manipulated to increase payment to VHA networks, no such consistent gaming behavior was observed. More research is needed to better understand the effects of financial incentives in other healthcare payment systems.

(Am J Manag Care. 2007;13:513-522)

For author information and disclosures, see end of text.

Financial incentives within healthcare payment systems are increasingly recognized as powerful drivers of hospital service use and may dictate the healthcare services provided to patients, regardless of clinical need.[1-4] Because of the direct relationship between the financing of healthcare services and the utilization of such services, changes in the payment system (eg, the institution of diagnosis-related groups for Medicare prospective payment[5]) can dramatically transform clinical and administrative practices.[6,7] For example, Kahn et al concluded from their comparison of Medicare hospitalizations between 1981-1982 and 1985-1986 that although overall patient outcomes (eg, 30-day mortality, in-hospital mortality) were not negatively affected by the change to a prospective system, patients were more likely to be discharged prematurely.[8,9] Chan et al showed that the economic incentives in Medicare's system of payment for rehabilitation hospitals encouraged a dramatic increase in length of stay during the base year.[10] This is because payments were capped at the base year level, with incentives given if charges per patient were reduced in subsequent years.

Although more is known about the response of individuals[11-13] or hospitals[8-10,14] to payment system incentives, less is known about responses of a healthcare delivery system to incentives in the payment system. Thus, our objective was to study the response of a large federal healthcare system to incentives to increase payments to a network (consisting of multiple integrated hospitals and outpatient clinics in a geographical area). We studied the Veterans Health Administration (VHA). The VHA is the largest integrated healthcare system in the United States, with a medical care and research budget of $32.52 billion in fiscal year (FY) 2005.[15] The VHA healthcare system cares for its enrollees in Department of Veterans Affairs (VA) Medical Centers and VA Community Based Outpatient Clinics, which employ salaried physicians. The VHA budget is appropriated by Congress as a global budget and distributed to its 21 VHA networks via a form of capitation. Capitation payments are 1 of 2 flat payments per enrolled veteran meeting vesting criteria, and are paid whether the veteran uses the services or not. It is important to note that in the VHA the networks all compete for a fixed budget. Thus, if 1 network is able to increase its budget relative to other networks, those other networks will receive a smaller share of the total budget. There are only 2 ways to increase a network share of this

In this issue
Take-away Points / p521
www.ajmc.com
Full text and PDF

budget. One way is to increase the number of patients cared for; a potentially more effective way to increase this share is to ensure that patients that are already cared for in the system qualify for the higher of the 2 capitation payments.

The incentives to engage in this second behavior (termed *gaming*) arise when the difference between the high and low payment classes is larger than the marginal cost of moving a patient over the utilization threshold (ie, when the payment/cost difference is larger for the higher payment class). Thus, there may be an incentive to move patients into those classes when this action may not be completely justified by clinical need. Because the VHA changed its method of payment to networks in 1997, the opportunity arose for a natural experiment to examine whether the presence of financial incentives affected how patients were classified. We studied payment classes for which a response to incentives was possible and might be financially advantageous, and compared them with payment classes for which there was less discretion in assigning patients to these categories (ie, gaming behavior would be less likely to occur). This work is important because in a healthcare system with a fixed global budget, efforts to increase payment for patients in this way may result in overuse of unnecessary healthcare services, effectively restricting access for others who need such services. If all units of a healthcare system are competing for a fixed budget, the incentive for networks and facilities in a globally budgeted healthcare system is to overprovide certain services without considering costs to the group. That may result in unequal distribution of services over time in the face of equivalent patient need. For example, patients needing care later in the budget cycle are competing for the residual of the fixed budget.[16]

Because the VHA pays healthcare providers via salary, providers do not experience incentives that are aligned with the interests of the network. In this situation, the ways that networks or hospitals can exert influence on the healthcare providers who deliver care include structural changes in service availability (eg, expanding services and hiring staff for lucrative clinical programs, downsizing less lucrative ones) and/or increasing efforts to recruit certain types of patients through outreach or screening activities (a version of cream skimming or preferred risk selection).[17]

We are not aware of studies evaluating responses to financial incentives in healthcare systems where payment to physicians and administrators is via salary and global budgeting. If salaried or globally budgeted healthcare delivery systems exhibit responses to incentives noted in other settings, this would suggest that even more attention should be paid to aligning incentives for improved healthcare than is currently being paid in the healthcare policy arena[18,19] through pay-for-performance and other types of programs.[11] Better understanding of responses to financial incentives is important, both to prevent unintended adverse effects and because using healthcare payment systems to reward quality or meet other societal goals has tremendous untapped potential.[20]

The measures we selected for this study follow from our hypotheses regarding the effects of gaming. We hypothesized that in profitable patient payment classes that are defined by meeting certain thresholds for health services utilization (such as bed days of care [BDOC]), the proportion and the profitability of patients in these classes would increase over time, and the illness burden of new patients would decrease over time as patients who are less ill enter the class. These changes would occur as utilization of services was increased for patients who were near meeting the utilization threshold necessary for higher payment. In concurrent control classes, which were not defined by utilization, no such behavior would occur. Because we expected that the influence the networks or hospitals can exert on the healthcare providers to effect changes in capitation classification are structural or involve recruitment of patients, we did not expect that such changes would be evident immediately after the VHA changed its payment system in 1997, but would show gradual responses over time. Therefore, we studied several years of data to assess the presence of the hypothesized response.

METHODS

VHA Payment System

The Veterans Equitable Resource Allocation (VERA) system was implemented in April 1997 to allocate the congressionally appropriated VHA medical care budget to the VHA networks. VERA assigns patients to 1 of 45 (as of FY 2002) patient class categories. Assignment to a patient class category may be based strictly on diagnosis (eg, stroke class), a combination of utilization and diagnosis criteria (eg, BDOC for chronic posttraumatic stress disorder [PTSD]), or strictly utilization (eg, home-based primary care [HBPC]). The patient classes are designated as being either VERA basic or VERA complex. Patients in VERA basic classes (eg, ear, nose, and throat) are accessing the healthcare system for routine healthcare and do not require use of VHA special care programs, whereas patients in VERA complex classes (eg, traumatic brain injury [TBI]) utilize special care programs and may require long periods of inpatient care and rehabilitation. Each year a basic and a complex "price" is calculated.

For example, in FY 1998 the VERA basic patient allocation was $2604, and the VERA complex patient allocation was $36 960. Because of this wide disparity in payments, although only 4% of total VHA users are complex care patients, they account for more than one third of the allocated monies.

During the years of study, the majority of VHA healthcare funds were allocated by utilization criteria. For example, criteria for patients with PTSD included the requirement of at least 60 BDOC. If a PTSD patient had 59 BDOC or less, the patient was in the basic care price group ($2604). If the patient had 60 BDOC, they were in the complex care group ($36 960). This type of payment differential may provide incentives to increase certain utilization parameters to move patients from the basic care to the complex care price (regardless of need), particularly when the ratio of price paid to actual cost favors such action.

This study was approved by the Baylor College of Medicine institutional review board.

Study Population

We studied all VHA users from FY 1998 to FY 2002. VHA healthcare users were identified as those veterans who received any services from the VHA during that FY. Veterans with recorded ages of <17 or >120 years (<.01% of patients) were removed from the dataset.

We studied 2 groups of patients: those assigned to payment classes that might be subject to gaming and those assigned to payment classes that are not likely to be susceptible to gaming. To identify payment classes that were vulnerable to gaming, we identified complex care classes that had explicit utilization criteria in the patient classification definition, including a minimum number of outpatient visits or inpatient treatment days. In addition, we only included classes that had a constant class definition over time to ensure that observed responses were not due to a change in the patient classification definition. Finally, we only included classes that were profitable at a national level during the initial year studied, and thus would be worth gaming. We refer to these payment classes as the utilization-based classes.

Payment classes unlikely to be manipulated consisted of complex care classes that had no utilization criteria in the definition and again had the same patient classification definition for at least 3 years. Control group payment classes could be, but did not have to be, profitable in the initial year. We refer to these payment classes as the diagnosis-based classes.

Of the 26 VERA complex care classes, only 5 met our inclusion criteria for features that could be gamed and only 4 met the criteria for the control group. None of the 19

Table 1. Patient Classification Criteria for the Utilization-based and Diagnosis-based Classes

VERA Complex Class	Patient Classification Criteria	Type of Class*
PTSD	*ICD-9-CM* code 309.8 *and* more than 60 inpatient days of PTSD treatment	Utilization
HBPC	At least 10 visits for home health care services	Utilization
DOM	At least 31 combined bed days of care in a domiciliary or community nursing home setting *and* twice as many days in a long-term care setting as in acute care *and* more days in a domiciliary than in a community nursing home	Utilization
Substance Abuse	*ICD-9-CM* code 291, 292, 303, 304, or 305 *and* at least 180 days of inpatient care	Utilization
Blind rehabilitation	At least 1 overnight stay in the blind rehabilitation bed section	Utilization
Stroke	Primary diagnosis of stroke, cerebrovascular accident, or occlusion (*ICD-9-CM* codes 430-436) *or* diagnosis of late effects of cerebrovascular disease (*ICD-9-CM* code 438)	Diagnosis
TBI	Primary diagnosis of TBI (*ICD-9-CM* codes 800, 801, 803, 804, 851, 852, 853, 854) *or* patient with a rehabilitation procedure (V57) *and* a secondary diagnosis of TBI (including *ICD-9-CM* codes 310.2, 905.0, and 907.0)	Diagnosis
SCI para-old	*ICD-9-CM* code 344.1 or 907.2 *and* patient was previously classified as a SCI paraplegic–new injury in the prior year	Diagnosis
SCI quad-old	*ICD-9-CM* code 344.0 *and* patient was previously classified as a SCI quadriplegic–new injury in the prior year	Diagnosis

*Utilization-based classes include criteria such as a minimum number of outpatient visits or inpatient treatment days. Diagnosis-based classes do not include any utilization criteria.
VERA indicates Veterans Equitable Resource Allocation; *ICD-9-CM*, *International Classification of Diseases, Ninth Revision, Clinical Modification*; PTSD, posttraumatic stress disorder; HBPC, home-based primary care; DOM, domiciliary; TBI, traumatic brain injury; SCI para-old, spinal cord injury paraplegic–old injury; SCI quad-old, spinal cord injury quadriplegic–old injury.

Table 2. Characteristics of New Patients in the Utilization-based and Diagnosis-based Classes, Fiscal Years 1998-2002*

VERA Complex Class*	No. New Patients†	Mean (SD) Age, y	Median (SD) Cost, $	Mean (SD) Diagnostic Cost Group Relative Risk Score	Mean (SD) Network Profitability Index
PTSD (U)					
1998	3856	50.0 (5.7)	23 382 (20 845)	1.56 (1.17)	1.42 (0.31)
1999	2057	50.6 (6.1)	28 250 (21 356)	1.70 (1.26)	1.17 (0.30)
2000	1983	51.7 (6.0)	31 344 (20 369)	1.75 (1.25)	1.22 (0.24)
2001	1930	52.3 (6.5)	32 316 (26 346)	1.82 (1.27)	1.10 (0.15)
2002	1419	53.2 (6.5)	28 452 (21 021)	1.69 (1.21)	1.27 (0.19)
HBPC (U)					
1998	8009	68.0 (13.0)	18 678 (28 801)	3.33 (2.82)	1.36 (0.21)
1999	7353	68.6 (13.0)	17 928 (30 939)	3.38 (2.83)	1.35 (0.26)
2000	9545	69.3 (12.9)	15 772 (29 176)	3.10 (2.73)	1.67 (0.28)
2001	9543	69.7 (12.5)	15 188 (31 557)	3.11 (2.75)	1.65 (0.28)
2002	9721	69.6 (13.0)	15 253 (29 272)	2.91 (2.57)	1.71 (0.26)
DOM (U)					
1998	3178	46.6 (9.5)	26 930 (16 077)	1.97 (1.25)	1.27 (0.21)
1999	3435	46.8 (9.1)	27 197 (16 343)	1.97 (1.21)	1.26 (0.21)
2000	3390	47.7 (8.9)	29 813 (20 393)	2.06 (1.21)	1.30 (0.25)
2001	3391	48.0 (8.6)	31 754 (19 306)	2.11 (1.14)	1.26 (0.21)
Substance abuse (U)					
2000	1524	48.0 (8.5)	38 301 (26 324)	2.18 (1.22)	1.03 (0.20)
2001	1613	48.4 (8.1)	43 904 (29 682)	2.28 (1.16)	0.92 (0.16)
2002	1724	48.5 (7.5)	43 272 (37 544)	2.16 (1.02)	0.90 (0.21)
SCI para-old (D)					
1998	1291	57.6 (14.5)	18 894 (44 273)	3.47 (2.59)	1.14 (0.29)
1999	1040	59.0 (14.5)	21 553 (45 805)	3.85 (2.76)	1.09 (0.30)
2000	1101	58.4 (14.0)	22 408 (48 941)	3.89 (2.85)	1.12 (0.28)
2001	1175	58.8 (13.7)	19 661 (56 240)	3.68 (2.70)	1.23 (0.47)
2002	911	59.2 (14.4)	19 838 (34 926)	3.71 (2.67)	1.45 (0.47)
SCI quad-old (D)					
1998	984	56.9 (15.0)	20 237 (56 383)	3.76 (3.42)	0.94 (0.26)
1999	939	58.2 (14.5)	24 917 (64 295)	4.09 (3.51)	0.80 (0.20)
2000	850	57.6 (14.3)	24 967 (59 070)	4.15 (3.39)	0.97 (0.32)
2001	946	58.4 (14.3)	26 018 (66 399)	4.08 (3.08)	0.98 (0.32)
2002	802	58.4 (13.7)	21 762 (36 099)	3.60 (2.70)	1.30 (0.23)
TBI (D)					
1998	896	59.7 (17.5)	19 165 (35 043)	3.54 (3.02)	1.24 (0.27)
1999	961	57.7 (18.5)	19 868 (36 637)	3.74 (3.04)	1.24 (0.29)
2000	792	60.0 (18.6)	27 315 (38 674)	4.04 (3.14)	1.10 (0.15)
2001	814	59.2 (18.7)	29 709 (44 077)	3.94 (3.07)	1.10 (0.29)
2002	811	62.1 (17.3)	21 544 (34 956)	3.54 (2.90)	1.33 (0.31)

(Continued)

Table 2. Characteristics of New Patients in the Utilization-based and Diagnosis-based Classes, Fiscal Years 1998-2002 *(Continued)**

VERA Complex Class*	No. New Patients†	Mean (SD) Age, y	Median (SD) Cost, $	Mean (SD) Diagnostic Cost Group Relative Risk Score	Mean (SD) Network Profitability Index
Stroke (D)					
1998	13 683	69.3 (10.6)	16 983 (31 371)	3.49 (2.66)	1.35 (0.16)
1999	8669	69.7 (10.8)	22 202 (32 103)	4.19 (2.59)	1.16 (0.13)
2000	8539	69.8 (10.9)	23 633 (35 244)	4.16 (2.57)	1.23 (0.16)
2001	8469	70.0 (11.0)	23 866 (37 652)	4.07 (2.53)	1.20 (0.17)
2002	5804	68.9 (11.4)	19 789 (29 633)	3.83 (2.37)	1.47 (0.21)

*Utilization-based classes include criteria such as a minimum number of outpatient visits or inpatient treatment days. Diagnosis-based classes do not include any utilization criteria.
†The proportion of total patients in the Department of Veterans Affairs healthcare system that was new to any of the classes for any study year was less than 1%.
VERA indicates Veterans Equitable Resource Allocation; SD, standard deviation; PTSD, posttraumatic stress disorder; U, utilization-based class; HBPC, home-based primary care; DOM, domiciliary; SCI para-old, spinal cord injury paraplegic–old injury; D, diagnosis-based class; SCI quad-old, spinal cord injury quadriplegic–old injury; TBI, traumatic brain injury.

VERA basic care classes met the study inclusion rules. The utilization-based classes included HBPC, chronic PTSD, domiciliary, substance abuse, and blind rehabilitation. The domiciliary classification definition changed for FY 2002, so we only reported results for FY 1998 to FY 2001. We report results for complex substance abuse for FYs 2000-2002 because FY 2000 was the first fiscal year the group was profitable at the national level. The diagnosis-based classes included spinal cord injury (SCI) quadriplegic–old injury, SCI paraplegic–old injury, TBI, and stroke. Table 1 lists the patient classification definitions.

Measures

We examined 3 different measures for each class: the proportion of VHA users who enter the class (new users), the average profitability of the new users, and the average illness burden of the new users. We restricted the analyses to new patients to eliminate the confounding effect of different rates of patients leaving the classes due to mortality and other reasons.

For each study year, we calculated the proportion of new users in a given payment class at the network level as the total number of new patients in that payment class divided by the total number of patients in the network. Second, we calculated the network-level profitability index (PI) of each complex class as the sum of the payments for those new patients assigned to the particular payment class divided by the sum of the network-level costs associated with those patients. For example, if a VHA network for a given FY has 50 VERA complex substance abuse patients that generate an overall VERA payment of $1.85 million but only cost $1.25 million to care for, the resulting PI for this group of patients is 1.48.

Third, an illness burden measure for each veteran user was calculated using a diagnosis-based risk adjustment methodology (Medicare algorithm) developed and licensed by the DxCG Company (Boston, Mass).[21] In the diagnostic cost group algorithm, more than 15 000 *International Classification of Diseases, Ninth Revision, Clinical Modification* codes are partitioned into 545 clinically homogeneous groups called DxGroups. Patients may fall into multiple DxGroups. DxGroups are clustered into 118 clinically homogeneous condition categories according to similarities in resource use. Again, patients may fall into multiple condition categories. Condition categories that are related and affect the same organ system are then arranged hierarchically to create 118 hierarchical condition categories. Patients are assigned the most severe condition they have within a hierarchy. A patient may have more than 1 hierarchical condition category, but only 1 from any given hierarchy. The DxCG software uses regression models to assign predicted costs to hierarchical condition categories, and then individual total predicted costs are calculated as the sum of the predicted hierarchical condition category costs. An individual's diagnostic cost group relative risk score (RRS) is the individual's total predicted cost divided by the average predicted cost of the algorithm population. Thus, individuals with higher illness burden have a higher RRS. This software has been used successfully for case-mix adjustment in a growing number of studies[22] and is the method currently used to risk-adjust Medicare+Choice capitated plans.[23-25] Using DxCG software version 6, we assigned

RRSs to all VHA users in each year from FY 1998 to FY 2002. For each FY, all patient diagnoses in that year were collected from VHA databases that record diagnoses (outpatient clinic file; patient treatment files including acute care, extended care, and non-VHA care; and the fee basis file for care received outside of the VHA but paid for by the VA). In previous work, we have shown that the diagnostic cost group methodology has strong clinical validity in the VHA population, with risk of mortality increasing monotonically with more severe diagnostic cost group categories.[26]

Statistical Analysis

We used a retrospective matched pair design to determine the behavior over time of each of the 3 study measures (proportion of new users, network-level PI, and RRS) for each of the utilization-based classes compared with the diagnosis-based classes. For each network, each utilization-based group and diagnosis-based group combination constituted a matched pair. For each of our study measures, our primary interest was whether the difference between the classes in a pair (eg, utilization-based network-level PI minus diagnosis-based network-level PI) increased or decreased over time. For each network, the difference between the classes among patients in that network was calculated. Using the difference ensured that our results would not be affected by systemic changes over the entire VHA patient population (eg, changes in illness burden as measured by RRS). We hypothesized that the diagnostic cost group RRSs for the utilization-based classes would decrease over time relative to those for the diagnosis-based classes. Similarly, we predicted that both network-level PI and overall proportion of patients in a given class would increase in the utilization-based classes compared with the diagnosis-based classes. An increase in network-level PI would occur if patients entering that class had increasingly lower costs over time, indicating a lower illness burden.

Table 3. Statistics on Differences Between the Utilization-based and Diagnosis-based Classes for the Diagnostic Cost Group Relative Risk Score

Utilization-based Class	Diagnosis-based Class	Time Parameter in Repeated Measures Analysis (*P*)	Mean of OLS Slopes (*P*)*
PTSD	SCI para-old	0.042 (NS)	0.013 (NS)
PTSD	SCI quad-old	0.111 (NS)	0.083 (NS)
PTSD	Stroke	–0.015 (NS)	0.00057 (NS)
PTSD	TBI	0.007 (NS)	0.031 (NS)
HBPC	SCI para-old	–0.189 (.0003)	–0.153 (NS)
HBPC	SCI quad-old	–0.018 (NS)	–0.083 (NS)
HBPC	Stroke	–0.144 (<.0001)	–0.165 (<.0001)
HBPC	TBI	–0.156 (NS)	–0.135 (NS)
Substance abuse	SCI para-old	0.074 (NS)	0.082 (NS)
Substance abuse	SCI quad-old	0.259 (NS)	0.253 (NS)
Substance abuse	Stroke	0.153 (NS)	0.153 (NS)
Substance abuse	TBI	0.284 (NS)	0.271 (NS)
DOM	SCI para-old	–0.132 (NS)	–0.103 (NS)
DOM	SCI quad-old	–0.046 (NS)	–0.048 (NS)
DOM	Stroke	–0.146 (<.0001)	–0.144 (<.0001)
DOM	TBI	–0.153 (NS)	–0.155 (NS)

*Significance tests are based on the Wilcoxon signed rank test.
OLS indicates ordinary least squares; PTSD, posttraumatic stress disorder; SCI para-old, spinal cord injury paraplegic–old injury; NS, not significant; SCI quad-old, spinal cord injury quadriplegic–old injury; TBI, traumatic brain injury; HBPC, home-based primary care; DOM, domiciliary.

We selected 4 utilization-based classes (PTSD, HBPC, domiciliary, and substance abuse) and 4 diagnosis-based classes (SCI paraplegic–old injury, SCI quadriplegic–old injury, stroke, and TBI) for a total of 16 utilization/diagnosis pairs. We evaluated a total of 48 comparisons, 16 pairs and 3 outcome measures per pair. Accordingly, we used 0.05/48 = 0.00104 as the significance level in all tests. For each FY and each comparison, there were 21 data points (one for each VA network) with the following exceptions: (1) for the domiciliary analyses, we removed 2 of the 21 networks from the study years (FYs 1998-2001) because they had fewer than 5 patients in the payment class; (2) for the substance abuse analyses, we removed 1 VHA network because of low patient count across the study years examined.

The analyses were done at the network level, and the dependent variable was the difference between the utilization-based and diagnosis-based classes in the study measure. Because the repeated measures at a network over time were expected to be correlated, we used a repeated measures analysis with unstructured covariance. The only fixed effects were the intercept and time. The sign and significance of the time parameter indicated whether the difference variable was changing over time.

Because the significance of parameter estimates in the repeated measures analysis requires an assumption of normality, we used a nonparametric analysis to check the results. A derived variable analysis, also known as a summary statistic analysis,[27] was used for this purpose. The derived variable is a generated variable that characterizes the time history of the difference between the utilization-based and diagnosis-based variables at each network.[27-29] The derived variable used here was the ordinary least squares (OLS) estimated slope.[30] For each network, the OLS slope is the coefficient of time in an OLS model in which time is the single independent variable, and the difference (utilization-based class – diagnosis-based class) in the study measure (eg, RRS) is the dependent variable. The OLS slope provides an indication of whether the values are increasing or decreasing over time and the rate of increase or decrease.

For each outcome measure (difference between the utilization-based class and the diagnosis-based class) and each network, the OLS slope was computed. For each measure, the mean of the OLS slopes over all networks was computed. A Wilcoxon signed rank test was used to determine whether the mean was significantly different from 0.30. The derived analysis is entirely nonparametric.

Table 4. Statistics on Differences Between the Utilization-based and Diagnosis-based Classes for the Profitability Index

Utilization-based Class	Diagnosis-based Class	Time Parameter in Repeated Measures Analysis (*P*)	Mean of OLS Slopes (*P*)*
PTSD	SCI para-old	–0.112 (NS)	–0.104 (NS)
PTSD	SCI quad-old	–0.126 (<.0001)	–0.126 (<.0001)
PTSD	Stroke	–0.0645 (NS)	–0.072 (NS)
PTSD	TBI	–0.038 (NS)	–0.045 (NS)
HBPC	SCI para-old	0.045 (NS)	0.037 (NS)
HBPC	SCI quad-old	–0.023 (NS)	0.015 (NS)
HBPC	Stroke	0.090 (.0002)	0.069 (.0009)
HBPC	TBI	0.082 (.0003)	0.096 (.0001)
Substance abuse	SCI para-old	–0.221 (.0002)	–0.230 (<.0001)
Substance abuse	SCI quad-old	–0.199 (<.0001)	–0.217 (.0002)
Substance abuse	Stroke	–0.128 (<.0001)	–0.173 (<.0001)
Substance abuse	TBI	–0.156 (NS)	–0.189 (NS)
DOM	SCI para-old	0.035 (NS)	0.006 (NS)
DOM	SCI quad-old	–0.027 (NS)	–0.030 (NS)
DOM	Stroke	0.026 (NS)	0.037 (NS)
DOM	TBI	0.068 (NS)	0.077 (NS)

*Significance tests are based on the Wilcoxon signed rank test.
OLS indicates ordinary least squares; PTSD, posttraumatic stress disorder; SCI para-old, spinal cord injury paraplegic–old injury; NS, not significant; SCI quad-old, spinal cord injury quadriplegic–old injury; TBI, traumatic brain injury; HBPC, home-based primary care; DOM, domiciliary.

RESULTS

Between FY 1998 and FY 2002, the total number of VHA patients funded by the VHA payment system increased from 3 195 795 to 3 780 217 patients, an increase of 18.3%. Table 2 displays the characteristics of the new patients in each of the classes we examined. The number of new patients across the classes we studied ranged from 31 897 in FY 1998 to 22 916 in FY 2002, representing a very small percentage of the overall cohort, 1.0% and 0.6%, respectively. Again, although the total proportion of patients is relatively small, the budget impact of these patients is large. Specifically, all patients in these complex payment classes accounted for on average 19% of the healthcare budget for the 5 study years we examined.

Across the study years, the average age of patients was lowest in the domiciliary and highest in the stroke and HBPC classes. New patients in the HBPC, stroke, TBI, SCI paraplegic–old injury, and SCI quadriplegic–old injury were on average almost twice as ill (as measured by diagnostic cost group RRS) as patients in the PTSD and domiciliary classes. Among the utilization-based classes, patients in the substance abuse class had the highest median cost, ranging from $38 301 in FY 2000 to $43 272 in FY 2002. Among the diagnosis-based classes, SCI quadriplegic–old injury had the highest median costs, but the median costs for these patients in FY 2002 were approximately one half of the median costs for patients in the substance abuse class.

Table 3, Table 4, and Table 5 show the results for diagnostic cost group RRS, PI, and proportion of new users, respectively. In each table, the first 2 columns list the utilization-based and diagnosis-based classes. The third column provides the time parameter in the repeated measures analysis. A significant positive value indicates that the difference (utilization-based class – diagnosis-based class) is increasing over time. The last column in each table provides the results of the derived variable analysis.

Reading horizontally across the tables reveals that the 2 tests show consistent results with a single exception (the HBPC vs SCI paraplegic–old injury RRS comparison). A few of the comparisons were significant, but they do not show a consistent pattern of responses to incentives. For example, 6 of 16

Table 5. Statistics on Differences Between the Utilization-based and Diagnosis-based Classes for the Proportion of New Users

Utilization-based Class	Diagnosis-based Class	Time Parameter in Repeated Measures Analysis (*P*)	Mean of OLS Slopes (*P*)*
PTSD	SCI para-old	–0.0001 (<.0001)	–0.00015 (<.0001)
PTSD	SCI quad-old	–0.0001 (<.0001)	–0.00016 (<.0001)
PTSD	Stroke	0.0002 (<.0001)	0.0004 (<.0001)
PTSD	TBI	–0.00006 (.0006)	–0.00016 (<.0001)
HBPC	SCI para-old	0.0001 (NS)	0.0001 (NS)
HBPC	SCI quad-old	0.0001 (NS)	0.0001 (NS)
HBPC	Stroke	0.0005 (<.0001)	0.0007 (<.0001)
HBPC	TBI	0.0001 (NS)	0.0001 (NS)
Substance abuse	SCI para-old	0.00005 (NS)	0.00005 (NS)
Substance abuse	SCI quad-old	0.00002 (NS)	0.00002 (NS)
Substance abuse	Stroke	0.0006 (<.0001)	0.0005 (<.0001)
Substance abuse	TBI	0.00002 (NS)	0.00002 (NS)
DOM	SCI para-old	0.00002 (NS)	0.00001 (NS)
DOM	SCI quad-old	0.00001 (NS)	0.0000054 (NS)
DOM	Stroke	0.0002 (0.0002)	0.0006 (<.0001)
DOM	TBI	0.00003 (NS)	0.00001 (NS)

*Significance tests are based on the Wilcoxon signed rank test.
OLS indicates ordinary least squares; PTSD, posttraumatic stress disorder; SCI para-old, spinal cord injury paraplegic–old injury; NS, not significant; SCI quad-old, spinal cord injury quadriplegic–old injury; TBI, traumatic brain injury; HBPC, home-based primary care; NS, not significant; DOM, domiciliary.

comparisons of PI were significant. However, 4 of these had negative values for the time parameter, indicating decreasing profitability in the utilization-based class relative to the diagnosis-based class. This result was contrary to our hypothesis. Several of the differences in proportion of new users were significant, but some were negative. The only comparison pair that indicated possible gaming across all 3 variables was HBPC and stroke. Because HBPC did not exhibit the same pattern compared with other control classes, this may have been a random occurrence.

DISCUSSION

We studied changes in the way patients were classified and coded over a 5-year period to assess effects of incentives at the payment system level in a healthcare system with a global budget. To control for other factors that were present during the study period, we examined how variables that might indicate gaming moved in the utilization-based classes relative to the diagnosis-based classes. We found that although it was profitable and possible to change the classification of patients by utilization criteria, we found little evidence for such gaming behavior.

It is important to note that in the VHA the networks all compete for a fixed budget. Thus, if 1 network is able to increase its budget relative to other networks, those other networks will receive a smaller share of the total budget. There are only 2 ways to increase a network share of this budget. One way is to increase the number of patients cared for; a potentially more effective way to increase this share is to ensure that patients that are already cared for in the system qualify for the higher of the 2 capitation payments. Such gaming has been shown in other studies. Chan et al showed that the economic incentives in Medicare's system of payment for rehabilitation hospitals encouraged a dramatic increase in length of stay during the base year.[10] This is because payments were capped at the base-year level, with incentives given if charges per patient were reduced in subsequent years. Kahn et al concluded from their comparison of Medicare hospitalizations between 1981-1982 and 1985-1986 that patients were more likely to be discharged prematurely after the implementation of the prospective payment system.[8,9]

Our findings may not be surprising. First, the hospitals we studied were federal hospitals. Second, we assessed whether there was an effect at the network level, where the incentives may not be powerful enough to produce changes in categorization behavior. We also examined effects at the hospital level, but because of very small sample sizes for some of the groups, we did not report them. Physicians employed by the VHA are salaried, and so gain no additional income from participating in gaming, and they have everything to lose if they do not follow stringent compliance and auditing standards.[31,32] Finally, although it was still profitable to increase the number of patients in the study groups, it may be that increases in patient numbers in these classes had already taken place prior to our study years.

Of course, we cannot make causal inferences in this observational study. It is possible that our findings could be due to other trends. However, we would expect that our diagnosis-based classes also would be subject to such effects. Thus, the comparison of utilization-based and diagnosis-based classes should have revealed gaming if it were present.

A large body of empirical research, as well as common sense, supports the assertion that the behavior of institutions and individuals is determined in part by how they are paid. Society generally demands that professionals such as physicians place the good of their patients ahead of financial considerations. It is likely that this trust has been warranted, but it would be naive to believe that financial considerations have no impact on how health professionals care for individual patients. For policy makers, reimbursement reforms can constitute a relatively accessible lever for system change. It is sometimes easier to modify payment arrangements than to change other organizational features.[33]

We found that in a system with a global healthcare budget where providers are paid by salary, but where utilization and thus patient categorization could be manipulated to increase payment to networks, no such consistent gaming behavior was observed. We are not aware of other data evaluating responses to financial incentives in healthcare systems where payment to physicians and administrators is via salary and global budgeting. Better understanding of responses to financial incentives is important, both to prevent unintended adverse effects and because using healthcare payment systems to reward quality or meet other societal goals has tremendous untapped potential.[20] More research is needed to better understand the effects of financial incentives in healthcare financing arrangements.

Take-away Points

Over a 5-year period, we evaluated a capitation-style healthcare system's response to financial incentives by comparing characteristics of patients categorized to utilization-based and diagnosis-based payment classes.

- We found no clear evidence that the system was being manipulated to increase healthcare payments.
- Further research is needed to assess the impact of financial incentives on the delivery of healthcare.

Acknowledgments

We thank Mark Kuebeler, MS, and Michael Thompson, MS, for expert data management and Abel Esquivel for administrative assistance in the preparation of this manuscript.

Author Affiliations: From the Division of Health Policy and Quality, Houston Center for Quality of Care and Utilization Studies, a Health Services Research and Development Center of Excellence, Houston VA Medical Center, and Section for Health Services Research, Baylor College of Medicine, Houston, Tex (LAP, THU, KP); and the Department of Epidemiology and Public Health, University of Miami, Miami, Fla (MMB).

Funding Sources: This work was supported in part by a VA Health Services Research and Development Investigator-Initiated Research Award (grant 04-349) and National Heart, Lung, and Blood Institute grant RO1 HL079173-01. Dr Petersen is a Robert Wood Johnson Foundation Generalist Physician Faculty Scholar (grant 045444) and an American Heart Association Established Investigator awardee (grant 0540043N). Dr Byrne holds a National Cancer Institute career development award.

The views expressed are solely of the authors, and do not necessarily represent those of the Department of Veterans Affairs.

Author Disclosure: The authors (LAP, THU, MMB, KP) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter discussed in this manuscript.

Authorship Information: Concept and design (LAP, THU, MMB, KP); acquisition of data (LAP); analysis and interpretation of data (LAP, MMB, KP); drafting of the manuscript (LAP, MMB,KP); critical revision of the manuscript for important intellectual content (LAP, MMB, KP); statistical analysis (LAP, KP); obtaining funding (LAP); administrative, technical, or logistic support (THU); supervision (LAP).

Address correspondence to: Laura A. Petersen, MD, MPH, Health Services Research and Development (152), Houston Veterans Affairs Medical Center, 2002 Holcombe Blvd, Houston, TX 77030. E-mail: laurap@bcm.tmc.edu.

REFERENCES

1. Petersen LA, Normand SL, Leape LL, McNeil BJ. Regionalization and the underuse of angiography in the Veterans Affairs Health Care System as compared with a fee-for-service system. *N Engl J Med.* 2003;348:2209-2217.

2. Leape LL, Hilborne LH, Bell R, Kamberg C, Brook RH. Underuse of cardiac procedures: do women, ethnic minorities, and the uninsured fail to receive needed revascularization? *Ann Intern Med.*1999; 130:183-192.

3. 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;138:273-287.

4. 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;138:288-298.

5. Chulis GS. Assessing Medicare's prospective payment system for hospitals. *Med Care Rev.* 1991;48:167-206.

6. Baker LC. Association of managed care market share and health expenditures for fee-for-service Medicare patients. *JAMA.* 1999;281:432-437.

7. Feder J, Hadley J, Zuckerman S. How did Medicare's prospective payment system affect hospitals? *N Engl J Med.* 1987;317:867-873.

8. Kahn KL, Keeler EB, Sherwood MJ, et al. Comparing outcomes of care before and after implementation of the DRG-based prospective payment system. *JAMA.* 1990;264:1984-1988.

9. Rogers WH, Draper D, Kahn KL, et al. Quality of care before and after implementation of the DRG-based prospective payment system. A summary of effects. *JAMA.* 1990;264:1989-1994.

10. Chan L, Koepsell TD, Deyo RA, et al. The effect of Medicare's payment system for rehabilitation hospitals on length of stay, charges, and total payments. *N Engl J Med.* 1997;337:978-985.

11. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145:265-272.

12. Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff (Millwood).* 2004;23:127-141.

13. Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev.* 2006;63:135-157.

14. Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. *N Engl J Med.* 1988;318:352-355.

15. Department of Veterans Affairs FY 2005 Annual Performance and Accountability Report. Office of Management. http://www.vagov/budget/Report/2005/HTML/indexhtml. Accessed July 31, 2007.

16. Kuhn M. *Quality in Primary Care: Economic Approaches to Analyzing Quality-Related Physician Behavior.* London, England: Office of Health Economics Research; 2003.

17. Culyer AJ, Newhouse JP. *Handbook of Health Economics.* Amsterdam, the Netherlands: Elsevier; 2000.

18. Berwick DM, DeParle NA, Eddy DM, et al. Paying for performance: Medicare should lead. *Health Aff (Millwood).* 2003;22:8-10.

19. Milgate K, Cheng SB. Pay-for-performance: the MedPAC perspective. *Health Aff (Millwood).* 2006;25:413-419.

20. Dudley RA, Miller RH, Korenbrot TY, Luft HS. The impact of financial incentives on quality of health care. *Milbank Q.* 1998;76:649-686, 511.

21. Ellis RP, Pope GC, Lezzoni L, et al. Diagnosis-based risk adjustment for Medicare capitation payments. *Health Care Financ Rev.* 1996;17:101-128.

22. Rosen AK, Loveland S, Anderson JJ, et al. Evaluating diagnosis-based case-mix measures: how well do they apply to the VA population? *Med Care.* 2001;39:692-704.

23. Centers for Medicare and Medicaid Services. Medicare Advantage (Medicare+Choice, Average Adjusted per Capita Costs) rates and statistics. http://www.cms.hhs.gov/MedicareAdvtgSpecRateStats/. Accessed November 28, 2006.

24. Ash AS, Posner MA, Speckman J, Franco S, Yacht AC, Bramwell L. Using claims data to examine mortality trends following hospitalization for heart attack in Medicare. *Health Serv Res.* 2003;38:1253-1262.

25. Zhao Y, Ash AS, Ellis RP, Slaughter JP. Disease burden profiles: an emerging tool for managing managed care. *Health Care Manag Sci.* 2002;5:211-219.

26. Petersen LA, Pietz K, Woodard LD, Byrne M. Comparison of the predictive validity of diagnosis-based risk adjusters for clinical outcomes. *Med Care.* 2005;43:61-67.

27. Diggle PJ, Heagerty P, Liang K, Zeger S. *Analysis of Longitudinal Data.* New York: Oxford University Press; 2002.

28. Davis CS. *Statistical Methods for the Analysis of Repeated Measurements.* New York, NY: Springer; 2002.

29. Everitt BS. The analysis of repeated measures: a practical review with examples. *The Statistician.* 1995;44:113-135.

30. Snedecor GW, Cochran WG. *Statistical Methods.* Ames, Iowa: Iowa State Press; 1989.

31. Adams DL, Norman H, Burroughs VJ. Addressing medical coding and billing part II: a strategy for achieving compliance. A risk management approach for reducing coding and billing errors. *J Natl Med Assoc.* 2002;94:430-447.

32. Carter D. Medicare cost controls and program compliance: the rationale of physician claims edits. *J Med Pract Manage.* 2002;18:115-119.

33. Rosen B. Professional reimbursement and professional behavior: emerging issues and research challenges. *Soc Sci Med.* 1989;29:455-462. ■