

CMPO Working Paper Series No. 03/089

Incentives in HMO's*

*Health Maintenance Organisations

Martin Gaynor

Carnegie Mellon University, National Bureau of Economic Research and
CMPO Associate Member

James B Rebitzer

Case Western Reserve University, National Bureau of Economic Research
and the Jerome Levy Institute

And

Lowell J Taylor

Carnegie Mellon University

June 2002

Abstract

We study the effect of physician incentives in an HMO network. Physician incentives are controversial because they may induce doctors to make treatment decisions that differ from those they would choose in absence of incentives. We set out a theoretical framework for assessing the degree to which incentive contracts do in fact induce physicians to deviate from a standard guided only by patient interests and professional medical judgement. Our empirical evaluation of the model relies on details of the HMO's incentive contracts and access to the firm's internal expenditure records. We estimate that the HMO's incentive contract provides a typical physician an increase, at the margin of \$0.10 in income for each \$1.00 reduction in medical utilisation expenditures. The average response is a 5% reduction in medical expenditures. We also find suggestive evidence that financial incentives linked to commonly used "quality" measure may stimulate an improvement in measured quality.

Address for Correspondence

National Bureau of Economic Research; Heinz
School, Carnegie Mellon University, Pittsburgh,
PA 15213-3890, USA
mgaynor@cmu.edu

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 within organizations. Using data derived from the internal records of firms, economists have documented how explicit and implicit incentives shape organizational structure and performance.¹ Our research contributes to this new empirical literature by studying incentives in a setting that has received relatively little attention from economists, but which has important implications for public policy: physician incentives in managed care organizations.²

Americans spend more than one trillion dollars annually on health care. Because physicians play a central role in the allocation of these vast resources, policies and practices that influence physician decision making have been the subject of high stakes litigation and intense public controversy. Much of this attention has focused on managed care organizations, because managed care organizations are the dominant form of health insurance in the United States (with 92 percent of all individuals possessing employer-provided health insurance enrolled in some form of managed care, Gabel et al., 2000). A key component of every managed care organization is a system of financial and non-financial incentives that reward physicians for limiting medical expenditures. These incentives are controversial to the extent they induce physicians to take actions on behalf of patients that differ from those

¹ Examples include: Baker, Gibbs and Holmstrom (1994) on the internal labor market of a large corporation; Lazear (2000) on piece rates in Safelite Glass; Ichniowski, Shaw and Prenzushi (1997) on human resource practices in steel mills; Landers, Rebitzer and Taylor (1996) on career incentives and work hours in law firms; Gaynor and Gertler (1995) on incentives and risk spreading in medical practices. Prendergast (1999) and Gibbons (1997) offer comprehensive reviews of this literature.

² See the surveys by Dudley *et al.* (1998) and Gold (1999) on physician incentives in managed care specifically, and Newhouse (1996) on reimbursement incentives in health care generally.

they would take in the absence of incentives.

This paper studies physician responses to cost and quality incentives in a managed care setting.³ The HMO we study operates a network of independent, primary care providers (PCPs). In this type of network, HMOs contract directly with physicians in independent practices (or with associations of such practices) to provide medical services.⁴

Our analysis exploits two noteworthy features of the HMO's incentive system. The first feature is that every primary care physician in the network was assigned the role of a "gatekeeper" who regulates access to medical resources. Thus the HMO required that each enrollee (as HMO members were called) select a PCP, and the chosen PCP was then responsible for all the medical expenditures incurred by the enrollee. The contract specified financial rewards for PCPs with average medical utilization expenditures below a target level. The second key feature of the HMO's incentive system is that the incentive contracts were group-based, i.e., they rewarded 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. As we discuss in more detail below, these PODs were haphazardly assembled by the HMO with little attention given to the possibility that group incentives are likely to be more powerful the smaller the size of the POD.

The approach we take to understanding the HMO's incentive system integrates a theoretical model with data provided by the HMO. Our theory offers a framework

³ Firms in this industry adopt a variety of organizational forms, such as health maintenance organizations (HMOs), preferred provider organizations (PPOs), and point-of-service plans (POSs). All of these plans seek to limit medical expenditures by selectively contracting with health care providers, and most also attempt to control care via financial incentives for doctors. See Cutler, McClellan, and Newhouse (2000) for recent work decomposing the savings achieved by managed care plans between price and quantity reductions.

⁴ These types of networks are referred to as "Independent Provider," or "IP" networks. These rarely studied entities comprise one of the largest segments of the managed care market, roughly 40 percent of total HMO enrollment in 1998 (InterStudy, 1999).

for assessing the degree to which incentive contracts induce physicians to deviate from expenditures they would have made if guided only by patient interests and professional medical judgements. We use this framework to analyze the spending patterns in the HMO's network of primary care physicians. Our empirical strategy relies on variation in the size of physician panels and variation over time in the contractual arrangements to identify cross-sectional and time-series variation in incentive intensity. We study the effect of incentives by examining correlations between average medical expenditures and incentive intensity.

The paper is organized as follows. In the next section, we describe in more detail the relevant institutional features of the HMO's incentive system. In Section 3 we set out a model of how physicians might behave in such a system. Our model clarifies the benefits that group incentives offer in the HMO context. Indeed, we find that the use of group incentives can lead to a more efficient allocation of resources among the HMO's patients than incentives geared towards individual physicians.

Section 4 contains empirical results. The data we use were taken from the records the HMO used to administer its incentive system. These data track the medical expenditures for each enrollee and attributes these expenditures to the enrollee's PCP. Using these data we are able to characterize how incentives influence marginal expenditures.

In the last year of our study the HMO introduced financial incentives for PCP panels that met certain "quality" targets. We turn, in Section 5, to an analysis of the effect of these quality incentives. The final section of the paper contains a summary and conclusions.

2. The Incentive System

As noted above, the HMO managed a network of independent, primary care providers. The HMO contracted directly with individual providers (or with associa-

tions of such practices) to provide medical services. The HMO provided contractually specified financial rewards to PCPs who met spending targets and, in the final year of our study, to those who also scored well on certain quality indicators. Although the terms of these contracts could and did change from one year to the next, all PCPs in the network in a given year worked under identical incentive contracts.

2.1. The System from 1991 through 1996

For the years 1991 through 1996, the HMO used the same incentive contract for all its primary care physicians. Under the terms of the contract, the HMO agreed to reimburse PCPs for services they provided at 125 percent of Medicare's fee schedule.⁵ Only 80 percent of these fees, however, were paid immediately to the physician. The remaining 20 percent were "withheld" and paid out at the end of the year to physicians whose PODs had average medical expenditures below target levels. In addition to getting their "withhold" returned, panels coming in under target were eligible for a "gain-sharing" bonus which returned to the POD one-half of the difference between actual and targeted utilization expenditures. Thus, physicians in PODs with average costs below target could expect a substantial "bonus" of no less than twenty percent of their total fees.

Spending targets were constructed as follows. Each enrollee was assigned an actuarially determined monthly budget based on the enrollee's age and gender, the specialty of the primary care physician, and forecasts of changes in the demand for and price of medical services.⁶ The target for the POD was constructed by summing up the budgeted expenditures for all enrollees for whom the POD's physicians were responsible. All health care utilization expenditures were included in the spending target except expenditures for dental care, pharmacy, mental health and substance

⁵ This fee schedule is known as the resource based relative value scale (RBRVS).

⁶ Primary care specialties include internists, family practice, pediatrics and obstetrics/gynecology.

abuse services, neonatal intensive care, and any expenses above \$15,000 accrued by an individual enrollee in a year. This last feature, known as a “stop-loss” provision, was important as it removed incentives to withhold care from the sickest (or most expensive) patients.

A critical feature of the HMO’s incentive system was the construction of the PODs. In their written material, the HMO emphasized that doctors decided which PODs they wish 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 PODs 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 POD), but little attention was given to the possibility that the intensity of incentives might be diluted as the number of physicians in the POD increased. In other cases, physicians joined panels because they were in the same practice or because they were located close to or had other social/professional relationships with other panel members. As a result of the way these panels were constructed, otherwise similar PCPs were grouped into quite different PODs. Panel sizes ranged from 3 to 30. Some PODs include pre-existing group practices, while others did not.⁷ In some PODs all physicians had the same specialty, while others were a mix of different primary care specialties.

An obvious concern in evaluating the behavioral impacts of the HMO’s group incentives is the selection of physicians into PODs. We know that physicians did move across panels, but as we show in the following empirical analysis, there does not appear to be a significant relationship between a physician’s patient expenditures

⁷ The HMO did not require that panels be composed of corporate entities or partnerships, but such entities were sometimes included in the construction of a POD. Only 11 percent of the PODs in our sample were composed of a single entity for the purposes of receiving payment from the HMO.

and the decision to change panels. We also note that the HMO did not and could not steer patients towards low-cost physicians or PODs in its network. Once a patient signed up with the HMO, she was free to use any physician in the network. Similarly, the HMO did not engage in “economic credentialing.” This means that as matter of policy and practice, the HMO did not try to force high cost PCPs out of its network. Once a PCP agreed to join the network and once the PCP passed the network’s screening process, continued participation in the network was not influenced by the level of expenditures chosen on behalf of the PCP’s enrollees.

Another issue in assessing the impact of incentives is whether the PCP “gatekeeper” could plausibly exert a substantial influence on medical expenditures. After all, the fees accruing to PCPs for the services they deliver account for a relatively small fraction of total medical utilization costs.⁸ This issue was especially important in the HMO we study because once an enrollee was referred by the PCP to a specialist, the specialist was free to order further tests and procedures without the approval of the enrollee’s PCP.

In informal conversations we learned that at least some PCPs in the network believed they could exert a substantial influence on medical expenditures. Three strategies stood out in these discussions. The first was to keep patients with congestive heart failure, diabetes, asthma or other chronic diseases out of the hospital by teaching patients how to better manage these diseases themselves. The second was to keep patients from using hospital emergency rooms by having extended office hours and very good message handling systems. The third strategy was to reduce specialist referrals by learning more about which types of cases needed to be referred and which could be safely handled by PCPs themselves. One physician, for example, claimed his POD dramatically reduced expensive referrals to dermatol-

⁸ In the HMO we study, PCP fees were approximately 20 percent of total utilization costs.

ogists when the PCPs underwent training to spot which conditions needed to be referred and which could be safely treated by the PCP.

2.2. Changes in 1997

The incentive system we have described was essentially unchanged from 1991 through 1996. In 1997 the rules changed. New Federal regulations required that the HMO reduce the intensity of cost-control incentives. Federal law mandated that no more than 25 percent of physician pay could be “at risk.” The HMO responded by reducing the withhold to 10 percent, while continuing to use the same fee schedule. In addition, they replaced the “gain-sharing” bonus with a new system that capped bonuses at 15 percent of income and, importantly for what follows, linked bonuses to both cost and quality targets. The quality targets consisted of specialty specific preventive care measures (e.g., proportion of patients immunized for pediatricians, proportion receiving mammograms for gynecologists), and some other measures such as patient satisfaction, office inspections, and patient turnover.⁹

3. A Simple Theory of Group Incentives

The essence of the HMO’s incentive system is easily summarized. For each patient, a target expenditure level is established. PCPs who influence expenditure decisions are members of groups (PODs). If, on average, the physicians in the *group* hold expenditures at or below the target, the physicians receive a substantial bonus. In this section we lay out a model which allows us to understand the nature of this group incentive and to derive testable predictions concerning its effect on physician behavior.

It is helpful at the outset to think about the HMO’s goals. The HMO’s long-run

⁹ In the new system introduced in 1997, the HMO ranked by expenditures all PODs who came in under their spending targets. The top quartile of these PODs got 7.5 percent of fees as a bonus. This same set of PODs were then ranked by quality and the top quartile got 7.5 percent of fees as a bonus. PODs that did less well in cost containment or quality got smaller bonuses.

viability in the marketplace hinges on its ability to balance two objectives—on the one hand providing health care that is viewed as high quality, and on the other hand providing health insurance at a reasonable price. Firms and their employees who contract with the HMO presumably understand that the HMO will be providing incentives to physicians to prevent excessive health expenditures. These clients must be confident that the incentive eliminates “wasteful” expenditures (e.g., reduces the use of high cost medical procedures that have low expected value to patients) while still encouraging physicians to act in a fashion that does not substantially deviate from the best interests of their patients.

3.1. A Model of Physician Behavior

To fix the central idea about the HMO’s role in the market, consider a physician who derives utility from money (y) and from the medical services (m) provided to patients. 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} .¹⁰

We conceive of \hat{m} as representing the level of medical services 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 and because physicians value good relationships with their patients. Thus in some cases, \hat{m} may represent the acknowledged medical best practice for treatment of a patient with a specific illness. Providing a less expensive and less effective treatment would cause disutility for the physician even if the patient could not judge for himself what best practice was. 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

¹⁰ Our model is tailored to the specific institutions and issues we examine. See McGuire (2000) for an excellent review of the general literature on models of physician behavior.

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 any alternative level of services, if those services were provided at no cost. Patients who pay for medical services, though, would prefer to contract for a level of service somewhat lower than \hat{m} .¹¹ An important role for the HMO is to set in place incentives that accomplish the goal of inducing doctors to set medical services below \hat{m} , but not by "too much." This task is complicated to a considerable degree by the fact that neither patients nor the HMO administration know \hat{m} .

As an especially transparent example consider a doctor with n patients ($i = 1, \dots, n$), whose utility is given by a simple quasi-linear form,

$$U = y - \sum_{i=1}^n v(m_i - \hat{m}_i),$$

with $v' < 0$ and $v'' > 0$ when $m_i < \hat{m}_i$.¹² Since an HMO in a competitive market will want to maximize patient satisfaction for any given level of medical expenditures, we begin by considering how this doctor would ration medical care subject to a fixed budget ($\sum m_i = M$), but no further financial incentives. The doctor would follow the rationing rule,

$$-v'(m_i^* - \hat{m}_i) = \lambda, \quad i = 1, \dots, n. \quad (1)$$

Not surprisingly, the doctor would allocate limited resources so that the marginal disutility from rationing care (relative to the patient's ideal) is equated across patients.

¹¹ Since \hat{m} is a patient's unconditional optimum, it lies on a flat spot in the patient's objective function. 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.

¹² Given these assumptions, preferences are convex and utility is increasing in both y and m_1 over the relevant range ($m_i < \hat{m}_i$).

In what follows, we show that the HMO can induce the physician to follow this same optimal allocation rule by adopting an easily implemented bonus scheme.

We assume that the HMO management cannot observe \widehat{m}_i , the ideal for patient i , or even the level of medical expenditures intended by the physician, m_i . Instead the HMO observes

$$x_i = m_i + \varepsilon_i,$$

where ε is a mean-zero random variable reflecting the fact that the physicians in the HMO have imperfect control over medical expenditures. The mean of this random variable has probability and cumulative distribution functions ϕ and Φ , respectively.¹³ Think of the random variable as representing the fact that physicians do not have complete control over medical expenditures.

Suppose that the HMO offers a per-patient bonus, B , that is awarded only if, average expenditures on medical services in a year are below a specified target, i.e., only if $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i < X_T$, where X_T is the target. The HMO also directly compensates observed x_i at the expected opportunity cost to the doctor (think of m_i as being scaled in dollar value of the opportunity cost of services).¹⁴ The implied budget constraint is

$$E(y) = nB\Phi(X_T - \frac{1}{n} \sum m_i).$$

When the doctor maximizes utility $[y - \sum_{i=1}^n v(m_i - \widehat{m}_i)]$ subject to this budget constraint, the resulting first-order conditions are

$$-v'(m_i^* - \widehat{m}_i) = B\phi(X_T - \bar{m}^*), \quad i = 1, \dots, n, \quad (2)$$

where m_i^* is the optimal choice.

¹³ In our empirical work below, we assume that ε has a normal distribution, but here we need only assume that the mean of the random variable meets the monotone likelihood ratio property.

¹⁴ In addition, there might be a lump-sum transfer between the doctor and the HMO to insure that the participation constraint is met, but this is not important to our discussion here since it does not affect incentives.

Figure 1 illustrates the doctor’s problem. The budget constraint slopes downward in the (y, m_i) plane. The curve labelled $V = V_0$ is an indifference curve for the physician, with the “U” shape reflecting the physician’s dislike for deviation from the \widehat{m}_i ideal. As depicted in Figure 1, the optimal choice is $m^* < \widehat{m}$.

We have three observations. The first is the equivalence of the first-order conditions (1) and (2). The HMO’s bonus incentive induces the same allocation rule an individual doctor would adopt if presented the task of rationing a fixed level of medical services to a given pool of patients. Secondly, we observe that although the bonus scheme we describe is discrete, the effective *expected* budget constraint is likely a smooth downward sloping function in each argument $E(y) = y(m_1, \dots, m_n)$ because of the stochastic nature of the outcomes. Third, the HMO can clearly influence the doctor’s level of medical service provision (m_i^*) by adjusting the incentive payments (B) and by altering the target spending level (X_T). However, the doctor’s chosen service levels also depend on the ideal levels (\widehat{m}_i) which are unobservable by the HMO, and indeed are stochastic from the HMO’s perspective. Thus for any chosen bonus level and target, the gap between m_i^* and \widehat{m}_i will depend on the unobservable clinical needs and preferences of the patients.

From equation (2) it is apparent how the stochastic component of the problem affects the extent to which the physician deviates from the ideal. Specifically, m_i^* will be close to \widehat{m}_i when the doctor lands in either tail of the expenditure distribution. When a doctor has an unusually sick (hence expensive) patient load, achieving the target is unlikely. In this case, the marginal effect of any treatment decision on the probability of making the target (ϕ) is low. Similarly for a doctor with an unusually healthy (hence inexpensive) patient population, the marginal effect of any treatment decision on the probability of making the target (ϕ) is also low. In both these cases, the physician can more easily “afford” to be generous in treating any patients who do need medical attention.

3.2. Why Group Incentives?

The rationale for implementing a bonus scheme on a group basis—rather than a physician-level basis—follows directly from the observations made in the previous paragraph. We have seen that a simple physician-level bonus plan provides doctors with the incentive to reduce medical expenditures in a sensible manner, by equalizing across patients the marginal disutility of rationing medical services. If the HMO uses a physician-level incentive bonus, however, this marginal disutility of allocating medical care is not necessarily equalized *across doctors*, even when doctors have identical preferences. Thus two identical patients (with the same \hat{m}_i) will receive different levels of medical services m_i^* if they choose different primary care physicians who have patient populations with different average medical costs.

The HMO can seek to reduce this disparity by adjusting the physician’s target on the basis of observable patient characteristics (like age and gender), but this is a partial remedy because risk adjusting is imperfect. An additional step for reducing the disparity in treatment is to group physicians, and provide a bonus based on observed group outcomes. Suppose doctors are placed into a POD of N doctors, and given a group incentive. If the average observed level of medical services for the group is less than a given target, a per-patient bonus B is awarded. We assume that doctors allocate the bonus in proportion to their share of patients (this is typically the case in the HMO we study). We assume also that doctors play Nash strategies.

Let there be N doctors, designated by j , with n_j patients treated by doctor j . Each doctor’s share of the group’s patients is then $s_j = \frac{n_j}{n}$, where $n = \sum_{j=1}^N n_j$. Physician j will choose the quantity of medical services that solves

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

where \bar{m}_j is the average level of medical expenditures for doctor j and \bar{m}_{-j} is the

mean for other doctors in the group. The first-order conditions for this problem are

$$-s_j B\phi(X_T - \bar{m}) - v'(m_{ij}^* - \hat{m}_{ij}) = 0, \quad i = 1, \dots, n. \quad (3)$$

We assume that the second order conditions,

$$s_j B\phi'(X_T - \bar{m}) - v''(m_{ij}^* - \hat{m}_{ij}) < 0, \quad i = 1, \dots, n. \quad (4)$$

also hold.

Notice that pooling patients across doctors can reduce the difference across patients in the marginal disutility of rationing care. Indeed if all doctors have equal size patient loads (so that s_j is the same for all physicians in the POD) and also have identical preferences concerning patients, the marginal disutility of rationing care is equalized across all patients in the POD.¹⁵

In the HMO we study there is some indication that the designers of the system understood the potential gains from group incentives, but their understanding of the economic logic of these gains was imperfect.¹⁶ As a result, they allowed for the formation of PODs which differed substantially in size and in the patient shares of PCPs within PODs.

This variation is fortunate for our purposes because our theory provides a clear prediction about how medical expenditures should vary with share. In particular, application of the implicit function theorem to (3) gives the intuitive comparative static result,

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

¹⁵ The larger the POD, the greater the number of patients for whom this allocation property holds. Why then doesn't the HMO organize all the physicians in the network into a single POD? As the first-order conditions make clear, the larger the POD, the more dilute will be any bonus incentive. As we have noted, the bonus size is limited by law. Even it were not, physician risk aversion (which we do not model here) would mitigate against allowing the size of the bonus (and hence the size of the POD) to become excessively large.

¹⁶ They clearly understood that in small panels random variation could cause some PODs to be too far above or below cost targets. This, for example, was the rationale for grouping otherwise unconnected PCPs into PODs.

Thus, all else equal, physicians who have a small share of the PODs patients will provide higher levels of medical services than physicians with larger shares. Similarly, per patient expenditures will be higher in large PODs (where physicians generally have relatively low shares) than in small PODs. We turn next to physician-level and POD-level data to evaluate these predictions.

4. Empirical Evidence

In this section we ask whether the HMO's incentive system had the effect of reducing medical expenditures. Our strategy is to examine the correlation between medical utilization spending and the intensity of incentives to limit these expenditures.

The data we use comes from the HMO's internal records and is available at the PCP level and the POD level. We have data from the roughly 1000 PCPs and 100 PODs in the network over the years 1994-1997. Section 4.1 focuses on the individual PCP-level data. We begin the section by imposing some additional structure on our theoretical model in order to derive a simple and estimable empirical model. Taking this "structural" approach to the data allows us to test our basic hypothesis while also uncovering some otherwise hard to observe aspects of the incentive system.

In section 4.2, we analyze POD-level rather than PCP-level data. While the POD-level equations lack the clear structural interpretation of the PCP level estimates, they do exploit the fact that the HMO kept far more detailed records at the POD than the PCP level. We use these additional data to check the robustness of the results in section 4.1. In addition, we use the POD-level data to analyze the types of expenditures for which the HMO incentive system matters.

4.1. Physician-Level Analysis

4.1.1. Econometric Specification

The theoretical model in the previous section leads to a simple specification for

a physician-level expenditure equation. Recall that all doctors are assumed to have preferences described by

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

with $v' < 0$ and $v'' > 0$ for $m < \widehat{m}$. For the purposes of empirical implementation we introduce an assumption about the form of the utility function: let $v(m_{ij} - \widehat{m}_{ij})$ be the function $\alpha(m_{ij} - \widehat{m}_{ij})^2$.

Given this specification, the first order conditions (3), for doctor j who has share s_j within the panel, can be written

$$s_j B\phi_j + 2\alpha(m_{ij}^* - \widehat{m}_{ij}) = 0, \quad i = 1, \dots, n.$$

These first-order conditions can be solved for m_{ij}^* ,

$$m_{ij}^* = -\frac{1}{2\alpha} [s_j B\phi_j] + \widehat{m}_{ij}. \quad (5)$$

Neither we nor the HMO observe m_{ij}^* , but the HMO does record realized expenditures, x_{ij} , which we have assumed are related to m_{ij}^* by

$$x_{ij} = m_{ij} + \varepsilon_{ij}.$$

Substituting this last equation into (5) we obtain,

$$x_{ij} = \widehat{m}_{ij} - \frac{B\phi_j}{2\alpha} s_j + \varepsilon_{ij}. \quad (6)$$

We assume that the sum $\widehat{m}_{ij} + \varepsilon_{ij}$ is equal to a linear combination of observed characteristics $(\beta_0 + \beta_2 Z_{ij})$ plus an error term u_{ij} .

We have data on average annual medical expenditures by physician. Let x_{jt} be the mean of x_i in year t for doctor j and u_{jt} be the mean of u_i for the same physician in year t . We assume u_{jt} is orthogonal to s_{jt} .¹⁷ The equation we estimate is thus

¹⁷ In short we are assuming that unobserved components of physician j 's expenditures are uncorrelated with physician j 's share of the POD's members. Our POD-level analysis relaxes this assumption.

$$x_{jt} = \beta_0 + \beta_1 \phi_j s_{jt} + \beta_2 Z_{jt} + u_{jt}, \quad (7)$$

where ϕ_j is specific to the POD in which doctor j is a member. We do not observe ϕ_j directly, but it can be estimated, as discussed shortly. If our assumptions are correct, we can give β_1 a structural interpretation, $\beta_1 = -\frac{B}{2\alpha}$. The value of B is determined by the incentive contract that is common across physicians in a year, while α is an unobserved parameter that determines the relative importance of income and deviations from “ideal” medical expenditures for physician utility.

Our identification strategy is straightforward: the slope of physician j ’s budget constraint is $B\phi_j s_j$. This varies across physicians because of variation in $\phi_j s_j$ (especially s_j , as it turns out). Doctors with a share near 0 essentially have a flat budget constraint, and will choose m_{ij}^* very close to \hat{m}_{ij} for their patients, while doctors with a relatively large share have a relatively steep budget constraint, and will choose reduced medical expenditures relative to \hat{m}_{ij} . This variation allows us to map physician responses to incentives, and in turn to identify the parameter α .

We can estimate equation (7) using a simple two-stage procedure outlined below.¹⁸ An obvious alternative strategy is to simply apply ordinary least squares to the equation,

$$x_{jt} = \gamma_0 + \gamma_1 s_{jt} + \gamma_2 Z_{jt} + u_{jt}. \quad (8)$$

We interpret the parameter γ_1 as an “average” of $\frac{-B\phi}{2\alpha}$, and we expect it to be negative.

4.1.2. Estimation Results

Columns 1 and 2 of Table 1 present estimates based on expenditure equations (7) and (8) respectively. We derive the estimates in column 1 as follows: First, we

¹⁸ The simplifying assumptions of our model enable us to implement an unusually transparent structural estimation procedure. There is relatively little work in the incentives literature that applies structural estimation, and papers that do take this approach (e.g., Ferrall and Shearer, 1999) generally employ more complicated methods.

estimate a probit model predicting the probability that PODs meet the incentive. Second, we use these estimates to form $\hat{\phi}_j$ for each doctor.¹⁹ Equation (7) is then estimated using OLS, treating $\hat{\phi}_j$ as “data.”²⁰ In both equations Z consists of dummy variables for year and specialty effects. Finally, due to the changes in the incentive system in 1997, we interact the key incentive variable with a 1997 dummy. These 1997 interaction terms are estimated quite imprecisely and we ignore them in the following discussion.

The estimated coefficients on *Physician’s Share of POD’s Member-months* * ϕ (equivalent to $\hat{\beta}_1$ in equation 7) and *Physician’s Share of POD’s Member-months* (equivalent to $\hat{\gamma}_1$ in equation 8) are both negative and statistically significant at conventional levels, consistent with our hypothesis that a doctor with a larger share of a group’s patients should have lower expenditures.

According to our structural interpretation of column (1), *Physician’s Share of POD’s Member-months* * $\phi = \hat{\beta}_1$, thus $\hat{\beta}_1$ is estimated to be -128.7. From equation (6), $\hat{\beta}_1 = -\frac{B}{2\alpha}$. Since the bonus B is approximately \$3.00, the implied value of α is 0.012.²¹

Recall that the first order conditions for physicians are

$$-s_j\phi_j B = 2\alpha(m_{ij}^* - \hat{m}_{ij}).$$

The left-hand side describes the slope of the “budget constraint,” and hence the “power” of the incentive scheme. The right-hand side is the marginal rate of sub-

¹⁹ The probit estimates are contained in a table in the Appendix.

²⁰ Because ϕ_j is estimated, conventional standard errors are not appropriate. We use a two-step, clustered bootstrap procedure to arrive at the standard errors. In step one we estimate $\hat{\phi}_j$ for 100 replications (clustered at the POD level) of the probit predicting whether or not the panel exceeded it’s target. For each of these 100 estimates, we ran 100 replications of the individual medical expenditure equation (clustered at the PCP level). The standard errors of the coefficients taken from the resulting 10,000 replications are reported in column 1 of Table 2.

²¹ This estimate of B assumes a bonus of a little more than 20 percent of an average PCP’s monthly expenditure of \$14.61 per member per month, approximately the HMO’s withhold for an average physician.

stitution between income and deviations in medical expenditures from the “ideal” levels that would prevail in the absence of cost considerations. If the doctor has a near zero share of the POD’s patients, medical expenditures will approximately equal \widehat{m}_{ij} .²² If the doctor has a share of 0.1 (the sample mean) and ϕ of 0.33 (also the sample average), the slope of the budget line ($s_j\phi_jB$) is -0.10 . In this case a physician who prescribes a medical procedure that costs \$100 has an expected decline in income of \$10. We estimate that for this doctor, the absolute value of $(m_{ij}^* - \widehat{m}_{ij})$ is \$4.29 per member per month. This represents savings of approximately 5 percent on average medical expenditures per-member per-month of \$81. Finally, if the physician has a share of 0.2, so that the slope of the budget constraint is 0.20, medical expenditures decline by \$8.58. Physicians are clearly responsive to financial incentives.

Figure 2 illustrates an average physician’s estimated indifference curve for a utility level of 1000 (picked arbitrarily), and the optimal levels of medical treatment for that physician when share equals 0.1 (the sample average) and when share is 0.3. A physician with a share of 0.1 will reduce medical treatment 5 percent below the “ideal” level. Increasing the share to 0.3 would induce a reduction of 15 percent.

The coefficient on *Physician’s Share of POD’s Member-Months* in Column 2 (-40.5) has a straightforward “reduced form” interpretation. A physician with the average patient share in our sample (0.1) will have medical expenditures that are \$4.05 (per member per month) lower than a physician who has a share of 0. Notice that this inference is virtually identical to the estimate based on the structural interpretation of column (1).

The savings implied by these estimates are substantial. The doctors in our sample serviced just over 800,000 HMO member months in 1996. If all physicians

²² 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).

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.

4.2. *POD-Level Analysis*

The results in Table 1 suggest that physician incentive contracts do have the effect of reducing medical utilization expenditures. An obvious concern with these estimates, however, is that our measures of physician share may be correlated with unobserved features of enrollees or physicians that have an independent influence on expenditures. In this section we use the more extensive POD-level data to investigate this potential omitted variable bias. In addition, we use the POD-level data to investigate the types of utilization costs that are most influenced by incentives.

4.2.1. *Econometric Specification*

If incentives matter, we can expect that medical expenditures increase with the number of physicians in the POD (since the average doctor’s share (s_j) declines with the number of physicians in the POD). This reasoning also suggests that reductions in the size of incentive bonuses should attenuate the relationship between POD size and spending levels.

Our basic regression for testing these hypotheses is:

$$x_{kt} = \delta_0 + \delta_1 \ln(N_{kt}) + \delta_2 \ln(N_{kt}) * D_{1997} + \delta_3 Z_{kt} + u_{kt}, \quad (9)$$

where x_{kt} is the observed medical expenditures per member per month in POD k at time t , $\ln(N_{kt})$ is the log of the number of physicians in POD k in year t and D_{1997} is a dummy variable equal to 1 for the year 1997.²³ Z is a vector of variables that

²³ The POD-level equations do not have the natural structural interpretation of the PCP level equations. Our theoretical exposition suggests that the relationship between

capture the influence of other factors that potentially influence medical expenditures (and δ_3 is a row vector of parameters).²⁴

In terms of the parameters of the expenditure equation, our first hypothesis suggests that $\delta_1 > 0$. Because of the Federally mandated reduction of incentive intensity in 1997, we anticipate $\delta_2 < 0$ (the second hypothesis above).

4.2.2. Estimation Results

Column (1) of Table 2 presents estimates of equation (9) for the years 1994-1997.²⁵ 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*1997 interaction. Point estimates suggest that in 1997, an increase in POD size from 10-12 physicians led to an increase in expenditures of only \$1.18 per member per month. This finding is consistent with the reduction in incentive intensity that occurred 1997.

One of the “control variables” included in vector Z is *Target Expenditures*. This target is the forecast of a POD’s “expected expenditures” based on enrollee age and gender, and PCP. The estimated coefficient of 0.99 on *Target Expenditures* indicates an approximately 1 to 1 relationship between expected POD medical expenditures and actual POD expenditures. The equation in column (1) also includes the variable

POD size and medical expenditures will be non-linear and concave. We use the log specification because it offers a convenient interpretation of the coefficients. Other specifications produced similar results.

²⁴ In the POD-level regressions Z includes variables measuring: the number of patients in the POD (measured in patient member-months), POD network affiliation, specialty composition of the POD, and percent of members enrolled in Medicaid in 1997. Details on these variables are provided in Data Appendix 1.

²⁵ 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.

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 urban settings 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.

Column (2) in Table 2 introduces an alternative dependent variable, *Medical Expenditures Net of Stop Loss*. Under the HMO's incentive system, physicians are not held responsible for expenditures above \$15,000 per patient in a year. We would, therefore, worry about our interpretation of the coefficients on *Log Physicians in POD* and *Log Physicians in POD*1997* if they changed dramatically when we net out "stop loss" expenditures. We observe instead, however, that the coefficients in column (1) are virtually identical to those in column (2).

Column (3) introduces yet another dependent variable, *Fraction Withhold Returned*. The HMO cares about regulating total medical expenditures. The physicians, in contrast, care about whether they can win back the withhold. Consistent with our theoretical reasoning, we find that as the number of physicians in a POD increases, the panel is less likely to win back their withhold. Point estimates indicate that an increase in panel size from 10 to 12 physicians reduces the fraction withhold returned by 0.04 or roughly 10 percent of the mean. The coefficient on *Log Physicians in POD*1997* is imprecisely measured and not statistically significant.

The estimates in columns (4) through (6) of Table 2 add additional control variables to capture heterogeneity across PODs and their enrollee populations. These variables are described in detail in Data Appendix 1, but their coefficients are suppressed in Table 2 to facilitate exposition. The addition of these control variables in columns (4) through (6) had no effect on the signs and statistical significance and little effect on the magnitudes of the coefficients on our key variables of interest, *Log*

Physicians in POD and *Log Physicians in POD*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.

An alternative approach to controlling for unobserved POD heterogeneity is to estimate our equations with POD-level fixed effect variables. These fixed effect estimates are presented in columns (7) through (9) of Table 2. The coefficients on the key variables, *Log Physicians in POD* and *Log Physicians in POD*1997* are not qualitatively changed from those in equations (1) through (6), although the magnitude of the effect of number of physicians on expenditures 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 *Log Physicians in POD* and *Log Physicians in POD*1997* in columns (7) through (9) as reflecting changed behaviors due to the changing degree of incentive intensity. An alternative possibility is that it is selection rather than behavioral changes that are driving our fixed effect results.²⁶ It may be, for example, that *Log Physicians in POD* has a positive correlation with *Total Medical Expenditures* because panels that reduce the number of physicians do so by expelling 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.

Table 3 offers an investigation of the importance of selection effects. It relies on individual expenditure data for the HMO's primary care physicians for the years

²⁶ Barro and Beaulieu (2000) offer evidence of the importance of selection effects under conventional insurance.

1994–1996. We use these data in Panel A to see if those leaving shrinking groups have spending that is, 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 PCPs in the panel they were leaving. This difference between movers and stayers is both small in magnitude and statistically insignificant.²⁷ In Panel B of Table 3, we compare total medical expenditures between incumbent physicians and new entrants in the panel. Of the 746 new entrants in the years 1995 or 1996, we observe that entrants had expenditures that are, on average, 1 percent higher than those in the panel they were entering. This difference is far too small to account for the results in Table 2. It is also statistically insignificant.

In Table 4, we explore the plausibility of our results by examining the impact of incentives on different types of medical expenditures. Data on the components of spending are only available for 1994–1996. Thus we begin by re-estimating our base model (column (1) of Table 2) for these years. The estimated coefficient *Log Physicians in POD*, is very close to those estimated over the full sample of years (1994–1997).

Under the HMO’s incentive system PCPs can reduce medical expenditures in two distinct ways. First, they can reduce the amount of income that accrues to them via fees by reducing the number of office visits they have with patients. Alternatively PCPs can cut expenses that accrue as fees to other health care providers, e.g., by reducing specialist referrals. Given the payout formula in the HMO’s incentive contracts, we would expect physicians to prefer the second method of reducing

²⁷ 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.

expenditures to the first.

Column (2) of Table 4 presents estimates of the determinants of *Expenditures Not Paid to Self (pmpm)*. This variable measures the average per member per month medical expenditures for procedures and treatments that do not earn revenues for the primary care physician. We find that the coefficient on *Log 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 *Log Physicians in POD* in column 3 is positive, but small in magnitude and 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 generates fees for other physicians. Physicians do not seem to cut expenditures on services that generate fees for themselves.

Columns (4) and (5) of Table 4 compare the determinants of spending due to in-patient and out-patient services and procedures. The dependent variable in column (4), *Total In-Patient (pmpm)*, tracks spending due to procedures conducted in a hospital on an in-patient basis. We observe that the coefficient on *Log Physician in POD* in column (4) is small and not statistically different from zero. Column (5) estimates the determinants of out-patient expenses. The variable *Total Out Patient* includes expenditures incurred in a hospital (on an out-patient basis), a clinic, or a doctor's office. In column (5) of Table 4, we find that *Log Physicians in POD* has a positive and statistically significant effect on out-patient spending. The magnitude of this effect is large. 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 in-patient expenses, but are

important for out-patient expenses. This finding makes sense. Patients admitted to hospitals are likely to be sicker than other patients and therefore are less likely to have purely elective procedures. Patients in a hospital are also more likely to have their care decided by a specialist rather than the by the primary care physicians and it is the PCPs who are most affected by the HMO's incentive system.²⁸ In contrast, out-patient procedures are more likely to be elective and more likely to be controlled by the PCP. We would therefore expect the incentive system to work most powerfully on these latter expenditures.²⁹

5. Incentives Along Other Dimensions

The results in the previous sections provide evidence that HMO incentives do motivate PCP “gatekeepers” to reduce medical expenditures. Much of the controversy surrounding HMO incentive systems involves the claim that such cost-control efforts require physicians to sacrifice the quality of medical care. Care quality in medicine is notoriously hard to measure, but some, admittedly imperfect, quality metrics have become standard in the industry. Because our data include some of these standard quality metrics, we have the opportunity to examine the quality issue, at least with regard to *measured* quality.

The HMO first incorporated quality measures into its incentive system in 1997. As discussed above, this change was part of a general overhaul of the system to bring it into compliance with the then new Federal regulations governing physician incentive plans.

²⁸ Our findings that in-patient expenditures are unresponsive to physician incentives is consistent with recent research on HMOs by Cutler, McClellan and Newhouse (2000). They find that HMOs reduce the costs of treating heart attack patients (who are largely in-patients) by reducing the money paid to providers (presumably by hard bargaining) but not by altering the amount or type of care these patients received.

²⁹ Stop loss provisions are also more likely to pertain for hospitalized patients, further reducing the PCP's incentives to reduce spending.

Under the new system, quality incentives were linked to cost-control incentives. Panels with a surplus in their enrollee accounts were rank ordered based on care quality, and bonuses were paid out according to rank. A key feature of this system is that quality payouts were conditional on keeping patient expenditures below the targeted levels for that year.

In constructing quality measures, the HMO gave PODs points based on two types of quality measures. The first type of measure focused on preventive care and included such things as the percent of women aged 50-65 who had mammograms or the percent of children aged 3-5 who had annual well-child evaluations. This type of measure is referred to as a HEDIS measure because the indicators were derived from a set of standardized quality measures put forward by the Health Plan Employer Data and Information Service (HEDIS).³⁰ In addition to the selected HEDIS measures of preventive care, the HMO constructed a second quality measure based on indicators of patient satisfaction, the accessibility of the PCPs after hours and direct office inspections. Details on the plan's HEDIS and non-HEDIS quality measures are presented in Data Appendix 2. As we describe in the appendix, the non-HEDIS measures proved to be both difficult to collect and of doubtful relevance to panel quality. For this reason, we focus our analysis on the HEDIS measures.³¹

Table 5 analyzes the determinants of medical expenditures and quality in 1997. Column (1) of Table 5 presents a regression of *Total Medical Expenditures* on HEDIS Quality Points. We find a statistically significant and negative relationship between

³⁰ HEDIS is sponsored, supported and maintained by the National Committee for Quality Assurance (NCQA), the organization responsible for accrediting managed care organizations.

³¹ For two of the non-HEDIS measures (membership retention and extended hours), the indicators were so unreliable that all panels were simply given the maximum number of points at the end of the year. The third non-HEDIS measure (patient satisfaction) gave maximum points to all but two panels. The fourth measure (whether the panel was accepting new HMO enrollees) has little connection to quality of care. The fifth measure (focused office reviews) was based on criteria that were not clearly explained to physicians.

expenditures and HEDIS points, i.e. the panels with the best quality scores have the lowest average expenditures. In columns (2) and (3) we find that this negative correlation is statistically significant for stop-loss expenditures but not for other medical expenditures.³²

Under the incentive contract in place in 1997, payouts for measured quality were influenced by the same “free riding” problems as payouts for spending-containment. In addition, panels were not eligible for quality bonuses unless they also kept medical expenditures under target levels. For both these reasons, we would expect quality measures to deteriorate as the number of physicians in the panel increased. Improvements in quality are reflected in an increase in the HEDIS Quality Points measure. Thus, we expect that HEDIS Quality Points should decrease as the number of physicians in the panel increases.

Column (4) of Table 5 analyzes the determinants of the HEDIS Quality Points earned by a panel. We find that the number of quality points earned by a panel falls as the number of physicians in the panel grows. The coefficient is statistically and economically significant. Increasing the number of physicians from 10 to 12 reduces HEDIS Quality Points by 1.19, or 7 percent of the mean.

Column (5) of Table 5 re-estimates the HEDIS Quality Points equation in column (4), but introduces additional variables that measure characteristics of panels, their physicians and enrollees. Introducing these variables has little effect on the magnitude and statistical significance of the coefficient on Log Physicians in POD.

The results in columns (4) and (5) are consistent with the notion that the HMO’s incentive system is moving physicians to offer both lower cost care and care with higher measured quality. The HEDIS quality measures emphasize preventive care and reward those panels that are best at educating and tracking their enrollees.

³² We want to emphasize that we are not attempting to estimate a causal relationship in columns (1) - (3), but rather exploring the correlation between medical expenditures and quality.

These same organizational capacities may also help panels reduce expenditures. Alternatively, it may be that low cost panels have high measured quality because they are especially committed to winning incentive payouts. These panels, for example, may find ways to make sure their flu immunization and mammography rates are high without otherwise improving the true quality of care they deliver. The quality measures included in the HMOs incentive system are quite crude, and they should be treated as highly imperfect indicators of the quality of care actually delivered. It is entirely possible that physicians are responding to the HMO's incentives by producing higher measured quality and shirking on non-metered aspects of quality. Nonetheless, we are not aware of any previous estimates of the impact of physician incentives on any aspects of quality. These results suggest that physicians do respond positively to incentives for metered quality and they raise the possibility that properly designed incentive systems needn't necessarily produce a reduction in quality.

6. Concluding Remarks

This paper analyzes the effect of contracts that give physicians in an HMO network a financial incentive to reduce medical expenditures. Our study focuses on a particular HMO that runs a network of independent primary care physicians. We find that cost-containment incentives have a statistically and economically significant effect on expenditures for patients' medical services. Indeed, it appears that the incentive system in place at the HMO between 1994 and 1996 reduced medical utilization expenditures by about 5 percent. This reduction was concentrated in spending for out-patient rather than in-patient services.

Using a similar methodology, we analyzed the effect of quality incentives introduced in 1997. Even though we have data for only one year and the quality metrics are crude, we find suggestive evidence that physicians respond to quality incentives

by improving quality—at least for the preventative care indicators metered and rewarded under the incentive system.

Given our focus on a single organization over a short period of time, we cannot be certain about the extent to which our results generalize to other managed care settings. Nevertheless, our findings suggest that critics of managed care are right to be concerned about the effects of physician incentives. In the HMO we study, these incentives clearly do induce physicians to take actions on behalf of patients that differ from those they would choose in the absence of incentives.

At the same time, our findings suggest that supporters of managed care are right to be concerned about policy initiatives that make it difficult for HMOs to implement effective incentive systems. HMO incentive systems are found to limit expenditures, and thus can be expected to serve an important role in reducing the price of health care coverage. Moreover, in the HMO we examine there is no evidence that the financial incentives adversely affected standard quality indicators along a number of (crudely) measured dimensions.

In conducting this research we relied on three key elements: (1) detailed knowledge of the structure of the organization’s incentive system, (2) a model of behavior tailored to this setting, and (3) detailed data on the outcomes. Without all three legs of this tripod, our ability to draw inferences about the incentive system would have been seriously limited. For example, the incentives in place from 1994 to 1996 involved a bonus that exceeded 20 percent of physician income. Intuitively, this might sound like a rather high-powered incentive. After all, each additional medical procedure ordered for a patient holds some probability of bumping the physician over target, potentially costing the physician many thousands of dollars. In fact, our empirical model shows that for a typical physician in the HMO network, expected income falls \$0.10 for every \$1.00 of additional expenditures, implying a marginal rate of substitution of -0.1. This tradeoff between income and expenditures induces

physicians to reduce expenditures by 5 percent.

Our experience in this project leads us to conclude that the development of an empirically grounded economic theory of incentives depends on research that combines the use of detailed knowledge of the organizational context, a carefully developed model that is appropriate to the context, and rich data drawn from within the organizations under study. The search for research settings with these key ingredients will, we believe, shape the continued development of organizational economics.

References

- Baker, George, Gibbs, Michael and Holmstrom, Bengt. 1994. "The Internal Economics of the Firm: Evidence from Personnel Data," *Quarterly Journal of Economics* 109(4), 881–919.
- Barro, Jason R. and Beaulieu, Nancy Dean. 2000. "Selection and Improvement: Physician Responses to Financial Incentives," unpublished manuscript, Harvard Business School.
- Coase, R. H. 1937. "The Nature of the Firm," *Economica* 52(5), 386–405.
- Cutler, David, McClellan, Mark and Newhouse, Joseph P. 2000. "How Does Managed Care Do It?" *Rand Journal of Economics*, 31(3), 526–548.
- Dudley, R. Adams, Miller, Robert H., Korenbrot, Tamir Y., and Harold S. Luft. 1998. "The Impact of Financial Incentives on Quality of Health Care," *Milbank Quarterly*, 76(4), 649–686.
- Ferrall, Christopher and Bruce Shearer. 1999. "Incentives and Transactions Costs Within the Firm: Estimating an Agency Model Using Payroll Records," *Review of Economic Studies*, 66, 2, 309–338.
- Gaynor, Martin, and Gertler, Paul. 1995. "Moral Hazard and Risk Spreading in Medical Partnerships," *Rand Journal of Economics*, 26(1), 591–613.
- Gibbons, Robert. 1997. "Incentives and Careers in Organizations," in David 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.
- Gold, Marsha. 1999. "Financial Incentives in Managed Care: Current Realities and Challenges for Physicians," *Journal of General Internal Medicine*, 14(Supplement 1), S6–S12.
- Ichniowski, Casey, Katheryn Shaw, and Giovanna Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines," *American Economic Review*, 87(3), 291–313.
- InterStudy, 1999, *The InterStudy Competitive Edge, Part II: The HMO Industry Report 9.1*, St. Paul, MN: Interstudy Publications.
- Landers, Renee M., James B. Rebitzer, and Lowell J. Taylor. 1996. "Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms," *American Economic Review*, 86(3), 329–48
- Lazear, Edward P. 2000. "Performance, Pay and Productivity," *American Eco-*

conomic Review, 90(5), 1346–61.

McGuire, Thomas G. 2000. “Physician Agency,” in Anthony J. Culyer and Joseph P. Newhouse, eds., *Handbook of Health Economics*. Volume 1A. Amsterdam: North-Holland.

Newhouse, Joseph P. 1996. “Reimbursing Health Plans and Health Providers: Selection Versus Efficiency in Production,” *Journal of Economic Literature*, 34(3), 1236–1263.

Prendergast, Canice. 1999. “The Provision of Incentives in Firms,” *Journal of Economic Literature*, 37(1), 7–63.

Table 1
Determinants of Physician Level Medical Expenditures

	[1]	[2]
	Total Medical Expenditures	Total Medical Expenditures
<i>Physician's Share of POD's Member-months * ϕ</i>	-128.651 [-2.172]	
<i>Physician's Share * ϕ * 1997</i>	-295.823 [-1.228]	
<i>Physician's Share of POD Member-months</i>		-40.463 [-2.00]
<i>Physician's Share * 1997</i>		-138.404 [-1.56]
Constant	yes	yes
PCP Specialty Fixed Effects	yes	yes
Year Dummies	yes	yes
Observations	2282	2282
Number of Physicians in Sample	961	961
R-squared	0.017	0.018

Numbers in [] are t-statistics.

t-statistics in column 1 bootstrapped, allowing for clustering by physician.

t-statistics in column 2 are OLS, with robust t-statistics allowing for clustering by physician.

The dependent variable is the annual dollars of total medical expenditures pmpm for an individual physician.

The mean and standard deviation of the dependent variable are respectively 81.73 and 176.10.

For variable *Physician's share** ϕ , the variable ϕ is calculated in two steps. First we use a probit to estimate the probability that a POD comes in below it's target level of expenses. Second, we use the results to estimate Φ , the predicted probability that the panel beats it's target costs and ϕ , the p.d.f.

The mean and standard deviation for *Physician's Share* are .099 and .108 respectively.

The mean and standard deviation for *Physician's Share* * ϕ are 0.033 and 0.039 respectively.

PCP specialties are internal medicine, family practice and pediatrics.

Table 2
POD Level Expenditure Equations

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	Fixed Effect (7)	Fixed Effect (8)	Fixed Effect (9)
Dependent Variable	Total Medical Expenditures (pmpm)	Expenditures Net Stop Loss (pmpm)	Fraction Withhold Returned	Total Medical Expenditures (pmpm)	Expenditures Net Stop Loss (pmpm)	Fraction Withhold Returned	Total Medical Expenditures (pmpm)	Expenditures Net Stop Loss (pmpm)	Fraction Withhold Returned
<i>Log Physicians in POD</i>	36.168	31.911	-0.269	27.346	24.494	-0.272	46.102	46.39	-0.314
(robust t-statistics)	(2.955)	(3.544)	(3.157)	(2.290)	(2.937)	(3.420)	(3.017)	(3.569)	(2.519)
<i>Log Physicians in POD*1997</i>	-29.669	-29.808	0.12	-26.879	-27.531	0.149	-27.862	-23.459	0.117
	(2.507)	(3.208)	(1.245)	(2.129)	(2.697)	(1.628)	(1.928)	(1.908)	(0.995)
<i>Target Expenses (pmpm)</i>	0.99	0.823	-0.001	0.756	0.66	0.001	0.6	0.536	0.001
	(6.938)	(6.074)	(0.904)	(3.629)	(3.184)	(1.177)	(5.829)	(6.121)	(1.405)
<i>Years POD is in Sample Since 1994</i>	9.111	6.615	-0.025	6.443	4.962	-0.013	28.607	23.679	-0.549
	(2.434)	(2.028)	(0.534)	(1.409)	(1.283)	(0.254)	(0.978)	(0.952)	(2.301)
<i>Year is 1995</i>	-14.516	-4.418	-0.011	-12.198	-5.741	-0.013	-30.865	-19.133	0.592
	(1.651)	(0.667)	(0.148)	(1.356)	(0.895)	(0.166)	(1.015)	(0.740)	(2.387)
<i>Year is 1996</i>	-51.834	-29.038	0.051	-47.079	-28.777	0.048	-88.932	-60.931	1.204
	(4.481)	(3.201)	(0.492)	(3.754)	(2.740)	(0.445)	(1.524)	(1.227)	(2.530)
<i>Year is 1997</i>	-8.909	4.868	-0.4	-17.176	-4.726	-0.283	-71.918	-50.294	1.085
	(0.244)	(0.149)	(1.117)	(0.437)	(0.130)	(0.821)	(0.948)	(0.779)	(1.754)
Constant	yes	yes	yes	yes	yes	yes	yes	yes	yes
Additional POD Level Variables	yes ¹	yes ¹	yes ¹	yes ²	yes ²	yes ²	yes ¹	yes ¹	yes ¹
POD Fixed Effects	no	no	no	no	no	no	yes	yes	yes
Observations	353	353	349	340	340	339	353	353	349
Number of PODs	118	118	116	109	109	109	118	118	116
Adjusted R-squared	0.457	0.464	0.066	0.454	0.488	0.154	-0.058	-0.039	-0.384

All equations estimated using data from 1994-1997. The median POD is in the sample for three of the four years.

¹ Variables include: the log of POD enrollees & it's interaction with 1997 dummy.

² Variables in note 1 (above) plus POD network affiliation, POD specialty mix, and % members enrolled in Medicaid in 1997. Descriptions of these variables are available in the data appendix.

Table 3
 Selction vs. Incentives: Do High Cost Physicians Exit Shrinking Panels and
 Do High Cost Physicians Enter Growing Panels?

Panel A

		Number of Physicians in POD Declines from Year t to Year t+1:			
		no	yes	All	
Physician Exits Panel After Year t	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	1.003	0.995
		Number of Physicians in Cell	678	338	1016
All		Relative Costs Year t	0.998	1.005	1.000
		Number of Physicians in Cell	935	379	1314

Panel B

		Number of Physicians in POD Increases from Year t-1 to Year t:			
		no	yes	All	
Physician Enters in Year t	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 of 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

"Relative Costs" is the physician's total medical expenses in year t relative to the panel's total medical expenses in year t. None of the cross-cell difference in the panel are statistically significant.

In panel A, t refers to years 1994 and 1995. In panel B, t refers to years 1995 and 1996.

In panel A, of the physicians in the sample, 298 exited the panel after years 94 or 95. The relative costs of these exiters was 2% above the mean of their POD for both shrinking and growing PODs.

In panel B, of the physicians in the sample, 746 entered the panel in years 95 or 96. The relative costs of these entrants was 1% above the mean of their POD. This 1% differential also held for growing PODs.

We restrict data to 1994-1997 because a change in the HMOs accounting systems made it difficult to track those who move across PODs between 1996 and 1997.

Table 4
Medical Utilization Expenditures by Category

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)
Dependent Variable	Total Medical Expenditures (pmpm)	Expenditures Not Paid to Self (pmpm)	Expenditures Paid to Self (pmpm)	Total In- patient (pmpm)	Total Out- Patient (pmpm)
Log Physicians in POD	26.227 (2.110)	24.062 (1.932)	2.165 (1.364)	5.055 (0.674)	9.512 (2.020)
Target Expenses (pmpm)	0.627 (2.156)	0.55 (2.334)	0.077 (1.274)	0.224 (3.331)	0.31 (1.852)
Years POD is in Sample	11.319	9.64	1.679	5.162	5.583
Since 1994	(1.250)	(1.144)	(1.420)	(1.356)	(1.156)
Year is 1995	-16.521 (1.512)	-14.995 (1.409)	-1.527 (1.361)	-14.621 (2.353)	-0.905 (0.201)
Year is 1996	-54.104 (2.918)	-50.957 (2.910)	-3.147 (1.264)	-30.488 (4.439)	-9.794 (1.030)
Constant	yes	yes	yes	yes	yes
Additional POD Level Variables ¹	yes (0.785)	yes (0.443)	yes (2.135)	yes (0.820)	yes (0.098)
Observations	236	236	236	236	236
Number of PODs	99	99	99	99	99
Adjusted R-squared	0.42	0.414	0.393	0.176	0.472

Absolute value of robust t-statistics in () for OLS estimates.

All equations estimated using data from 1994-1997.

¹ log POD enrollees, POD network affiliation, POD specialty mix, and % members enrolled in Medicaid in 1997. Descriptions of these variables are available in the data appendix.

Table 5
Determinants of HEDIS Quality Measures 1997

Dependent Variable	(1) Total Medical Expenditures (pmpm)	(2) Stop Loss (pmpm)	(3) Expenditures Net Stop Loss (pmpm)	(4) HEDIS Quality Points	(5) HEDIS Quality Points
HEDIS Quality Points	-0.805 (3.744)				
HEDIS Quality Points		-0.666 (3.343)			
HEDIS Quality Points			-0.138 (1.532)		
Log Physicians in POD				-6.628 (2.857)	-5.663 (2.157)
Target Expenses (pmpm)				-0.095 (1.779)	-0.03 (0.555)
Years POD is in Sample				4.529	4.425
Additional POD Level Variables ¹	no	no	no	no	yes
Observations	107	107	107	107	104
Adjusted R-squared	0.101	0.096	0.009	0.287	0.309

¹ Variables include: the log of POD enrollees, POD network affiliation, POD specialty composition, and % POD enrollees with Medicaid in 1997

Descriptions of these variables are available in the data appendix.

Data Appendix 1
Variable Names and Descriptions
(Tables 2, 4, 5)

Variable Name	Mean	Std. Dev.	Min	Max	Notes
Dependent Variables					
Total Medical Expenditures (pmpm)	89.56	65.60	8.82	583.09	The sum of expenditures due to in-patient and out-patient procedures, professional services and "other" expenditures, minus the amount paid by other insurers.
Total In Patient Expenditures (pmpm) [#]	28.76	33.46	0.00	263.99	sub-category of total medical expenses available for 1994-1996.
Total Out Patient Expenditures (pmpm) [#]	35.39	31.43	5.27	306.48	sub-category of total medical expenses available for 1994-1996.
Fraction Withhold Returned	0.50	0.49	0.00	1.00	Fraction of withheld fees returned to POD at end of year.
Expenditures Paid to Self (pmpm) [#]	14.49	11.28	0.00	107.52	Utilization Expenses that Accrue as Income to the referring PCP's.
Expenditures Not Paid to Self (pmpm) [#]	84.83	67.75	8.82	475.57	Total Medical Expenses - Expenditures Paid to Self.
Hedis Quality Points [@]	16.08	13.52	0.00	50.00	Described in second data appendix.
Other Variables					
Log Physicians in POD	2.16	0.53	1.10	3.56	Log of the number of physicians in POD.
Log Member-months	8.48	1.69	2.71	11.27	Log of the number of HMO membermonths at the panel.
Single Specialty POD ¹	0.60	0.49	0.00	1.00	POD physicians all in one specialty: either internal medicine, family practice, pediatrics or ob/gyn.
Target Expenses (pmpm)	79.42	36.51	14.92	265.79	Level of per member per month expenditures at which withhold is returned and below which bonuses are paid.
% Medicaid > 10 in 1997 ¹	0.30	0.46	0.00	1.00	Dummy variable = 1 when % medicaid > 10% of panel's patient population. HMO began accepting Medicaid in Feb. 1995, but it remained very small (and wasn't counted separately) until 1997.
Year is 1995	0.25	0.43	0.00	1.00	Dummy variable = 1 when year is 1995.
Year is 1996	0.26	0.44	0.00	1.00	Dummy variable = 1 when year is 1996.
Year is 1997	0.31	0.46	0.00	1.00	Dummy variable = 1 when year is 1997.
Years POD is in Sample Since 94	2.27	1.04	1.00	4.00	
POD In Network 1 ¹	0.03	0.18	0.00	1.00	Dummy variable indicating panel is part of larger network.
POD In Network 2 ¹	0.03	0.18	0.00	1.00	Dummy variable indicating panel is part of larger network.
POD In Network 3 ¹	0.06	0.24	0.00	1.00	Dummy variable indicating panel is part of larger network.
POD In Network 4 ¹	0.17	0.38	0.00	1.00	Dummy variable indicating panel is part of larger network.
POD In Network 5 ¹	0.08	0.27	0.00	1.00	Dummy variable indicating panel is part of larger network.
Fraction POD Internal Medicine ¹	0.30	0.41	0.00	1.00	
Fraction POD Pediatrics ¹	0.26	0.39	0.00	1.00	

Data Appendix 1
Variable Names and Descriptions
(Tables 2, 4, 5)

Variable Name	Mean	Std. Dev.	Min	Max	Notes
Fraction POD Family Practice ¹	0.25	0.38	0.00	1.00	
POD Only OB/GYN ¹	0.11	0.32	0.00	1.00	

Data available for 1997 only

@ Data available for 1994-96 only

¹ one of the "Additional POD Level Variables" in Tables 2, 4 and 5

Data Appendix 2
Probability of POD's Withhold Being Returned
Probit

Dependent Variable	All Withhold Returned
Log Physicians in POD (robust t-statistics)	-0.873 [3.65]
Log Physicians in POD*1997	0.437 [1.57]
Log Member-months	0.127 [1.39]
Log Membermonths*1997	-0.007 [0.05]
Target Expenses (pmpm)	0.002 [0.85]
Years POD is in Sample Since 1994	-0.102 [0.70]
Year is 1995	-0.082 [0.34]
Year Is 1996	0.241 [0.78]
Year Is 1997	-0.571 [0.57]
Single Specialty POD	0.016 [0.08]
% Medicaid > 10 in 1997	0.093 [0.36]
POD In Network 1	1.304 [3.34]
POD In Network 2	1.726 [3.94]
POD In Network 3	-0.81 [1.80]
POD In Network 4	0.762 [2.27]
POD In Network 5	0.499 [1.23]
Fraction POD Family Practice	-0.19 [0.38]
Fraction POD Internal Medicine	0.333 [0.63]
Fraction POD Pediatrics	0.561 [1.09]
POD Only OB/GYN	0.013 [0.03]
Constant	0.165 [0.22]
<hr/>	
Observations	339
Number of PODS	109
Wald chi2(20)	75.8
Prob > chi2	0
Pseudo R ²	0.1448
Robust z statistics in brackets	

Dependent variable equal to 1 if POD got 100% of it's withhold returned and 0 otherwise.

Data Appendix 3

Quality Incentives 1997

Quality incentives were the sum of points awarded for two types of quality measures: quality of care/preventative care measures and other measures of patient satisfaction. The first category was referred to within the HMO as HEDIS measures. These measures varied by specialty.

HEDIS Quality Measures (50 points maximum)

Pediatrics:

- One point given for each percentage point above 70% compliance with age specific immunizations for children aged birth to 2 years.
- One point given for each percentage point above 60% compliance with annual well child examination for children aged 3-5 years.

Internal Medicine:

- One point given for each percentage point above 60% of a mammogram in the past 24 months in women age 50-65.
- One point given for each percentage point above 50% of patients with high risk medical conditions or age 65+ having received an annual influenza vaccine.

Family Practice:

- One point given for each percentage point above 70% compliance with age specific immunizations for children aged birth to two years.
- One point given for each percentage point above 60% of a mammogram in the past 24 months in women age 50-65.

OB/GYN

- One point given for each percentage point above 60% of a mammogram in the past 24 months in women age 50-65.
- One point given for each percentage point above 65% of women 18-65 with a Pap test in the past three years.

Non- HEDIS Quality Measures (50 points maximum + Medicaid Supplement)

- Panels given 10 points and then points were deducted for families requesting a change in PCP for perception of poor quality of care or poor provider/patient relationships. Deductions varied with size of panel. In 1997, problems with indicator led to all panels being assigned 10 points. This indicator was improved and continues to be used in the HMOs incentive system.

- Points awarded based on patient satisfaction results from the annual patient satisfaction survey. This measure proved to be uninformative. All but 2 panels in 1997 got the full 10 points for this measure.
- Points awarded to PCPs offering regular office hours outside of M-F, 8AM to 5PM. In 1997 problems with indicator led to all panels being assigned 10 points.
- Fraction of physicians open for accepting new patients from the HMO times 10.
- Points awarded on the basis of a focused office review, a kind of "white glove" inspection of office facilities including checks of medical records.
- 10 quality points awarded to PCPs with more than 10% of their enrollees in Medicaid.

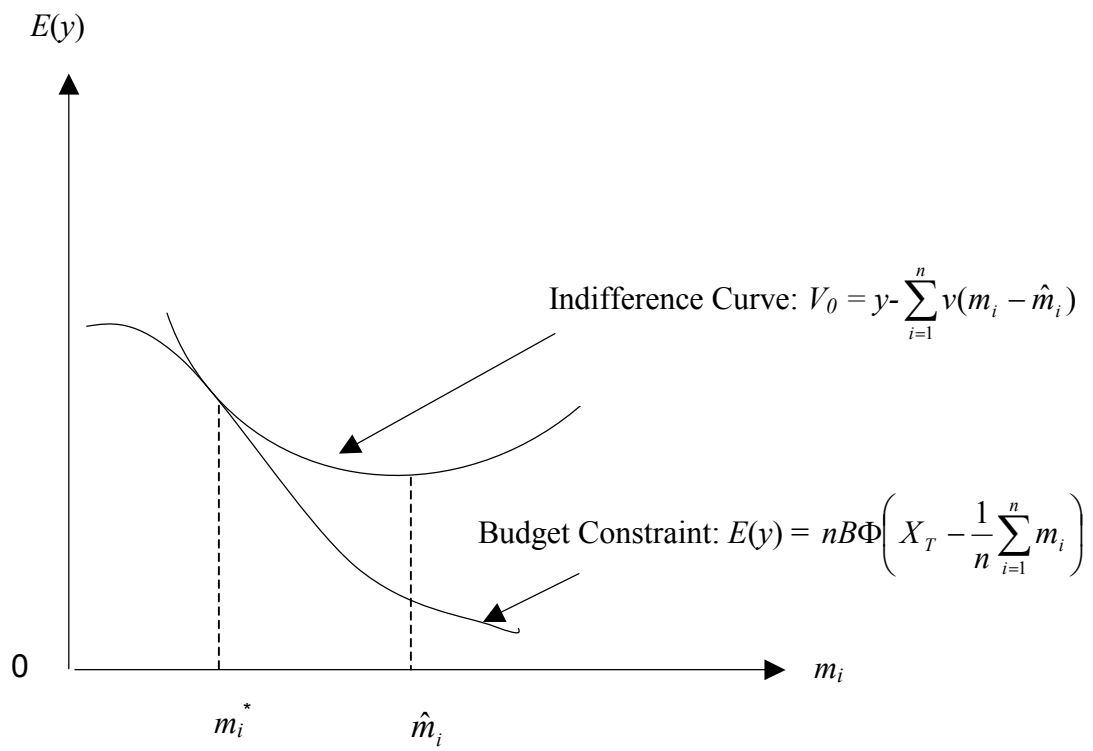


Figure 1: The Physician's Optimum Medical Expenditures For Patient i

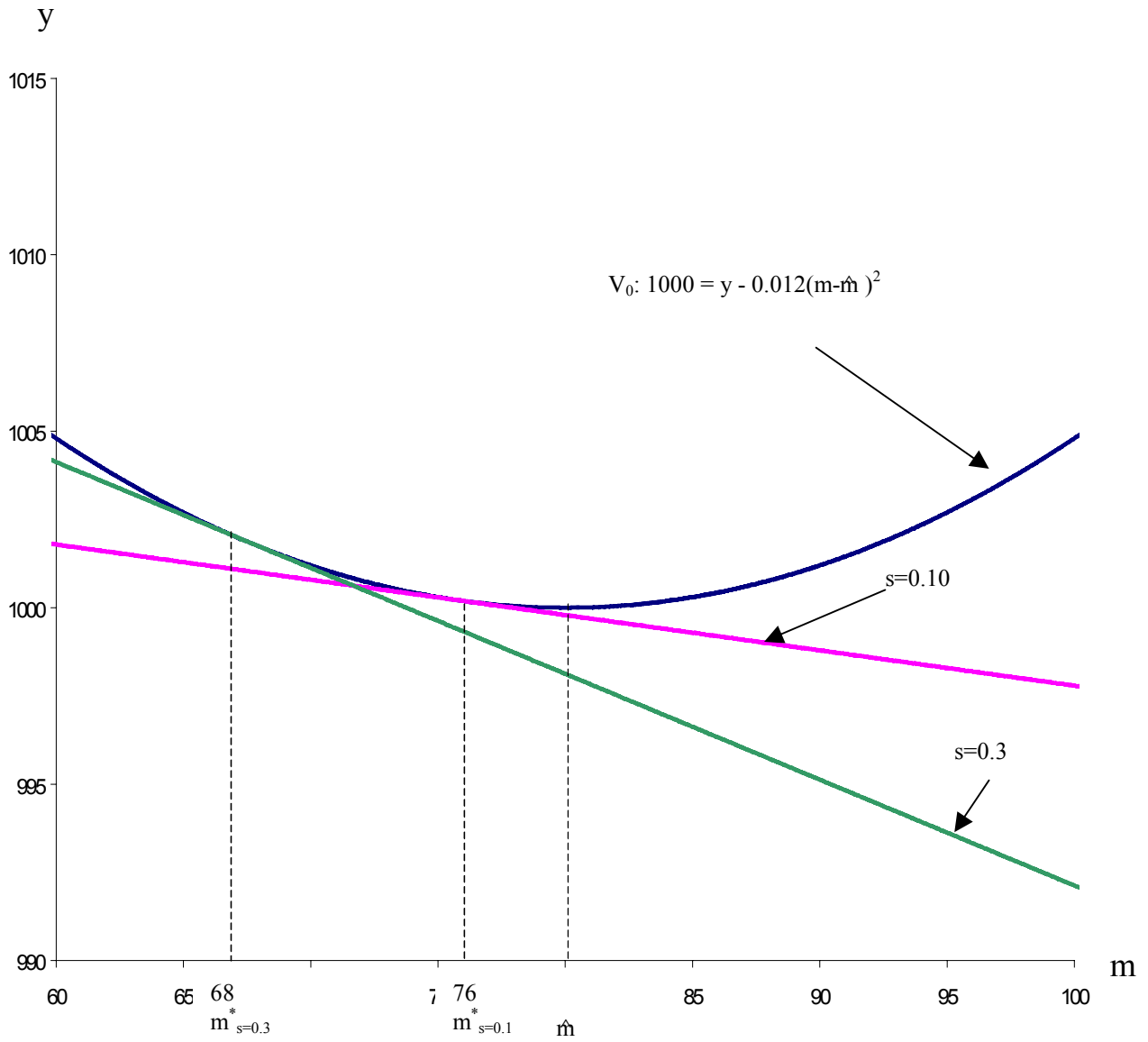


Figure 2: Optimal m for differing values of s_j