Getting Real Performance Out of Pay-for-Performance

SEAN NICHOLSON, MARK V. PAULY,
ANITA YA JUNG WU, JAMES F. MURRAY,
STEVEN M. TEUTSCH, and MARC L. BERGER

*Cornell University; University of Pennsylvania; Nationwide Better Health;
Merck; Eli Lilly*

Context: Most private and public health insurers are implementing pay-for-performance (P4P) programs in an effort to improve the quality of medical care. This article offers a paradigm for evaluating how P4P programs should be structured and how effective they are likely to be.

Methods: This article assesses the current comprehensiveness of evidence-based medicine by estimating the percentage of outpatient medical spending for eighteen medical processes recommended by the Institute of Medicine.

Findings: Three conditions must be in place for outcomes-based P4P programs to improve the quality of care: (1) health insurers must not fully understand what medical processes improve health (i.e., the health production function); (2) providers must know more about the health production function than insurers do; and (3) health insurers must be able to measure a patient's risk-adjusted health. Only two of these conditions currently exist. Payers appear to have incomplete knowledge of the health production function, and providers appear to know more about the health production function than payers do, but accurate methods of adjusting the risk of a patient's health status are still being developed.

Conclusions: This article concludes that in three general situations, P4P will have a different impact on quality and costs and so should be structured differently. When information about patients' health and the health production function is incomplete, as is currently the case, P4P payments should be kept small, should be based on outcomes rather than processes, and should target

---

Address correspondence to: Sean Nicholson, 123 MVR Hall, Cornell University, Ithaca, NY 14853 (email: sn243@cornell.edu).

The Milbank Quarterly, Vol. 86, No. 3, 2008 (pp. 435–457)
© 2008 Milbank Memorial Fund. Published by Wiley Periodicals Inc.

physicians' practices and health systems. As information improves, P4P incentive payments could be increased, and P4P may become more powerful. Ironically, once information becomes complete, P4P can be replaced entirely by "optimal fee-for-service."

Keywords: Pay-for-performance, reimbursement, quality.

SINCE THE PROPOSAL FROM MEDICARE TO TEST AND invest in pay-for-performance (P4P) programs building on the groundswell of private-sector initiatives, there has been an explosion of interest in and debate on the merits and methods of P4P. The P4P expectations span the continuum from being the answer to the current woes of the health care system to outright skepticism and cynicism about another flawed attempt for a quick fix that avoids the need for fundamental change and true reform.

Is P4P the future, or will it fade and eventually pass into the graveyard of past failed health care reform efforts? As often happens, truth probably lies somewhere in between. In this article, we offer some ideas for reflection as part of the debate. Our goal is not to praise or condemn. Rather, we hope to temper expectations while offering some important considerations that will maximize P4P's potential for success and identify some principles for considering where P4P can have the most constructive impact on health. Even though P4P will undoubtedly influence the cost of care, it is unrealistic to expect that we can depend only on incentives to systematically improve quality and, at the same time, reduce costs.

Although a majority of payers have instituted pay-for-performance programs (Rosenthal *et al.* 2006), it is too early to tell whether P4P programs, currently based largely on process measures, are improving quality and/or reducing costs. Several articles in the trade press have proclaimed the Rewarding Results and Integrated Healthcare Association programs to be a success, but these conclusions are based primarily on the growing size of performance payments rather than improved outcomes produced by the payments. The nine randomized controlled evaluations of P4P programs that have been published (Adams 2004) were of programs instituted in the 1980s and 1990s and measured provider performance with a single indicator, whereas the second wave of P4P programs tracks multiple indicators (Baker 2004). The effect of this second wave of P4P programs is mixed, and most of the studies suffer from concerns regarding selection into voluntary programs, small sample size,

and/or absence of a valid control group.¹ In a recent journal issue devoted to evaluating P4P programs, Mark Chassin concluded that “none of the designs employed in these early studies is robust enough to provide convincing evidence that any of the P4P programs was responsible for any of the improvements cited” (Chassin 2006, p. 123S).

P4P programs can be structured to pay physicians and hospitals an amount above the basic payment for a service when they provide medical processes believed to improve health outcomes, when they invest in structures (e.g., electronic medical records systems) believed to improve health outcomes, or when they improve risk-adjusted outcomes. Currently, most P4P programs provide incentives for providing specific processes, such as eye exams for diabetic patients or beta blockers for heart attack patients. We argue in the next section that P4P would be more effective if payers made adequate direct payments for processes whose importance is well understood and used P4P programs to provide incentives for better outcomes. Our proposed pay-for-performance system, therefore, can be viewed as paying for outcomes rather than for processes. Our formulation is most relevant to fee-for-service settings, typical of the great preponderance of health care in the United States.

Two broad types of information are required for an outcomes-based P4P system to work well: (1) measures of health outcomes that are appropriately adjusted for the severity of a patient’s initial illness and that can be positively affected by clinical management; and (2) information about specific treatments and clinical management that can transform people with an initially poor health status into those with better outcomes, that is, knowledge of what might be called the *health production function*.

We propose that an outcomes-based P4P system should be designed differently, in accordance with how much of the required information is actually available. In three situations an outcomes-based P4P will have differing impacts on quality and so should be structured differently in each situation. These three situations differ in how well payers can measure a patient’s health and how much payers and providers know about the health production function; that is, how specific clinical management regimens and treatments translate into changes in health.

The next section of this article describes the three conditions that must be present in order for P4P to substantially improve the quality of medical care. We then present data to demonstrate that only two of the three conditions now prevail, discuss consequences associated

with implementing P4P in this setting, and recommend design changes that can minimize negative consequences. In the following section, we describe a paradox: if P4P programs provide sufficiently strong incentives for providers and payers to expand the bounds of evidence-based medicine, P4P might eventually become irrelevant. We next discuss whether P4P is likely to develop organically from the market or whether government assistance will be needed. Finally, we summarize the article and offer some concluding thoughts.

When Pay-for-Performance Could Be Relevant and Powerful: The Right Combination of Information and Ignorance

Pay-for-performance will be most effective in situations in which health insurers and providers of medical care have the right combination of information and ignorance. Specifically, the following three conditions must be in place: (1) health insurers must not fully understand which medical processes improve health, which we refer to as the *health production function*; (2) individual providers must know more about the health production function than insurers do; and (3) health insurers must be able to measure a patient's initial risk-adjusted health, his or her ultimate health, and each provider's contribution to the change in health.

Consider a situation in which medicine could be codified into a "cook-book," so that if physicians ordered certain tests and performed certain procedures, a patient's health would improve as much as possible. Armed with this cookbook, health insurers could implement an "optimal" fee-for-service system by increasing fees sufficiently to encourage physicians to use the effective processes and reducing fees (possibly to zero) for ineffective processes. We describe such a system later. Although health insurers could achieve the same results using a process-based P4P system that gives providers extra payments if they use effective processes, the point is that when health insurers have complete knowledge of the health production function, such a P4P system will not provide any benefit beyond a well-designed fee-for-service system.

Now consider the more probable situation in which health insurers know that all providers should use certain medical processes (e.g., blood sugar tests for diabetic patients); there is no evidence to justify evaluating other processes; and insurers do not know the best way to combine

many possible processes to produce the best health outcomes. Some providers, however, have more complete knowledge regarding how to produce good health outcomes based on their unique clinical experience, knowledge of individual patients, willingness to experiment, and/or superior management systems, such as information technology or care pathways.

In this situation, health insurers could use a two-part reimbursement strategy: increase fees for processes known to be effective and base part of a provider's payment on the improvement in the risk-adjusted health of the provider's patients. The higher fee will encourage all providers to use evidence-based medicine, and the outcomes-based P4P component will create incentives for providers to use processes that they know or believe will, on average, improve health. This strategy may encourage physicians to address issues they have traditionally avoided, such as patients adhering to the recommended treatment and coping with cultural/ethnic disparities. Providers who know more about or invest more time in learning about the health production function will, on average, receive higher payments. Providers will also have incentives to invest in systems (e.g., electronic medical records, clinical decision support systems, computerized physician order entry systems, care pathways) and programs that improve outcomes (e.g., personnel to follow the chronic care model), and they will be rewarded for their innovation.

Based on this discussion, the "performance" measure in the P4P system ought to be an intermediate (e.g., HbA1c measure for a diabetic patient) or a final health outcome measure (e.g., mortality), not a measure of process (e.g., whether patients received some recommended medical service) or a provider's investment in a "structure" (e.g., electronic medical records, computerized physician order entry systems). If a health insurer knows that all providers should use certain processes, then insurers should pay for these processes on a fee-for-service basis, raising the fee if necessary to provide stronger incentives for the processes to be carried out in the optimal way. By tying a proportion of provider payments to changes in health outcomes, health insurers will encourage providers to pursue treatments that providers believe produce good outcomes, even though health insurers have not yet determined how to do so. Once a process has become validated scientifically and widely accepted as being a vital component of a treatment plan, health insurers can increase its fee and reduce the proportion of payments tied to health outcomes. From this point on, therefore, when we use the term *P4P*, we are referring

to a reimbursement program that provides supplemental payments to providers based on the risk-adjusted health outcome, or a change in the risk-adjusted health outcome, of a provider's patients.

What impact would P4P have in this situation? Health outcomes would certainly improve relative to those based on a straight fee-for-service system, particularly if the providers know a lot more about the health production function than the insurers do. If health insurers are really able to reduce fees for processes with unknown efficacy, providers will use only those processes that improve health outcomes because they will receive little profit (for hospitals) or rents (for physicians) for providing a service *per se*, but more profit or rent through the performance-based outcome payments.

Our own judgment is that even though the number of cases in which the structure of information is ideal for outcomes-based P4P but unsuited for "optimal" fee-for-service is quite small now, it could increase over time as evidence-based medicine and the method of measuring initial health and ultimate outcomes continue to develop. Still, we cannot be certain that the right combination of information and ignorance will materialize.

An outcomes-based P4P will not succeed if providers and health insurers have identical information about the health production function. Although providers would have incentives to improve health, they would not have the means to do so beyond what the fee-for-service payment schedule promotes. Furthermore, if the risk adjustment system could not determine whether a patient's health actually improved, providers would have little incentive to innovate.

Today's Situation: Possible Unintended Consequences of Pay-for-Performance

Only two of the three conditions required for an outcomes-based P4P to succeed now exist. Payers appear to have incomplete knowledge of the health production function, and providers appear to know more about the health production function than payers do. Nonetheless, accurate methods of risk adjusting a patient's health status are still being developed. Under these circumstances, P4P could generate unintended consequences if providers avoided the sickest patients in an attempt to improve their reported outcomes or required a payment premium to accept the risk inherent in P4P programs not completely adjusted for

risk. Although this does not mean that P4P should be abandoned, it does mean that the first wave of outcomes-based P4P programs should use small payments targeted at physicians' practices or hospital systems rather than at the individual physician or hospital.

Our assumption that payers have incomplete information about the health production function is based on two types of analysis: the strength of relationships between the currently recommended process of care measures and patient outcomes, and an evaluation of whether evidence-based medicine is comprehensive. Three recent studies demonstrate that despite the relationship between the recommended care processes that payers are using to denote high-quality care and risk-adjusted patient health outcomes, this relationship is fairly weak. Bradley and colleagues (2006) show that seven CMS/JCAHO process measures (e.g., administering a beta blocker when a patient is admitted to a hospital) for treating acute myocardial infarction collectively explained only 6 percent of the variation among hospitals' thirty-day risk-adjusted mortality rates. Likewise, Werner and Bradlow (2006) and Jha and colleagues (2007) found that differences among hospitals' use of processes, shown on the Center for Medicare and Medicaid Services' (CMS) Hospital Quality Alliance database, are associated with only small differences in thirty-day, one-year, or in-hospital risk-adjusted mortality rates for three different health conditions. The processes that health insurers believe are important to producing good outcomes appear to explain only a few of the providers' differences in outcomes.

Another conclusion from these three studies is that certain providers are producing superior health outcomes because, at least in part, they are using processes other than those being measured. That is, producing good outcomes is an art as well as science, and some providers are better at this than others. Although providers appear to understand the health production function better than payers do, there is evidence that their understanding also is incomplete. The American College of Cardiology (ACC) and the American Heart Association (AHA) recently identified five process measures believed to have an important impact on heart failure patients' outcomes. In fact, however, none of the five measures (e.g., evaluation of left ventricular systolic function) was significantly associated with sixty-day or ninety-day risk-adjusted mortality rates (Fonarow et al. 2007).

We assessed the current comprehensiveness of evidence-based medicine by examining the Institute of Medicine's recommended "starter set" of twenty-six measures for evaluating the quality of outpatient care.

Eighteen of these are process measures that record the proportion of appropriate patients who receive specific medical services (Institute of Medicine 2006). We chose these eighteen process measures, listed in table 1, based on five criteria: scientific soundness, feasibility, importance, alignment, and comprehensiveness. In order to determine the comprehensiveness of these measures, we estimated the proportion of

TABLE 1
Extent of Information on the Outpatient Health Production Function:
Institute of Medicine's Recommended "Starter Set" Process
Performance Measures

Recommended Service	Actual/Potential Outpatient Spending in the United States, 2004 (\$ millions)
Prevention	
a. Breast cancer screening	\$7,706 ^a
b. Colorectal cancer screening	\$9,180 ^a
c. Cervical cancer screening	\$2,260 ^a
d. Smoking cessation advice	\$539 ^a
e. Flu vaccines	\$1,790 ^a
f. Pneumonia vaccines	\$1,570 ^a
Coronary artery disease	
g. Lipid-lowering drugs	\$15,593
h. Beta blocker after heart attack	\$14,616
Heart failure	
i. ACE inhibitor/ARB therapy	Included in (h)
j. LVF assessment	\$474
Diabetes	
k. HBA1c tests	\$169 ^a
l. Blood pressure medication	Included in (h)
m. Lipid measurement	\$309 ^a
n. Eye exam	\$1,186 ^a
Asthma	
o. Prescribed inhaled corticosteroids	\$4,941
Depression	
p. Prescribed antidepressant medication	\$13,390
Prenatal care	
q. Screening for HIV at first or second prenatal visit	\$59 ^a
r. Anti-D immune globulin treatment for appropriate patients	\$18 ^a

(Continued)

TABLE 1—*Continued*

Recommended Service	Actual/Potential Outpatient Spending in the United States, 2004 (\$ millions)
Actual/potential spending associated with P4P measures	\$73,800
Total outpatient spending	\$694,400
P4P spending as a percentage of total	10.6

Notes: Prescription drug spending for the United States by class of drugs is reported on the IMS Health website, accessed April 20, 2006. Medicare physicians' fees and lab fees were derived from the CMS payment files for 2004. Medicare physicians' fees are based on the relative value unit assigned for a particular service. We assume that private health plans pay physicians 23 percent more than Medicare does, based on data in the Medicare Payment Assessment Commission's 2004 Report to Congress (Medicare Payment Assessment Commission 2004). Recommended patient populations for cancer screening and vaccinations are from the American Cancer Society (ACS) and the Centers for Disease Control and Prevention (CDC), respectively. For example, the CDC recommends that 71.5 million people receive a flu shot, and the ACS recommends that 69.0 million women receive a breast cancer-screening exam.

^aEstimated spending if recommended guidelines were adhered to perfectly; all other entries are actual expenditures.

Source: Institute of Medicine 2006.

total U.S. outpatient spending in 2004 represented by the medical services that are believed to improve quality.

The estimated spending reported in table 1 is based on a series of assumptions that are described more fully in the note following the table. In some cases, we reported the amount actually spent on a recommended service. For example, patients and health insurers together spent a total of \$4.9 billion in 2004 on prescribed inhaled corticosteroids to treat asthma. In other cases, we estimated how much would have been spent if recommended guidelines were followed precisely (i.e., potential spending), as the proponents of P4P programs desired. For example, if all women over the age of thirty-nine received an annual mammogram, as recommended by the American Cancer Society, patients and health insurers would spend an estimated \$7.7 billion on breast cancer screening. We estimated that these eighteen measures accounted for \$74 billion of the \$694 billion spent in 2004 on outpatient medical care, or only 10.6 percent of the total. Current efforts do not address or account for the other 89 percent of spending that may well improve the quality of medical care.

Unfortunately, the third element, our ability to measure risk-adjusted outcomes, which is required to make the outcomes-based P4P effective,

currently is limited, and health insurers appear to be aware of this limitation. Only nine of the twenty-three prominent P4P programs in table 2 are using clinical outcome measures, whereas fourteen are using subjective, patient-reported outcome measures to evaluate the providers' performance. This limitation has hurt our ability to generalize and scale P4P across all health care services. What would happen if P4P systems

TABLE 2
Structure of Prominent P4P Programs

P4P Program	Type of Performance Measures	Performance Bonus as a Percentage of Payer's Reimbursement
Physicians		
Bridges to Excellence	Process, structure	Max.: 10%
Integrated Healthcare Association (IHA)	Clinical outcomes, patients' experiences, process, structure	Avg.: 1.5%
Blue Cross Blue Shield of CA	Clinical outcomes, patients' experiences	Avg.: 5%
PacifiCare Health System	Patients' experiences, structure	Avg.: 5%
Aetna/IHA	Patients' experiences, structure	Avg.: 3.5%
Anthem Blue Cross Blue Shield of NH	Patients' experiences, process, structure	Max.: 5%
Blue Cross of CA	Clinical outcomes, patients' experiences, process, structure	Avg.: 5.5%
Blue Cross Blue Shield of MO	Patients' experiences, process	Avg.: 8%
CIGNA	Patients' experiences, process, structure	Avg.: 4%
Hawaii Medical Service Association	Patients' experiences, process	Avg.: 5.5%
HealthGuard (PA)	Process	Avg.: 2.5%
Health Net (CA)	Clinical outcomes, patients' experiences, process	Avg.: 5.5%
Highmark Blue Cross Blue Shield of PA	Clinical outcomes, patients' experiences, process	Max.: 5%

(Continued)

TABLE 2—*Continued*

P4P Program	Type of Performance Measures	Performance Bonus as a Percentage of Payer's Reimbursement
Independence Blue Cross of PA	Patients' experiences, process, structure	Avg.: 5.5%
Independent Health (NY)	Patients' experiences, process	Avg.: 4%
Excellus/Rochester Individual Practice Association (RIPA)	Patients' experiences, process	Avg.: 10%
Employer Coalition on Health	Process	Avg.: 1%
Tri-River Healthcare Coalition (OH)	Process, structure	Avg.: 5%
	Overall average	5.1%
	Hospitals	
Blue Cross Blue Shield of MI	Clinical outcomes	Avg.: 3.3%
Anthem of Virginia (formerly Trigon) Quality-in-Sights Hospital Incentive Program	Clinical outcomes, process	Avg.: 1%
Empire Blue Cross and Leapfrog employers	Structure	Avg.: 2%
Independence Blue Cross (PA)	Clinical outcomes, structure	Avg.: 4%
Premier (Medicare)	Clinical outcomes, process	Avg.: 0.3%; max.: 2%
	Overall average	2.1%

Sources: Baker 2004; Brown 2004; Foels 2004; Leapfrog website; Medicare Payment Assessment Commission 2005; Rosenthal et al. 2004, 2005; Shermer 2004.

using outcomes measures (perhaps in addition to structure and process) were imposed on providers that did not have the information required to make such programs effective, as is currently happening in Medicare and in many private health plans?² That is, what would happen if payers did not have accurate and complete information about patients' health, given that many factors beyond the provider's control can affect the change in a patient's health? In this situation, some providers would

receive poor performance ratings even when they followed the optimal clinical management regimens and treatments, because their patients did not adhere to the treatments, their patients were sicker initially than they appeared to the risk-adjuster, or the patients experienced unusually bad luck (Hofer *et al.* 1999). Consequently, in some years the performance payments would be large, and in other years they would be small.

Because physicians (and hospitals to a lesser extent) are risk averse (*i.e.*, they value both income/profit and the stability of income/profit), they may begin to demand higher payments from insurers to offset the utility loss from the additional risk (*i.e.*, a risk premium). This has happened in other industries. For example, advertising agencies and consulting firms that offer a reduced fee in exchange for a percentage of cost savings or a percentage of the client's increased sales demand a higher expected payment from their clients (Lawton 2005; Lyons 2004).

In most of today's P4P programs, however, the magnitude of the incentive payments remains small, allocating only 2 to 4 percent of payments for a provider's structures, processes, and/or outcomes (see table 2). As a result, the increase in the provider's risk and the resulting risk premium has been small. But with small incentive payments, physicians' behavior is not likely to change much if they are not offered a "risk" premium.

This situation may change if payers begin to tie a substantial proportion of payments to providers' performance. Mark McClellan, the former administrator of CMS, predicted that by 2009, 20 percent of Medicare payments will be made via pay-for-performance programs (Landro 2004). If this happens, P4P may do more harm than good.

One way to reduce providers' risks while still encouraging them to implement structures and processes believed to improve outcomes is to measure the performance of many physicians' practices or that of a whole hospital system rather than the performance of an individual physician or hospital. Aggregating providers reduces the risk that stems from stochastic health outcomes. Such an aggregation, however, could allow individual physicians to free ride on their colleagues' behavior. Unless physician groups can effectively prevent free riding, perhaps through peer pressure, outcomes-based P4P at the group level may not enhance the quality of care as much as it could at the individual physician level.

The Ideal End State: Pay-for-Performance Is Irrelevant

Pay-for-performance will be effective and helpful in settings with the right combination of health insurers' and providers' information and ignorance. Suppose that external observers (and everyone else) know the ideal health production function and the severity of patients' initial illness. Consider an example often used by Brent James of Intermountain Health, in which certain people are aware of a disease management technique that is highly effective in the clinical management of diabetics at a given risk or severity level. Suppose also that these people have information about the volume of specific services consumed by these patients, like hospital admissions and outpatient visits. What would be the best way to pay providers, and what might be the adverse consequences of alternative ways?

It is easy to see what the best thing is to do. The insurer should pay for managing those patients with the severity levels for which it is appropriate. In addition, the insurer should set payments for other services close to the marginal cost of such care. For diabetes patients at other levels of severity (say, lower) whose disease does not require management, the payment schedule for such services should be zero.

In the face of such payment levels, to belabor the obvious, providers will furnish disease management services and, because other services are priced essentially at breakeven, be willing to supply the appropriate levels of outpatient and hospital care but no more. This is "optimal fee-for-service," in the sense that both what is defined as an insured service and the payment levels for such services are set at "optimal" levels. In this case, paying for performance based on outcomes would be unnecessary and therefore irrelevant. This is similar to Donabedian's point that if there is an established one-to-one causal relationship between a process and an outcome, then either could be used to measure quality (Donabedian 1980).

In James's example, insurers did not pay for the disease management program, and apparently they did overpay for inpatient care. As a result, he reached the discouraging conclusion that the health system that developed the disease management program faced dysfunctional incentives; that is, it would make more money dropping the disease management program and reaping profits on an even higher rate of hospital admissions. What we can say is that compared with a poorly

designed fee-for-service system, outcomes-based P4P might work better here. In theory, though, a perfect version of fee-for-service could be as good as P4P and, if providers are risk averse and health outcomes are to some extent beyond the control of providers' actions, strictly preferable to P4P. What if providers received a bundled payment covering a number of services, but the insurer wants to encourage greater use of only one of those services (e.g., beta blockers for patients hospitalized with congestive heart failure)? In theory, an insurer could implement "optimal fee-for-service" by eliminating the beta blockers from the inpatient case rate payment and paying a separate fee for administering the drug. If this is difficult to implement, the insurer could instead use a P4P program based on the process measure of administering a beta blocker. Our point is that the process-based P4P program does not achieve anything beyond what could be done with fee-for-service, although it may be easier to implement.

What is the kind of information needed to make outcomes-based P4P unnecessary? In the most general sense, it is, literally, evidence-based medicine. If we know what works for whom, we can simply pay for what works for those patients for whom it works and not pay (or at least not overpay) for other things. We are not advocating optimal fee-for-service as a realistic reimbursement system for the near future. Rather, we are using the system to highlight a paradox: if the information required for effective outcomes-based P4P becomes available, do we necessarily need P4P? That is, can we simply make the information publicly available and expect the same kind of automatic emergence of strong incentives as occurs elsewhere in the economy?

The development of the data needed for external or artificial P4P would make the planned implementation of such systems irrelevant; the data alone should be enough. Of course, the current evidence does not provide much confidence that patients will pay attention to outcomes and risk adjustment data. Schneider and Epstein (1998), for example, found that only 12 percent of patients undergoing coronary artery bypass surgery in Pennsylvania were even aware that the state had posted risk-adjusted outcomes data for surgeons and hospitals. Dafny and Dranove (2005) discovered that although consumers did respond to the publication of Medicare HMO report cards, the information they provided was small compared with the amount of information that consumers acquired from other, less formal sources.

Market-Based Insurance and Pay-for-Performance

Up to this point, our discussion of optimal fee-for-service almost seems to advocate a centrally planned system guided by decision makers concerned only about achieving efficiency. But this may not be the best model of the U.S. (or any other) health insurance system. Is there any prospect that something close to the ideal might emerge without such spiritual guidance? There are really two issues here. One is whether “formal” outcomes-based P4P might be expected to emerge in market settings. The other, more fundamental, issue is whether some kinds of insurances might obviate the need for formal P4P.

Let us begin with the second question. In markets for other services, there is no buzz about paying for performance, no conferences, and no governmental studies. However, the absence of explicitly targeted systems does not mean that good performance is not rewarded. Quite the contrary; incentives for good outcomes and good quality emerge as a part of the competitive process, without the need for hand-wringing, price setting, or planning. The story here is fundamentally a simple one: if a business (think of a restaurant or a cleaning service) offers better quality (in the sense of outcomes) than do other similar businesses with which it competes, it will potentially be rewarded for this improved performance in two ways: at the prices prevailing in the market, it will attract more patrons, and it will be able to charge a higher (premium) price than can other firms in the market offering only a standard quality. Moreover, these two influences will reinforce each other: as prices rise to make additional volume more profitable, it will become more profitable to improve quality in order to increase volume even further. No recourse to consultants, analysis, or estimation is required. What is required is that buyers be able to determine when quality advantages exist (through information sources like the *Zagat Guides* and *Consumer Reports*) and that they care enough about higher quality to pay for it.

What is the problem with medical care? There are two parts to the answer. One obvious one is that buyers may not be well informed, perhaps not even well enough informed to know that they ought to care about quality, or how much they should be willing to pay for a certain improvement in quality. The other reason is that prices are not determined in ordinary markets, because of the presence of insurance

that both shields final buyers from exposure to the actual payment made to sellers and inserts the insurer as a potential proxy buyer between the final consumer and the seller.

Solving the information deficit is necessary for formal outcomes-based P4P, as we have already noted. But then we have a paradox: if we have the information needed for good P4P, do we necessarily need P4P? Or can we just make the information publicly available and expect the same kind of automatic emergence of strong incentives as occurs elsewhere in the economy?

The answer to this part of the question depends on how the pricing or reimbursement system actually works. At one extreme, suppose the pricing system is like Medicare, which tries to set administered prices slightly above average costs. Such a system makes it illegal for consumers to offer higher prices than the administered price level even when they are eager to do so in order to get access to providers they regard as especially attractive. This system encourages the purchase of supplemental insurance that mitigates the effect of price variation in consumers' possible cost sharing. When all the ways in which ordinary markets reward higher quality are rendered inoperative, analysts observe—to their great surprise—that higher quality reaps only weak rewards.

At the other extreme, consider medical services that are often not covered by conventional service-benefit insurance, such as plastic surgery, vision-correction surgery, or in-vitro fertilization. Patients who are informed about providers that offer care that produces better outcomes (controlling for initial health status) would be expected to reward those providers with higher prices and/or more volume. Moreover, the amount of the financial reward would in equilibrium be calibrated, “as if by an invisible hand,” to be adequate to produce just that level of improvement in quality whose benefits justify its cost. Tu and May (2007) report that in fact, LASIK vision-correction surgery does have tiered pricing, with discount providers charging about 30 percent less and premium-price providers about 30 percent more than the average provider. Although patients who see premium-price providers have relatively high satisfaction rates, as would be expected, there is evidence that useful price and/or quality information is not readily available to all consumers.

The upshot is that at this level of abstraction, the development of the data needed for external or artificial P4P would make the planned implementation of such systems irrelevant; that the data alone should be enough. Of course, the existing evidence does not provide much

confidence that patients will pay attention to outcomes and risk adjustment data. But then, the existing data have not been all that convincing, so perhaps patients have been wiser than policymakers in ignoring data that can be gamed, quality measures that are limited to process, or outcome measures that are irrelevant to patients' values. More to the point, the information revolution that is supposed to be about to occur in concert with the spread of high-deductible health plans might solve the problem. Even the recent discussion of price transparency takes on greater relevance in a setting in which the buyers' choices might reward high-priced providers if they also furnish higher quality.

When will there be a case for externally designed and imposed outcomes-based P4P systems? One argument would be that the transactions costs of informing buyers constructively about outcomes and severity adjustment are higher than the cost of conveying that information to whatever entity might implement a formal mechanism. It is costly to inform consumers and may be cheaper to inform managers of a P4P system, especially if consumers cannot determine when outcomes improve and the value of those improvements. A second argument is that some insurers (think of Medicare) may not be willing to implement market-based insurance for other reasons, related to politics, and instead trust in the shopping capacity of their enrollees, or market strategy. A third argument is that the incentive to develop good measures of outcome and risk will be stronger if an insurer that uses those measures can keep the benefits for itself (in terms of competitive advantage). That is, if an insurer can develop a proprietary method for identifying high-performing providers and deny other insurers access to the information it generates, the first insurer may be able to earn higher profits. That information is a public good, so that the exclusion of others may be inefficient, but it may be the price that needs to be paid to create incentives to develop and maintain good methods for measuring performance itself. All these arguments are plausible, but none is guaranteed; we would need to explore how they work in practice.

To sum up: in two situations, outcomes-based P4P would be irrelevant. One is when information on the production function is available and definitive, and so an ideal fee-for-service system pays for only the ideal process. The other is when patients can make better use, than can health insurers or a central authority, of the severity-adjusted outcome data that P4P requires.

TABLE 3
Increasing and Diminishing Returns of P4P

Summary

Given these potential problems, can we still make a case for P4P in today's environment? Our answer is a qualified yes as long as P4P is structured intelligently. When the information about patients' health (the left side of table 3) is imperfect, performance payments should be small and targeted whenever possible at the physicians' practices or health system level in order to reduce providers' risk and the resulting upward pressure on costs. In most cases, providers' performance should be measured by patients' outcomes or changes in their outcomes. Although basing P4P payments on process measures may be helpful when specific processes cannot be easily removed from bundled payments, 74 percent of the P4P programs described in table 2 use process measures to identify high-performing providers, and the discussion nationally is focusing on

processes rather than outcomes. If a payer knows that a process improves health, it should increase the corresponding fee until providers prescribe the optimal amount. If a payer is not sure whether a process improves health, it should give providers incentives to make this determination by using outcomes-based P4P. Although the incentives would be modest, P4P can help spur the use of evidence-based medicine, which will in turn make P4P more powerful. In its infancy, P4P is likely to produce small gains in quality.

As knowledge of the health production function and outcomes measurement improves (the middle of table 3), payers can increase fees for clinical management that improves health, reduce fees for clinical management that has no effect or a negative effect on health, and increase the magnitude of outcomes-based performance payments. In this setting, P4P is at its most effective. Payers can accurately measure the change in health of each provider's patients, which reduces the risk that providers face and allows P4P to be targeted at the level of the individual physician or hospital.

If payers ever have complete information on the health production function, outcomes-based P4P would become irrelevant (the right side of table 3). Rather than basing payments on changes in outcomes, payers could implement "optimal fee-for-service," thereby increasing fees for clinical management regimens and treatments that are known to work and reducing (possibly to zero) fees for clinical management regimens and treatments that are known not to work. An entity that is able to evaluate the efficacy of processes and disseminate this information persuasively to providers would speed up the evolution to optimal fee-for-service. While quality would certainly improve, costs may go up or down depending on how much of the medical care currently being provided proves (once the evidence arrives) to increase costs without improving quality, how expensive it is to improve quality, and the value that consumers place on quality improvements. In all three conditions, however, insurers may not allocate the amount of money for quality improvement that consumers want.

Conclusions

Pay-for-performance is one of a number of strategies being employed today with the goal of enhancing the quality and efficiency of health

care delivery in the United States. Given the current state of affairs—its tendency to “underuse” measures and the relatively small number of incentives—P4P can play a constructive role in accelerating quality improvement efforts. As many advocates have noted, “It is a good place to start.”

P4P is not a panacea, however, and expectations must be realistic. P4P cannot address fundamental problems of our current health care system, such as coverage for the uninsured, quality improvement for individuals who choose not to seek care (e.g., because high deductibles may mean they effectively do not have coverage), and the high cost of administration. As the number of incentives increases, the potential for unintended negative consequences rises as well, such as the disproportionate effort placed on services that are measured and the increased risk for providers. Balancing this will be a more effective P4P as we learn more about how the health care system actually works.

In any case, it is unlikely that P4P will play a significant role in addressing the “overuse” and “misuse” of various clinical management regimens and treatments in the near future. Clearly, a comprehensive quality improvement agenda will require P4P to be combined with other strategies. These may include more stringent criteria for recertifying physicians, public reporting of provider performance at various levels, as well as the broader application of case management and disease management. Quality improvement efforts have gained momentum only in the last fifteen or so years, as marked by the creation of the Agency for Healthcare Research and Quality (AHRQ) and the National Committee for Quality Assurance (NCQA). These still are in their infancy, with a long way yet to go.

Endnotes

1. The Premier hospitals participating in the Medicare P4P demonstration program improved the quality of their care, relative to that of a group of similar hospitals, by between 2.6 and 4.1 percent in four clinical areas in the first two years of the program (Lindenauer et al. 2007). Partners Community HealthCare Inc. improved on three diabetes quality measures relative to the rest of Massachusetts after it instituted a P4P program, experienced a decline on one diabetes measure, but had no differential change for asthma (Levin-Scherz, DeVita, and Timbie 2006). After PacifiCare implemented a P4P program in California, physicians' practices in the state improved cervical cancer screening relative to practices in the Pacific Northwest that were not eligible for the program, but there was no difference between California and the Pacific Northwest in mammography screening or HbA1c testing for diabetics (Rosenthal et al. 2005). In a recent comprehensive literature review, Christianson, Leatherman, and Sutherland (2008,

- p. 7) conclude that “the findings from studies on the effect of payer initiatives that reward providers for quality improvements or the attainment of quality benchmarks are mixed.”
2. In 2005, the Medicare Payment Assessment Commission recommended the use of process, structure, and outcome measures for hospital care; outcome measures for home health care; and structure and process measures for physicians’ services (Medicare Payment Assessment Commission 2005). This recommendation is consistent with recent private health plan P4P initiatives that use a mix of process, structure, and patient satisfaction measures (Rosenthal et al. 2004).

References

- Adams, D. 2004. *Strategies to Support Quality-Based Purchasing: A Review of the Evidence*. AHRQ Technical Appendix, AHRQ publication no. 04-0057.
- Baker, G. 2004. The Evolution of Pay-for-Performance Models for Rewarding Providers. In *Case Studies in Health Plan Pay-for-Performance Programs*, edited by J. Gutman, Introduction, pp. 1–13. Washington, D.C.: Atlantic Information Services.
- Bradley, E.H., J. Herrin, B. Elbel, R.L. McNamara, D.J. Magid, B.K. Nallamothu, Y. Wang, S-L.T. Normand, J.A. Spertus, and H.M. Krumholz. 2006. Hospital Quality for Acute Myocardial Infarction. *Journal of the American Medical Association* 296(1):72–78.
- Brown, J. 2004. Blue Cross of California’s Physician Incentive Programs. In *Case Studies in Health Plan Pay-for-Performance Programs*, edited by J. Gutman, Chapter 7, pp. 105–115. Washington, D.C.: Atlantic Information Services.
- Chassin, M.R. 2006. Does Paying for Performance Improve the Quality of Health Care? *Medical Care Research and Review* 63(1):122S–25S.
- Christianson, J., S. Leatherman, and K. Sutherland. 2008. *Financial Incentives, Healthcare Providers and Quality Improvements*. London: Health Foundation.
- Dafny, L., and D. Dranove. 2005. Do Report Cards Tell Consumers Anything They Don’t Already Know? The Case of Medicare HMOs, working paper #11420. Cambridge, Mass.: National Bureau of Economic Research.
- Donabedian, A. 1980. *The Definition of Quality and Approaches to Its Assessment*. Ann Arbor, Mich.: Health Administration Press.
- Foels, T. 2004. Second Generation Pay-for-Performance: Comprehensive Chronic Disease Profiling. In *Case Studies in Health Plan Pay-for-Performance Programs*, edited by J. Gutman, Chapter 4, pp. 63–75. Washington, D.C.: Atlantic Information Services.

- Fonarow, G.C., W.T. Abraham, N.M. Albert, W.G. Stough, M. Gheorghiadu, B.H. Greenberg, C.M. O'Connor, K. Pieper, J.L. Sun, C. Yancy, and J.B. Young. 2007. Association between Performance Measures and Clinical Outcomes for Patients Hospitalized with Heart Failure. *Journal of the American Medical Association* 297(1):61–70.
- Hofer, T.P., R.A. Hayward, S. Greenfield, E.H. Wagner, S.H. Kaplan, and W.G. Manning. 1999. The Unreliability of Individual Physician "Report Cards" for Assessing the Costs and Quality of Care of a Chronic Disease. *Journal of the American Medical Association* 281(22):2098–2105.
- Institute of Medicine. 2006. *Performance Measurement: Accelerating Improvement*. Washington, D.C.: National Academies Press.
- Jha, A.K., E.J. Orav, Z. Li, and A.M. Epstein. 2007. The Inverse Relationship between Mortality Rates and Performance in the Hospital Quality Alliance Measures. *Health Affairs* 26(4):1104–17.
- Landro, L. 2004. Booster Shot: To Get Doctors to Do Better, Health Plans Try Cash Bonuses. *Wall Street Journal*, September 17.
- Lawton, C. 2005. Small Ad Agencies Turn to Revenue-Sharing Deals. *Wall Street Journal*, September 19.
- The Leapfrog Group website: <http://www.leapfroggroup.org/> (accessed June 16, 2008).
- Levin-Scherz, J., N. DeVita, and J. Timbie. 2006. Impact of Pay-for-Performance Contracts and Network Registry on Diabetes and Asthma HEDIS Measures in an Integrated Delivery Network. *Medical Care Research and Review* 63(1):14S–28S.
- Lindenauer, P.K., D. Remus, S. Roman, M.B. Rothberg, E.M. Benjamin, A. Ma, and D.W. Bratzler. 2007. Public Reporting and Pay for Performance in Hospital Quality Improvement. *New England Journal of Medicine* 356(5):486–96.
- Lyons, D. 2004. Skin in the Game. *Forbes* 173(3):78–80.
- Medicare Payment Assessment Commission. 2004. *March 2004 Report to Congress*. Washington, D.C.
- Medicare Payment Assessment Commission. 2005. Strategies to Improve Care: Pay for Performance and Information Technology. In *March 2005 Report to Congress*. Washington, D.C.
- Rosenthal, M.B., R. Fernandopulle, H.R. Song, and B. Landon. 2004. Paying for Quality: Providers' Incentives for Quality Improvement. *Health Affairs* 23(2):127–41.
- Rosenthal, M.B., R.G. Frank, Z. Li, and A.M. Epstein. 2005. Early Experience with Pay-for-Performance. *Journal of the American Medical Association* 294(14):1788–93.

- Rosenthal, M.B., S-L.T. Normand, R.G. Frank, and A.M. Epstein. 2006. Pay for Performance in Commercial HMOs. *New England Journal of Medicine* 355:1895–1902.
- Schneider, E.C., and A.M. Epstein. 1998. Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery. *Journal of the American Medical Association* 279(20):1638–42.
- Shermer, J. 2004. Blue Cross & Blue Shield of Michigan Participating Hospital Agreement Incentive System. In *Case Studies in Health Plan Pay-for-Performance Programs*, edited by J. Gutman, Chapter 5, pp. 77–90. Washington, D.C.: Atlantic Information Services.
- Tu, H.T., and J.H. May. 2007. Self-Pay Markets in Health Care: Consumer Nirvana or Caveat Emptor? *Health Affairs Web Exclusive*, February, w217–w226.
- Werner, R.M., and E.T. Bradlow. 2006. Relationship between Medicare's Hospital Compare Performance Measures and Mortality Rates. *Journal of the American Medical Association* 296(22):2694–2702.

Acknowledgments: Funding for this project was provided by Merck & Co., Inc.