

# Physician Incentives in HMOs

Martin Gaynor

*Carnegie Mellon University and National Bureau of Economic Research*

James B. Rebitzer

*Case Western Reserve University, National Bureau of Economic Research, and Levy  
Economics Institute*

Lowell J. Taylor

*Carnegie Mellon University*

We appreciate helpful comments from Ashish Arora, Simon Burgess, Sue Helper, Carol Propper, Sendhil Mullainathan, and seminar participants at the University of Bristol, Carnegie Mellon University, Case Western Reserve University, University of Chicago, University of Michigan, University of Maryland, NBER, University of North Carolina, Northwestern University, the 2000 American Economic Association annual meetings, and the CEPR workshop on health economics and policy held in Bergen, Norway in May 2002. Rebitzer's research was supported in part by a grant from the Levy Economics Institute. The usual caveat applies.

## **Abstract**

Managed care organizations rely on incentives that encourage physicians to limit medical expenditures, but little is known about how physicians respond to these incentives. We address this issue by analyzing the physician incentive contracts in use at a Health Maintenance Organization (HMO). By combining knowledge of the incentive contracts with internal company records, we examine how medical expenditures vary with the intensity of the incentive to cut costs. Our investigation leads us to a novel explanation for high-powered group incentives—such incentives can improve efficiency in the allocation of resources when the allocation process is based on the professional judgment of multiple agents. Our empirical work indicates that medical expenditures at the HMO are five percent lower than they would have been in the absence of incentives.

# 1 Introduction

Since Coase's (1937) landmark paper on the theory of the firm, economists have devoted increasing attention to the internal workings of organizations. An important development in this literature is the empirical study of incentive systems. Using the internal records of firms and survey data, economists have documented how explicit and implicit incentives shape individual careers and organizational performance.<sup>1</sup> This paper uses internal company records to analyze incentives in a setting that has received relatively little attention from economists but which has important implications for economics and for public policy: managed care organizations. More specifically, we study how primary care physicians in managed care networks respond to incentives to contain medical expenditures.

The empirical study of physician incentives in managed care networks is important for understanding the workings of the health care system in the United States. Managed care is, after all, the nation's dominant form of health insurance and every managed care organization relies on a system of financial and non-financial incentives that encourage physicians in their networks to control costs.<sup>2</sup> The problem of inducing cost-consciousness in physicians has, however, implications that extend beyond the managed care industry. As we discuss below, the HMO's incentive problem is a special case of a broader incentive design issue that can arise whenever principals rely on the judgment of skilled agents to ration resources.

The outline of the paper is as follows. The next section discusses the incentive design problem facing the HMO as well as the key features of the incentive system in place at the firm we study. In Section 3 we combine our knowledge of the HMO's incentive contracts with internal company data to estimate the response of physicians to cost-containment incentives. Section 4 concludes by assessing the broader economic questions raised by our findings.

---

<sup>1</sup>Ichniowski and Shaw (2003), Prendergast (1999), and Gibbons (1997) offer reviews of this literature.

<sup>2</sup>Ninety two percent of all individuals possessing employer-provided health insurance in the United States are enrolled in some form of managed care (Gabel et al., 2000)

## 2 The Incentive System

The HMO we study manages a network of independent, primary care providers—primarily internists, pediatricians and family medicine doctors.<sup>3</sup> The HMO contracts directly with individual providers, or with associations of such practices, to provide medical services. In its physician contracts, the HMO specifies financial rewards to doctors who meet spending targets. Although the terms of these contracts can change from one year to the next, and in fact did change during our period of analysis, in any given year all physicians in the network worked under identical incentive contracts.

### 2.1 Global Budgets and Group Incentives

For the purpose of our analysis, there were two important features of the HMO’s physician contracts.<sup>4</sup> First, doctors were given a global budget for patient expenditures and were free to allocate resources across patients as they wished. This budget was a function of the number of HMO members who identified the physician as their primary care provider as well as the age and gender of these patients.<sup>5</sup> Physicians were given substantial financial incentives for keeping annual expenditures below the budgeted amount. For the years 1991-1996, roughly 20 percent of a physician’s fees were “withheld” if the physician’s expenditures exceeded the budget target. In 1997, as a result of changes to Federal regulations, the proportion of such “at risk” fees was reduced to 10 percent.

The second key feature of the HMO’s incentive system was that the incentive contracts were group-based, i.e., bonuses were based on the performance of *panels* of primary care physicians rather than individual doctors. These panels of doctors (known as PODs) varied in size from 3 to 30 physicians. Physicians only received incentive payments if the panel’s

---

<sup>3</sup>Such networks, sometimes known as “independent provider” networks, comprise one of the largest segments of the managed care market, roughly 40 percent of total HMO enrollment in 1998 (Interstudy, 1999)

<sup>4</sup>A more detailed description of the HMO’s incentive contracts can be found in Gaynor, Rebitzer, and Taylor (2001)

<sup>5</sup>Physician budget targets included nearly all non-pharmacy and non mental-health related expenditures up to \$15,000 per patient per year.

collective expenditures for the year came in under budget. This “team based” incentive is somewhat surprising in the HMO setting we study; among the physicians in the PODs there are few of the interdependencies in production that are generally thought to motivate team compensation. Upon closer examination, however, group based incentives make economic sense for a reason rarely discussed in the economics literature—their potential to improve the efficiency of the health care rationing decisions made by the HMO’s physicians.

To understand the efficiency gains from group incentives, consider a risk neutral physician who derives utility from income ( $y$ ) and from the provision of medical services to patients ( $m$ ). It is natural to think about the doctor’s utility as decreasing in the extent to which  $m$  is less than some “ideal” level,  $\hat{m}$ . We write the utility of physician  $j$  treating  $i = 1, \dots, n$  patients as:

$$U_j = y_j - \sum_{i=1}^n v(m_{ij} - \hat{m}_{ij}), \quad (1)$$

with  $v(0) = 0$ , and  $v' < 0$  and  $v'' > 0$  when  $m_{ij} < \hat{m}_{ij}$  (so that  $U$  is increasing in both  $y$  and  $m$  when  $m_{ij} < \hat{m}_{ij}$ ).

We conceive of  $\hat{m}$  as representing the level of medical services—scaled in dollar expenditures—that the patient would want the physician to provide if the patient did not bear any of the costs of treatment. Patient preferences enter the physician’s utility function because norms of professional conduct require the physician to advocate for the patient’s interests. Thus in some cases  $\hat{m}$  may represent the acknowledged medical best practice for treatment of a patient with a specific illness. In other cases  $\hat{m}$  may be influenced by patient preferences that differ from the physician’s conception of best practice treatments. For example, a patient may insist on a medically unnecessary MRI to reduce anxiety about the presence of a brain tumor. A physician who decides not to provide the MRI would experience some disutility because of a failure to resolve the patient’s on-going anxiety.

Whatever the precise determinants of  $\hat{m}$ , patients would prefer that the physician provide  $\hat{m}$  to alternative levels of services, if those services were provided at no cost. Patients who

*pay* for medical services, though, would prefer a level of services somewhat lower than  $\hat{m}$ .<sup>6</sup> An important role for the HMO is to set in place resource allocation rules and incentives that accomplish the goal of inducing doctors to set medical services below  $\hat{m}_{ij}$  for each patient, but not by “too much.” This task is complicated to a considerable degree by the fact that the HMO administrators cannot know  $\hat{m}_{ij}$ ; rather they must rely on the physician’s superior information and as well as her professional judgment.

A simple policy that relies on physician judgment to allocate resources would be to award each physician a fixed sum of money, say  $M_j$ , to be allocated to that doctor’s patients (so that  $\sum_i m_{ij} = M_j$ ). A physician working under this budget cap would follow the rationing rule:  $v'(m_{ij} - \hat{m}_{ij}) = \lambda_j$ . This rule is desirable because it equates the marginal disutility from rationing care (relative to each patient’s “ideal”) across a physician’s patient population. Requiring each physician to work under such a “hard” budget cap, however, also creates allocative inefficiencies. A physician who, by random chance, has a population of patients with medical conditions demanding more resources will provide a lower level of care (relative to  $\hat{m}_{ij}$ ) than will a physician having an unusually healthy clientele. This type of cross-physician disparity in treatment is a potentially important problem; efficiency requires that the allocation of medical resources depends on an individual’s medical exigencies ( $\hat{m}_{ij}$ ) rather than random events associated with her physician’s patient population.

The HMO sought to implement an incentive system that realized the benefits of physician budget caps while reducing some of the inefficient random variation in treatment these caps could induce. Their solution was to impose a “soft” budget constraint over *groups* of physicians. Each panel of doctors was given a *target*—not an absolute maximum—expenditure level and the group was rewarded if they *collectively* met this target. This group incentive pooled expenditures across physician patient populations and therefore reduced the role of chance variation in the allocation of medical resources.<sup>7</sup> By implementing group-level in-

---

<sup>6</sup>Since  $\hat{m}$  is a patient’s unconditional optimum, a small reduction in  $m$  from this level has a first-order effect on medical care expenditures, but a second-order effect on the patient’s well-being.

<sup>7</sup>This feature of group incentives (i.e., that they can improve efficiency in the allocation of resources) is quite different from economic models of team compensation that emphasize interdependencies among team

centives, however, the HMO introduced an additional complication—the problem of moral hazard in teams. The role of moral hazard is easily illustrated in our context and forms the basis for the empirical study in Section 3.

## 2.2 Moral Hazard in Teams

Suppose that doctors do not have complete control over their expenditures so that actual expenditures by doctor  $j$  on patient  $i$  is  $x_{ij} = m_{ij} + \varepsilon_{ij}$  where  $\varepsilon_{ij}$  is a mean-zero random variable.<sup>8</sup> Suppose further that doctor  $j$  with  $n_j$  patients is placed into a panel of  $N$  doctors, and that they work under a group incentive such that a per-patient bonus  $B$  is awarded to each physician when the average level of actual expenditures (the mean of  $x_{ij}$ ) in the panel is less than a specified target  $X_T$ . This mean is a random variable, with a probability and cumulative distribution function,  $\phi$  and  $\Phi$ , respectively. We assume that doctors allocate the bonus in proportion to their share of patients (managers report that this is typically the case in the HMO we study), and assume further that doctors play Nash strategies. Then the equilibrium medical services provided by physician  $j$  solves:

$$\max_{\{m_{ij}\}} \left[ n_j B \Phi(X_T - \bar{m}) - \sum_{i=1}^{n_j} v(m_{ij} - \hat{m}_{ij}) \right],$$

where  $\bar{m}$  is the average level of medical expenditures for  $n$  patients treated by the group. The first-order conditions for this problem are:

$$-s_j B \phi(X_T - \bar{m}^*) - v'(m_{ij}^* - \hat{m}_{ij}) = 0, \quad i = 1, \dots, n, \quad (2)$$

where  $s_j = n_j / \sum_{j=1}^N n_j$  is the physician's share of patients within the panel.

---

members. One might also argue that pooling across physicians reduces the role of random chance in earning incentive payments and may thus be desirable for risk-averse physicians. This risk aversion story is better suited, however, for explaining why individual physician contracts might have low-powered incentives rather than for explaining the use of the high-powered group incentives we observed at the HMO.

<sup>8</sup>Realized outcomes  $x_{ij}$  may also diverge from intended spending because of uncertainty about what the right treatment is.

Notice that for a POD in which each physician has an equal share,  $v'(m_{ij}^* - \hat{m}_{ij})$  is equal across all patients; there is no cross-patient disparity in treatment. On the other hand, if  $s_j$  differs across patients, this gives rise to a different motivation for disparate treatment. Assuming the second order conditions are satisfied, we can differentiate (2) to provide the comparative static:

$$\frac{\partial m_{ij}^*}{\partial s_j} = \frac{B\phi(X_T - \bar{m}^*)}{\frac{s_j B}{n}\phi'(X_T - \bar{m}^*) - v''(m_{ij}^* - \hat{m}_{ij})} < 0. \quad (3)$$

Given a panel with  $n$  patients, physicians who have a small share of the panel's patients will provide higher levels of medical services than physicians with larger shares. Similarly, per patient expenditures will be higher in large panels (where physicians generally have relatively low shares) than in small panels.

The designers of the system had hoped that peer pressure in PODs would offset the free-riding problem.<sup>9</sup> Opinions differed at the HMO about the actual importance of mutual monitoring, but our findings in Section 3 suggest that peer pressure was not sufficient to overcome the problem of moral hazard in teams.

### 3 Empirical Evidence

In this section we analyze the effect of the HMO's incentive system on medical expenditures. Specifically, we examine the correlation between spending on medical utilization and the intensity of incentives to limit these expenditures.

#### 3.1 Why Did Physicians' Shares of Members and Panel Size Vary?

Following the theory, we use variation in a physician's share of her POD's patients and in the number of physicians in a panel to identify variation in incentive intensity. As a prelude to this analysis we briefly discuss why panel size and physician patient share varied in the

---

<sup>9</sup>Kandel and Lazear (1992) and Encinosa, Gaynor, and Rebitzer (1999) discuss group incentives and peer pressure in professional and medical settings.

HMO network.

In their written material, the HMO emphasized that doctors themselves decided which panel they wished to join, but the actual construction of the panels seems to have been quite haphazard. No one we spoke with could reconstruct the process by which physicians ended up in the panel they did, and no records were kept regarding the decision process. In some instances, it appears that the HMO arbitrarily assigned physicians to panels (typically to increase the number of enrollees in the panel), but little attention was given to the possibility that the intensity of incentives might be diluted as the number of physicians (as opposed to enrollees) in the panel increased. In other cases, physicians joined panels because they were in the same practice, were located close to or had other social or professional relationships with other panel members. As a result of the way these panels were constructed, otherwise similar primary care physicians were grouped into quite different panels. Panel sizes ranged from 3 to 30. Some panels include pre-existing group practices, while others did not.<sup>10</sup> In some panels all physicians had the same specialty, while others were a mix of different primary care specialties. Panel sizes did vary from one year to the next as new physicians entered or left panels, and we use this variation to estimate fixed effects models. As we show in the following empirical analysis, however, there does not appear to be a significant relationship between a physician's patient expenditures and the decision to change panels. One final institutional fact about the HMO is relevant to our empirical analysis: the HMO did not and could not steer patients towards low-cost physicians or panels in its network. Once a patient signed up with the HMO, she was free to use any physician in the network.<sup>11</sup>

The data we use comes from the HMO's internal records and data are available at the physician level and the panel level. We begin with the physician-level analysis because of its close relationship to our discussion in Section 2, and then turn to the panel-level analysis. This panel-level analysis has a number of advantages. Of these the most important is that

---

<sup>10</sup>Only 11 percent of the panels in our sample were composed of a single entity for the purposes of receiving payment from the HMO.

<sup>11</sup>Also, as a matter of policy and practice, the HMO did not try to force high-cost physicians or panels out of its network.

the HMO kept records on a number of key variables at the POD level but not at the physician level.

### 3.2 Physician-Level Analysis

Our basic specification is linear:

$$x_{ij} = \beta_0 + \beta_1 s_j + \beta_2 Z_{ij} + u_{ij}, \quad (4)$$

where  $x_{ij}$  is the observed expenditure by doctor  $j$  on patient  $i$  (recorded on a per-member per-month basis),  $s_j$  is the share of physician  $j$  in his or her panel, and  $Z_{ij}$  are other available patient characteristics. If incentives affect physician behavior,  $\beta_1$  will be negative.

The theoretical model in the previous section supplies a useful interpretation of the linear specification (4). If the component of physician utility that derives from providing medical services takes the form  $v(m_{ij} - \hat{m}) = \alpha(m_{ij} - \hat{m}_{ij})^2$ , the first-order condition (2) for the physician can be written

$$m_{ij}^* = \frac{-B\phi}{2\alpha} s_j + \hat{m}_{ij}. \quad (5)$$

While  $m_{ij}^*$  is not observed, the HMO does record realized expenditures,  $x_{ij}$ , which we have assumed are related, by  $x_{ij} = m_{ij}^* + \varepsilon_{ij}$ . If one assumes that the unobservable term,  $\hat{m}_{ij}$ , is linearly related to observed characteristics  $Z_{ij}$  plus an error term that is orthogonal to  $s_j$ , this gives equation (4). In this specification  $\beta_1$  is  $\frac{-B\phi}{2\alpha}$ . Both  $B$  and  $\phi$  are positive. The parameter  $\alpha > 0$  determines the relative importance of income and deviations from “ideal” medical expenditures for physician utility. If physicians are insensitive to the monetary incentive,  $\alpha$  will be very large, and  $\beta_1$  will be close to zero; there will be little correlation between the physician’s share in her panel and expenditures on patients.

To our linear specification, we also add a variable interacting physician share  $s_j$  and an indicator variable for the year 1997. Estimates are presented in Table 1. Consistent with

the hypothesis that doctors respond to incentives, the coefficient on physician share ( $\hat{\beta}_1$ ) is negative and statistically significant at conventional levels. If a physician has a patient share near 0 in the panel, incentives will be irrelevant and medical expenditures will approximately equal  $\hat{m}_{ij}$ .<sup>12</sup> If the doctor has a share equal to the sample average of 0.1,  $(m_{ij}^* - \hat{m}_{ij})$  is estimated to be approximately \$4.00 per member per month, a savings of five percent of the average medical expenditures per-member per-month of \$81. If, alternatively, the physician has a share of 0.2, medical expenditures are estimated to decline by approximately \$8.00. Physicians are clearly influenced by the financial incentives.<sup>13</sup>

The savings implied by this estimate of physician responsiveness are substantial. The doctors in our sample served just over 800,000 HMO member months in 1996. If all physicians had 0.1 shares of their POD's enrollees (the sample average), the total reduction in medical expenditures due to incentives was approximately \$3.2 million. If the HMO had placed doctors in smaller PODs, so that patient share was 0.2, expenditures would, according to our estimates, decline another \$3.2 million.

In 1997, changes in Federal regulations forced the HMO to reduce the bonus for meeting cost targets from 20 percent to 10 percent. We capture the effect of this change by interacting the *Physician Share* variable with a dummy variable for 1997. The coefficient on this variable has the “wrong” sign but is measured imprecisely. We discuss the interpretation of this result in greater length in the next section.

---

<sup>12</sup>Of course, there are no doctors in our sample with a share of 0, but there are a number of doctors with very small shares (below 0.01).

<sup>13</sup>In an independent practice HMO like the one we study, physicians treat patients from a number of insurance companies. It is possible that a doctor's response to the HMO's incentives may depend on the proportion of their patients drawn from the HMO. Although we know the number of HMO patients each physician treats, we unfortunately do not know their total number of patients. Attempts to use the total number of HMO patients as a proxy for the fraction of HMO patients in a physician's practice failed because this variable is also the numerator of our key variable of interest, *Physician's Share of POD Member Months*. These estimates are available from the authors on request.

### 3.3 Panel-Level Analysis

The results in Table 1 suggest that physician incentive contracts do have the effect of reducing medical utilization expenditures. In this section we use the more extensive panel-level data to estimate the effect of incentives. The richer data allow us to introduce a broader set of covariates into our regressions. In addition, the information available at the panel-level enables us to investigate which types of utilization costs are most influenced by incentives.

Our basic regression is

$$x_{jt} = \delta_0 + \delta_1 \ln(N_{jt}) + \delta_2 \ln(N_{jt}) * D_{1997} + \delta_3 Z_{ijt} + u_{it}, \quad (6)$$

where  $x_{jt}$  is the observed medical expenditures per member per month in POD  $j$  at time  $t$ ,  $\ln(N_{jt})$  is the log of the number of physicians in POD  $j$  in year  $t$  and  $D_{1997}$  is a dummy variable equal to 1 for the year 1997.<sup>14</sup>  $Z$  is a vector of variables that capture the influence of other factors that potentially influence medical expenditures.

If incentives influence physician behavior in the expected manner then  $\delta_1 > 0$ . Because of the Federally mandated reduction of incentive intensity in 1997, we also anticipate  $\delta_3 < 0$ .

Column (1) of Table 2 presents estimates of equation (6) for the years 1994-1997.<sup>15</sup> We observe that PODs with more physicians have higher medical expenditures. This coefficient is economically as well as statistically significant. An increase in POD size from 10 to 12 physicians is associated with an increase in expenditures of \$6.58 per member per month. This represents an increase of 7.3 percent over the sample mean. Consistent with the reduction of incentive intensity in 1997, we also observe a negative coefficient on the POD Size $\times$ 1997 interaction. Point estimates suggest that in 1997, an increase in POD size from 10 to 12 physicians would lead to an increase in costs of only \$1.18 per member per month.

---

<sup>14</sup>Theory suggests that the relationship between POD size and medical expenditures will be non-linear. We use the log specification because it offers a convenient interpretation of the coefficients. We find similar results with an alternative specification that uses dummy variables to represent low, medium, and large POD size categories.

<sup>15</sup>Our POD level estimates are from a time-series, cross-section panel, and we have adjusted standard errors to account for possible within-POD correlation of errors and cross-panel heteroscedasticity.

This finding is consistent with the reduction in incentive intensity that occurred in 1997.

One of the “control variables” included in vector  $Z$  is *Target Expenses*. This target is the forecast of a POD’s “expected expenditures” based on enrollee age and gender. The estimated coefficient of 0.99 on *Target Expenses* indicates an approximately 1 to 1 relationship between expected and actual POD medical expenditures. The equation in column (1) also includes the variable *Years that POD is in Sample Since 1994*. This variable is potentially important because during the time of this study, the HMO was expanding from an urban setting into suburban areas. The positive coefficient on *Years that POD is in Sample Since 1994* may indicate that the older, more urban, panels were likely treating more expensive populations of enrollees. Finally the estimates in column (1) include year fixed effects.<sup>16</sup>

The equation estimated in column (2) of Table 2 introduces additional control variables to capture heterogeneity across PODs and their enrollee populations. These are described in detail in Data Appendix 1 in Gaynor, Rebitzer, and Taylor (2001), but are suppressed in Table 2 to facilitate exposition. The addition of these “control” variables had no effect on coefficient signs and statistical significance and little effect on the magnitudes of the coefficients on our key variables of interest, *Ln Physicians in POD* and *Ln Physicians in POD*  $\times$  1997. These results suggest that our findings are not likely due to correlation between the number of physicians in a POD and other observed characteristics of the panel, its physicians or its patient population.<sup>17</sup>

An alternative approach to controlling for unobserved POD heterogeneity is to estimate our equations with POD-level fixed effect variables. The fixed effect estimates are presented in column (3) of Table 2. The coefficients on the key variables, *Ln Physicians in POD* and *Ln*

---

<sup>16</sup>A subtle feature of the HMO’s incentive system was that physicians were not held responsible for expenditures above \$15,000 per patient in a year. See Gaynor, Rebitzer and Taylor (2001) for estimates using a variety of different dependent variables.

<sup>17</sup>In the POD specifications and in the theoretical framework, it makes no difference if all the  $N$  physicians in a panel have  $1/N$  share or if certain physicians have a disproportionate share of the HMO business. We tested this hypothesis by re-estimating the equations in columns 1 and 2 of Table 2 with a measure of concentration of the POD’s HMO enrollees. This measure, the sum of the square of each physician’s share of HMO membership, did not have statistically significant relationship to *Total Medical Expenditures* and adding this variable did not qualitatively affect the results presented in Table 2.

*Physicians in POD* × 1997 are not qualitatively changed from those in column (1) although the magnitude of the effect of number of physicians on costs increases somewhat in absolute value. Thus, in our fixed effect estimates, a POD that increases the number of physicians from 10 to 12 will, for the years 1994-1996, experience an increase in *Total Medical Expenditures* of \$8.28 per-member per-month, falling to \$3.29 per-member per-month in 1997.

We interpret the fixed effect coefficients on *Ln Physicians in POD* and *Ln Physicians in POD* × 1997 as reflecting changed behaviors due to the changing degree of incentive intensity. An alternative possibility is that selection, rather than behavioral changes, is driving our fixed effect results. It may be, for example, that *Ln Physicians in POD* has a positive correlation with *Total Medical Expenditures* because panels that reduce the number of physicians do so by kicking out their most expensive physicians. Alternatively, it may be that panels with growing numbers of physicians are hiring new members who tend to be more expensive than incumbents.<sup>18</sup>

Table 3 offers an investigation of the importance of selection effects. It relies on individual cost data for the HMO's primary care physicians for the years 1994-1996. In Panel A we use these data to see if those leaving shrinking groups have patient expenditures that are, on average, higher than others in their panel. We observe that the 298 physicians who exited a panel at the end of 1994 or 1995 had expenditures that were, on average, 2 percent higher than other physicians in the panel they are leaving. This difference between movers and stayers is both small in magnitude and statistically insignificant.<sup>19</sup> In Panel B of Table 3, we compare total medical expenditures between incumbent physicians and new entrants in

---

<sup>18</sup>Selection may also help account for the other results in Tables 1 and 2 if physicians with preferences for less cost-conscious practice styles choose large PODs because incentive pressures are reduced. Patients with preferences for more costly care may then select physicians with reputations for more expensive practice styles. It is unlikely however that these patients would select physicians on the basis of POD size itself because the size and composition of PODs was typically known only by physicians and administrators—not patients. Since PODs were not generally composed of organizations visible to patients such as a practice it would be hard for a patient to infer the size of a POD from public information.

<sup>19</sup>To put the Panel A figures in perspective, remember that the point estimates in column (1) of Table 2 suggest that a panel decreasing from 12 to 10 members will experience a 7 percent reduction in Total Medical Expenditures. If the leavers in such a panel have expenditures that are, on average, 2 percent higher than the stayers, and if the only effect from reductions in physicians in the panel were due to selection, the reduction in expenditures would be only 0.4 percent.

the panel. Of the 746 new entrants in the years 1995 or 1996, we observe that entrants have expenditures that are, on average, 1 percent higher than those in the panel they are entering. This difference is far too small to account for the results in Table 2. It is also statistically insignificant.

### 3.4 Types of Expenditures

In the HMO's system, primary care providers were responsible for all medical expenditures up to \$15,000 per patient per year, even if these fees were due to charges from specialists or other providers. Under the rules prevailing prior to 1997, primary care physicians got a minimum of 80 percent of the fees for services they provided and zero percent of the fees other doctors charged. In addition, the system offered essentially no cost-cutting incentives to doctors who were not primary care physicians. Clearly the cost-cutting incentives were most powerful for services *not* provided by the primary care physician herself.<sup>20</sup> This feature of the system makes possible more refined tests of the impacts of incentives. We present these in Table 4.

Data on the sub-components of spending are only available for 1994–1996. Thus we begin our analysis in Table 4 by re-estimating our base model (column (1) of Table 2) for these years. The estimated coefficients *Ln Physicians in POD*, are very close to those estimated over the full sample of years (1994–1997).

Column (2) of Table 4 presents estimates of the determinants of *Expenditures Not Paid to Self*. This variable measures medical costs of procedures and treatments that do not earn revenues for the primary care physician. We find that the coefficient on *Ln Physicians in POD* is positive, statistically significant, and similar in magnitude to the coefficient in column (1). In contrast, the dependent variable in column (3) measures expenditures that do earn income for the primary care physician. The coefficient on *Ln Physicians in POD* in column

---

<sup>20</sup>The primary care physicians were very aware of this aspect of the incentive system. Reducing specialist referrals was considered an especially desirable and effective way to cut costs and earn incentive rewards. Indeed we attended meetings in which physicians compared notes on how to reduce such referrals without compromising patient care.

3 is not statistically significant. Taken together, the results in columns (2) and (3) suggest that in PODs with fewer physicians (and consequently higher powered incentives), physicians cut expenditures on services and procedures that generate fees for other physicians, but not for themselves.

Columns (4) and (5) of Table 4 compare the determinants of expenditures due to inpatient and outpatient services and procedures. The dependent variable in column (4), *Total Inpatient*, tracks medical expenditures due to procedures conducted in a hospital on an inpatient basis. We observe that the coefficient on *Ln Physicians in POD* in column (4) is small and not statistically different from zero. Column (5) estimates the determinants of outpatient expenses. The variable *Total Outpatient* includes costs incurred in a hospital (on an outpatient basis), a clinic, or a doctor's office. In column (5) of Table 4, we find that *Ln Physicians in POD* has a positive and statistically significant effect on outpatient spending. The magnitude of this effect is substantial. Increasing panel size from 10 to 12 physicians increases outpatient spending by \$1.71 per member per month or roughly 5 percent above the mean. Taken together, the results in columns (4) and (5) suggest that the HMO's incentives do not have much influence on inpatient expenses, but are important for outpatient expenses. This finding makes sense. Patients admitted to hospitals are likely to be sicker than other patients and are also more likely to be cared for by specialists rather than primary care physicians. In contrast, outpatient procedures are more likely to be elective and more likely to be controlled by the primary care physician. For both these reasons, the incentive system should work more powerfully on outpatient expenditures.

## 4 Conclusion

In this paper we examine the structure and performance of physician cost-containment incentives in an HMO network. Our results have implications for the management of these important organizations as well as for the general economic theory of incentive design.

From an economist's point of view, the most striking feature of the contracts we study

is the use of high-powered group-based incentives. Conventional wisdom holds that group incentives make most sense when production entails significant interdependencies among group members. These interdependencies, however, are largely absent in the organization we study. Similarly, while the reduction in income variation inherent in group-based incentives may benefit risk averse physicians, risk aversion would seem most likely to lead to use of low-powered incentives targeted at the individual level, rather than high-powered group incentives. Our investigation of the anomalous use of team-based incentives led us to a little-noticed characteristic of these systems—that they can enhance the efficiency of resource allocation decisions. The central economic logic of these gains from team incentives is a generally applicable point; these efficiencies can appear in any setting where the principal relies on the professional judgment of multiple agents to allocate resources.

Group incentives may enhance the efficiency of resource allocation decisions but they also introduce the well-known problem of moral hazard. The designers of the HMO's system understood the problem of moral hazard in teams, but hoped that peer pressure within physician panels would overcome the dilution of incentives that naturally occurs in group settings. Our finding that expenditures were substantially reduced in small physician panels suggests that these effects were not sufficient to eliminate all “free riding.” This finding is significant because many team based incentives rely to some extent on the power of peer pressure to overcome moral hazard. In subsequent research we hope to study more directly the role that peer pressure played in this system.

Beyond the issue of group incentives, other features of the HMO's incentive system are hard to defend on economic grounds. While the use of a bonus based on a target had the advantage of transparency, the payout was highly non-linear around target expenditures, likely introducing an unnecessary fragility into the bonus system. Just as the strike price is crucial to the incentive effect of CEO stock options, so the power of the HMO's incentive system was critically dependent on not setting the target budgets “too low” or “too high.”

At the managerial level, our work clearly demonstrates that the incentive systems HMOs

introduce can help contain medical costs (though we have not established that the savings thereby created exceed the cost of implementing these systems). An equally important unresolved issue is whether cost containment incentives compromise care quality. We plan to take up these issues in subsequent papers.

## References

- Coase, Ronald H. "The Nature of the Firm." *Economica* 52 (November 1937): 386–405.
- Encinosa, William; Gaynor, Martin; and Rebitzer, James B. "The Sociology of Groups and the Economics of Incentives: Theory and Evidence On Incentives in Medical Groups." Manuscript. Pittsburgh: Carnegie Mellon University, 1999.
- Gabel, Jon; Levitt, Larry; Pickreign, Jeremy; Whitmore, Heidi; Holve, Erin; Hawkins, Samantha; and Miller, Nick. "Job-Based Health Insurance in 2000: Premiums Rise Sharply While Coverage Grows." *Health Affairs* 19 (September/October 2000): 144–151.
- Gaynor, Martin; Rebitzer, James B.; and Taylor, Lowell J. "Incentives in HMOs." Working Paper 8522. Cambridge, MA: National Bureau of Economic Research, 2001.
- Gibbons, Robert. "Incentives and Careers in Organizations." in David M. Kreps and Kenneth Wallis, eds., *Advances in Economics and Econometrics, Theory and Applications: Seventh World Congress*. Volume 2. Econometric Society Monographs, no. 27. Cambridge, New York and Melbourne: Cambridge University Press, 1997.
- Ichniowski, Casey, and Shaw, Kathryn. "Beyond Incentive Pay: Insiders' Estimates of the Value of Complementary Human Resource Management Practices." *Journal of Economic Perspectives* 17 (Winter 2003): 45–58.
- InterStudy. *The InterStudy Competitive Edge, Part II: The HMO Industry Report 9.1*. St. Paul, MN: Interstudy Publications, 1999.
- Kandel, Eugene, and Lazear, Edward P. "Peer Pressure and Partnerships." *Journal of Political Economy* 100 (August 1992): 801–817.
- Lazear, Edward P. "Performance, Pay and Productivity." *American Economic Review* 90 (December 2000): 1346–61.

Prendergast, Canice. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (March 1999): 7–63.

Table 1: Determinants of Physician-Level Medical Expenditures<sup>a</sup>

Variable	Parameter Estimate <sup>b</sup>
Physician's Share of POD Member-months	-40.46 (-2.00)
Physician's Share $\times$ 1997	-138.40 (-1.56)
Constant Term	Yes
Physician Specialty Fixed Effects <sup>c</sup>	Yes
Year Fixed Effects	Yes
Number of Observations <sup>d</sup>	2,282
Number of Physicians in the Sample	961
R <sup>2</sup>	0.02

<sup>a</sup> Measured as annual total medical expenditures per-member per-month for an individual physician.

<sup>b</sup> Numbers in parentheses are t-statistics calculated to allow for clustering by physician.

<sup>c</sup> Physician specialties are internal medicine, family practice, and pediatrics.

<sup>d</sup> This is an unbalanced panel due to changes in the size of the HMO's physician network during the period of this study. 203 physicians were in the sample for all 4 years, 254 for 3 years, 204 for 2 years, and 300 for 1 year.

Table 2: POD-Level Expenditure Equations<sup>a</sup>

Variable	Parameter Estimates <sup>b</sup>		
	OLS (1)	OLS (2)	Fixed Effects (3)
Ln Physicians in POD	36.17 (2.96)	27.35 (2.29)	46.10 (3.02)
Ln Physicians in POD × 1997	-29.67 (2.51)	-26.88 (2.13)	-27.86 (1.93)
Target Expenses	0.99 (6.94)	0.76 (3.63)	0.60 (5.83)
Years POD is in Sample Since 1994	9.11 (2.43)	6.44 (1.41)	28.61 (0.98)
Year is 1995	-14.52 (1.65)	-12.20 (1.36)	-30.87 (1.02)
Year is 1996	-51.83 (4.48)	-47.08 (3.75)	-88.93 (1.52)
Year is 1997	-8.91 (0.24)	-17.18 (0.44)	-71.92 (0.95)
Constant Term	Yes	Yes	Yes
Additional POD-Level Variables	Yes <sup>c</sup>	Yes <sup>d</sup>	Yes <sup>c</sup>
POD Fixed Effects <sup>c</sup>	No	No	Yes
Number of Observations <sup>d</sup>	353	340	353
Number of PODs	118	109	118
R <sup>2</sup>	0.46	0.45	-0.06

<sup>a</sup> The dependent variable is annual dollars of total medical expenditures per-member per-month for the panel of physicians.

<sup>b</sup> Numbers in parentheses are t-statistics calculated to allow for clustering by POD.

<sup>c</sup> Variables include the natural log of POD enrollees and its interaction with the dummy variable for 1997.

<sup>d</sup> Variables include those in note (c) above, plus POD network affiliation and specialty mix and percent members enrolled in Medicaid in 1997. Descriptions of these variables are in the data appendix in Gaynor, Rebitzer, and Taylor (2001).

Table 3: Selection vs. Incentives<sup>a,b</sup>

Panel A <sup>c</sup>				
Physician Exits POD After Year t	Size of POD Declines from Year t to Year t+1			
		No	Yes	All
Yes	Relative Costs Year t	1.02	1.02	1.02
	Number Physicians in Cell	257	41	298
No	Relative Costs Year t	0.99	1.00	1.00
	Number Physicians in Cell	678	338	1016
All	Relative Costs Year t	1.00	1.00	1.00
	Number of Physicians in Cell	935	379	1314

  

Panel B <sup>d</sup>				
Physician Enters POD in Year t	Size of POD Increases from Year t-1 to Year t			
		No	Yes	All
No	Relative Costs Year t	1.00	0.98	0.99
	Number Physicians in Cell	678	338	1016
Yes	Relative Costs Year t	1.06	1.01	1.01
	Number Physicians in Cell	45	701	746
All	Relative Costs Year t	1.00	1.00	1.00
	Number of Physicians in Cell	723	1039	1762

<sup>a</sup> In Panel A, t refers to years 1994 and 1995. In Panel B, t refers to years 1995 and 1996. We restrict our analysis to 1994-96 because a change in the HMO's accounting system made it difficult to track those who moved across PODs between 1996 and 1997.

<sup>b</sup> Relative Costs is the physician's total medical expenses per member per month in year t relative to the panel's average medical expenses per-member per-month in year t. None of the cross-cell differences in the table is statistically significant.

<sup>c</sup> In Panel A, 298 of the physicians in the sample exited after 1994 or 1995. The relative costs of these exiters were 2% above the mean for both shrinking and growing PODs.

<sup>d</sup> In Panel B, 746 of the physicians in the sample entered in 1995 or 1996. Their relative costs were 1% above the means of their PODs. This 1% differential also held for growing PODs.

Table 4: Medical Utilization Expenditures by Category<sup>a</sup>

Variable	Total Medical Expenditures (1)	Expenditures Not Paid to Self (2)	Expenditures Paid to Self (3)	Total Inpatient (4)	Total Outpatient (5)
Ln Physicians in POD	26.23 (2.11)	24.06 (1.93)	2.17 (1.36)	5.06 (0.67)	9.51 (2.02)
Target Expenses	0.63 (2.16)	0.55 (2.33)	0.08 (1.27)	0.22 (3.33)	0.31 (1.85)
Years POD is in Sample Since 1994	11.32 (1.25)	9.64 (1.14)	1.68 (1.42)	5.16 (1.36)	5.58 (1.16)
Year is 1995	-16.52 (1.51)	-15.00 (1.41)	-1.53 (1.36)	-14.62 (2.35)	-0.91 (0.20)
Year is 1996	-54.10 (2.92)	-50.96 (2.91)	-3.15 (1.26)	-30.49 (4.44)	-9.79 (1.03)
Constant Term	Yes	Yes	Yes	Yes	Yes
Additional POD- Level Variables <sup>c</sup>	Yes	Yes	Yes	Yes	Yes
Number of Observations	236	236	236	236	236
Number of PODs	99	99	99	99	99
R <sup>2</sup>	0.42	0.41	0.39	0.18	0.47

<sup>a</sup> All estimates are OLS using data from 1994–1997. All dollar measures are per-member per-month.

<sup>b</sup> Numbers in parentheses are absolute values of t-statistics calculated to allow for clustering by POD.

<sup>c</sup> Variables are: Ln POD enrollees, POD network affiliation and specialty mix and percent members enrolled in Medicaid in 1997. Descriptions of these variables are in the data appendix in Gaynor, Rebitzer, and Taylor (2001).