

# Pay for Performance in Health Care: Methods and Approaches

Edited by

Jerry Cromwell, Michael G. Trisolini, Gregory C. Pope,  
Janet B. Mitchell, and Leslie M. Greenwald

March 2011

**RTI** Press

The logo for RTI International, featuring a stylized 'R' and 'T' inside a square, followed by the text 'RTI' in a large, bold, serif font, and 'INTERNATIONAL' in a smaller, all-caps, sans-serif font below it.

©2011 Research Triangle Institute.  
RTI International is a trade name of Research  
Triangle Institute.

All rights reserved. Please note that this  
document is copyrighted and credit must be  
provided to the authors and source of the  
document when you quote from it. You must  
not sell the document or make a profit from  
reproducing it.

Library of Congress Control Number: 2011921923  
ISBN 978-1-934831-04-5

**doi:10.3768/rtipress.2011.bk.0002.1103**  
**[www.rti.org/rtipress](http://www.rti.org/rtipress)**

**Suggested Citation**

Cromwell, J., Trisolini, M. G., Pope, G. C., Mitchell,  
J. B., and Greenwald, L. M., Eds. (2011). *Pay  
for Performance in Health Care: Methods  
and Approaches*. RTI Press publication  
No. BK-0002-1103. Research Triangle Park, NC:  
RTI Press. Retrieved [date] from  
<http://www.rti.org/rtipress>.

This publication is part of the RTI Press Book series.

RTI International  
3040 Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709-2194 USA  
[rtipress@rti.org](mailto:rtipress@rti.org)  
**[www.rti.org](http://www.rti.org)**

# Overview of Pay for Performance Models and Issues

Gregory C. Pope

For the purposes of this chapter, we define “pay for performance” (P4P) as *a set of performance indicators linked to an incentive scheme*. The performance indicators are the *performance* component of P4P, and the incentive scheme is the *pay* component. In health care, P4P contrasts with traditional fee-for-service (FFS) payment, which pays for quantity of services without regard to performance.

This chapter considers the elements that go into designing P4P systems. A very large number of specific P4P schemes can be formed from various combinations of the elements described in this chapter. Given the lack of compelling evidence for particular approaches, payers have experimented with many different approaches. P4P encompasses a large range of real-world programs that have not yet coalesced into a small number of accepted models. All P4P programs, however, are based on decisions about a common set of design elements.

This chapter presents measures of performance and the incentive schemes that payers (e.g., health plans or government programs such as Medicare or Medicaid) may attach to performance measurement. We identify the limits of the P4P model and offer alternative ways to reach the same goals. For concreteness and simplicity, throughout this chapter we focus mostly on situations in which payers apply incentives to health care provider organizations (including group practices, hospitals, and integrated delivery systems) and physicians or other clinicians. Payers, health plan sponsors, and policy makers can apply many of the same principles and even specific approaches in other situations (e.g., employers or the government giving incentives to health plans).

## Measuring Performance

P4P systems attempt to reward explicitly measured dimensions of performance. Performance measurement consists of several components: defining domains of performance, selecting domains to be measured, selecting indicators to measure each domain of performance, defining the unit for performance measurement and accountability, choosing data sources for measuring performance, and deciding whether participation will be voluntary or mandatory.

## Defining Domains of Performance

The first crucial step in designing a P4P system is defining the domains or dimensions of performance that the program might reward. In health care, performance domains might include clinical outcomes, clinical process quality, patient safety, access to and availability of care, service quality, patient experience or satisfaction, cost efficiency or cost of care, cost-effectiveness, adherence to evidence-based medical practice, productivity, administrative efficiency and compliance, adoption of information technology, reporting of performance indicators, and participation in performance-enhancing activities. We discuss these in turn below.

**Clinical outcomes.** The ultimate goal of health care is to maintain or improve patient health status. Clinical outcomes are, therefore, a desired performance domain. Outcome measures include mortality, morbidity, functional status, quality of life and quality-adjusted life years (QALYs), and avoidance of acute exacerbations of chronic conditions. However, using outcomes to measure quality faces challenges (Eddy, 1998). Some outcomes, such as mortality, are rare or observed only with a long time lag. Outcomes such as functional status can be expensive to measure in large populations. Also, outcomes can be influenced by many factors, and some important ones (e.g., patient adherence to recommended care) may be outside of physicians' control.

**Clinical process quality.** Given the limitations in using clinical outcomes to judge performance, process measures are currently the most widespread method that evaluators use to assess clinical quality. Examples of process measures include eye examinations, lipid tests for patients with diabetes, and mammograms for women in certain age groups. Compared with outcomes measures, process measures are often frequent and controllable. In recent years, the efforts of several national bodies—including the US Agency for Healthcare Research and Quality, the National Committee for Quality Assurance, the American Medical Association, and the National Quality Forum—have

substantially increased the number of available clinical guidelines and detailed quality process measure specifications. However, quality measurement in health care is not as straightforward as one might hope. Professional organizations, policy makers, and regulating bodies often base clinical guidelines and quality measures more on expert opinions than on the results of randomized controlled trials. While process quality measures may still be appropriate in many cases, there are not always well-established linkages between process quality measures and final outcomes of interest (see Chapter 4 for more on quality measures).

**Patient safety.** Reports of the large numbers of patients injured by medical care have stimulated interest in improving patient safety by reducing medical errors (Kohn et al., 1999). An example of a patient safety performance measure is the rate of hand washing among hospital patient care employees (a higher rate of washing reduces the rate of patient infections).

**Access and availability of care.** Measuring enrollee access to care may be especially important in settings, such as capitated payment systems, that have incentives to withhold services. The health plan, which controls the benefit design and provider network, is often a natural unit for measuring access.

**Service quality.** “Service quality” refers to nonclinical aspects of the patient experience that may be valuable to patients. Service quality can include such factors as patient waiting time to see physicians, patient telephone or e-mail access to provider organizations, convenience and length of office hours, and so forth.

**Patient experience or satisfaction.** Patient reports, which researchers usually obtain from patient surveys, provide evidence of provider organization or physician performance from the point of view of the patients who receive medical care. Typical domains include how individual physicians are rated for attributes such as communication; whether patients have difficulty getting referrals, tests, or care; whether patients receive needed care; whether patients receive care quickly; how well physicians communicate; how good physicians’ customer service is; and how provider organizations and physicians submit and process claims. The number of existing patients who have changed doctors or new patients who have selected doctors can also be used to infer patient experience or preferences.

**Cost efficiency or cost of care.** Cost efficiency refers to the cost of providing a given level of quality of care or health outcome. Together with the quality of care, cost efficiency defines the value of care (see Chapter 5 for more about efficiency). An example of a cost efficiency measure is the cost of producing

an extra QALY. Cost of care is the cost of producing an intermediate health care services output. Examples of cost of care measures include the rate of prescribing generic drugs by a physician or within a health plan, hospital days per 1,000 health plan enrollees, case mix–adjusted hospital average length of stay, and cost per episode of care. Because cost of care measures are much easier to quantify than cost efficiency measures, they tend to be much more prevalent than the latter measures (Hussey et al., 2009).

**Cost-effectiveness.** Cost-effectiveness refers to the relative cost of alternative interventions that produce desired outcomes such as improvement in health (e.g., QALYs). To reduce continued increases in health care costs, P4P programs may provide incentives for more cost-effective medical treatment patterns. For example, P4P might reward physicians who order fewer expensive diagnostic imaging tests that are not considered medically necessary by clinical practice guidelines.

**Adherence to evidence-based medical practice.** Medical practice encompasses many practice styles, some of which do not rely on evidence-based standards of care (Wennberg et al., 2004). Adhering to evidence-based standards of care may enhance physicians' quality and efficiency. P4P may reward physicians for following clinical practice guidelines in their treatment of patients (e.g., following an evidence-based decision algorithm when deciding on ordering advanced imaging tests for low-back pain).

**Productivity.** Productivity refers to the amount of output per unit input. Payers may wish to measure and explicitly reward productivity in situations in which base compensation for physicians or provider organizations is not tightly tied to work effort and generated output. For example, if physicians are salaried, a payer may want to find a way to reward productivity to stimulate work effort, efficiency, and provided services.

**Administrative efficiency and compliance.** Administrative compliance refers to performance outside the clinical and patient domains on indicators that may be relevant to payers. For example, a health plan might want to reward provider organizations and physicians based on their electronic submission of claims (invoices) for medical treatment, timely submission of claims, and low error rates in claims submission.

**Adoption of information technology.** Most payers consider measuring information technology (IT) critical to improving the coordination, quality, and efficiency of care. For example, payers might reward organizations and physicians based on physicians' use of electronic software to order prescriptions for their patients. This use may both lower costs and improve quality by reducing medication errors.

**Reporting of performance indicators.** Especially early in the implementation of a P4P system, complete reporting of requested performance indicators may be an important measure of performance. For P4P to be comprehensive, fair, and equitable, provider organizations and physicians must report performance indicators frequently and accurately. “Pay for reporting” is a first step toward improving the data to which payers apply incentives.

**Participation in performance-enhancing activities.** Payers may provide incentives for physicians to participate in performance-enhancing activities. Participation in such activities could include attending collaborative quality-improvement workgroup meetings and developing quality improvement action plans. The limitation of this “pay for participation” is that payers can measure the fact that participation occurred but not the performance outcomes of participation.

### **Selecting Performance Domains for Measurement**

Some P4P systems may be comprehensive and include many domains of performance; others may focus on only a single domain. Payers may implement systems in stages, starting with a single domain and gradually adding others. Payers may determine domains, specific performance indicators, and the relative size of rewards by considering numerous variables: the importance of individual domains; the goals of the program; the availability of meaningful measures; the potential for clinical improvement; existing problem areas; and cost, burden, and data availability (Dudley & Rosenthal, 2006; Sorbero et al., 2006).

**Importance and goals.** Some domains may be more important to the priorities of the sponsor of the P4P system (e.g., a health plan) or to its members or clients (e.g., enrollees) than others. Many P4P programs focus on clinical quality of care. For example, California’s Integrated Healthcare Association (IHA) P4P program weights clinical quality at 50 percent of total performance (McDermott & Williams, 2006). As another case in point: in six Rewarding Results demonstration sites, the weight on clinical quality ranges from 40 to 100 percent (Young et al., 2007). Patient satisfaction is also often weighted heavily; for example, the IHA program weights it at 30 percent. Early in the implementation phase, to facilitate implementation of the system, programs may place more weight on adopting IT and reporting systems and on reporting performance indicators. The California IHA program weights IT at 20 percent.

**Availability of meaningful performance measures.** The availability of meaningful (reliable, valid, and significant) performance measures varies across domains. Performance in domains for which a larger number of meaningful measures is available is likely to be assessed more accurately, facilitating inclusion of these domains in P4P programs.

**Potential for improvement.** Payers have fewer reasons to focus on domains in which performance is difficult to improve (e.g., domains in which performance is already high) than on domains in which the need and potential for improvement are substantial. A wide range of performance in a domain may indicate that it has considerable potential for improvement.

**Current problems or areas of poor performance.** P4P programs may emphasize areas in which current performance is poor or needs improvement. Focusing measurement and incentives on problem areas can lead to improvements in these areas.

**Cost, burden, and data availability.** P4P programs are more likely to include domains for which data are available or can be generated at low cost without undue burden on providers or health plan enrollees. For example, domains with measures for which programs can obtain data from existing computerized administrative data systems or health insurance claims are typically easier to implement than domains with measures that require new methods of medical chart abstraction or patient surveys.

### **Selecting Performance Indicators for Measured Domains**

Once P4P programs choose the domains they will include in their systems, they need to specify indicators of performance for each of these domains. Good performance indicators should be valid, reliable, important, relevant, specific, controllable, actionable, efficient, and cost-effective.

**Validity.** The indicators should be valid indicators of the performance dimension that they purport to measure. Programs may choose indicators that have been peer reviewed and endorsed by a national accreditation organization (for example, the National Quality Forum). If programs use process-of-care indicators, the indicators should be linked with the ultimate outcome of interest (e.g., patient mortality, morbidity, or functional status).

**Reliability.** The indicators should be reported as consistently as possible across participants and across time. The sample size of patients that the indicators use should be large enough for statistically reliable calculation of rates. The data underlying the measurement process should be reliable. Physicians may dislike P4P programs that they feel do not measure their performance accurately.

**Importance and relevance.** Indicators should measure an important or relevant aspect of the performance domain to which they correspond. An outcome indicator, for example, should measure a significant aspect of patient health, such as mortality or functional status, and there should be evidence that physicians' actions can appreciably affect it. A process indicator should measure a process that has a demonstrable link to health outcomes of interest and that is under the control of physicians.

**Specificity and controllability.** The indicators should be specific to the performance domains they measure. They also should be specific to factors under the control of the entity whose performance is being measured. Indicators should match accountability with control. For provider organizations and physicians, one advantage of process measures over outcome measures is that process measures often measure factors under the direct control of provider organizations and physicians, whereas patient and other characteristics may affect outcomes measures in ways that are difficult to adjust for.

**Actionability.** The indicators should provide information that provider organizations and physicians can act upon to improve performance.

**Efficiency.** The indicator set should be the smallest possible that is still broad enough to cover the performance domain. Too many quality measures may impose excessive data collection costs on provider organizations and physicians, and the sheer number of measures may cause a lack of focus in quality improvement activities. On the one hand, with many indicators, the potential reward from improving performance for any one indicator may be too small to justify the investment in doing so. On the other hand, having too few performance measures creates the risk that provider organizations or physicians will focus too narrowly on the selected measures while ignoring other dimensions that are important for overall performance in a domain.

**Cost-effectiveness and cost benefit.** P4P programs prefer indicators that have greater expected benefit of improved performance relative to their costs of collection and compliance. It should be possible to improve, collect, and report indicators in a cost-effective manner. The data needed to calculate the indicator should be available at a reasonable cost. The cost of complying with and reporting a performance indicator should correspond with the expected benefits of improved performance on the measure.

The availability of indicators that score highly on these criteria may vary greatly across performance domains and across particular settings. Hence, at the current time, implementing P4P programs that emphasize certain domains

(e.g., clinical quality) rather than others (e.g., cost efficiency) may be more feasible.

### **Defining the Unit for Performance Measurement and Accountability**

P4P systems differ in whose performance is measured. Performance may be measured for any or all of the following: provider organizations and physicians, disease management companies and other third-party care management organizations, and health plans. We discuss the issues for each of these target units of analysis below.

**Provider organizations and physicians.** Most commonly, health care P4P systems apply directly to provider organizations and physicians. These entities directly deliver services; therefore, they have the most direct control over important aspects of performance such as clinical quality. The provider organizations and physician entities that are held accountable in P4P programs may be classified into three broad categories.

- **Institutional providers.** Institutional providers include hospitals, nursing facilities, and home health agencies. Institutions are important targets of P4P for several reasons. First, a large percentage of health care spending occurs in institutions. Second, institutions are often large organizations with considerable resources. They are more likely to have sophisticated information systems that can capture and report performance measures. Also, they are more likely to have the management systems and organizational structures to respond to incentives to improve performance. Third, institutions typically treat large patient populations. Thus, events (e.g., treated patients) that are eligible for performance measurement occur frequently and allow statistically reliable and valid measurement of performance. Fourth, institutions facilitate attributing responsibility for care. For example, one and only one hospital is responsible for a given hospital stay.<sup>1</sup> An example of a P4P program in which institutions are the unit of accountability is Medicare's Premier Hospital Quality Incentive Demonstration, which rewards or penalizes hospitals for their performance on selected inpatient quality measures.
- **Physicians and other clinicians.** Clinicians, particularly physicians, control most health care spending because they make the decisions about whether to order or authorize care. For this reason, P4P programs tend

---

<sup>1</sup> This assertion assumes that a patient is not transferred from one hospital to another. Quality measures often exclude such cases.

to measure the performance of physicians and other clinicians. P4P programs often focus on the performance of primary care physicians (PCPs) because the PCPs may be responsible for managing patients' overall care. The United Kingdom's General Medical Services Contract, for example, rewards PCPs for their performance on 146 performance measures (Doran et al., 2006). Although fewer performance measures exist for specialist physicians, those physicians are also important because they control a considerable portion of health care spending, including many high-cost and possibly discretionary services. Payers sometimes hold specialists accountable for episodes of specialty care beginning with primary care referral or first contact with a patient.

Patients often receive treatment from several different physicians. Assigning responsibility to particular physicians is a problem for open access insurance arrangements that do not require enrollees to select primary care gatekeepers. Some P4P systems allow multiple physicians to earn incentive payments (e.g., all physicians who provided at least one or two primary care visits for the patient). These are often termed the *one-touch* or *two-touch* rules for assignment. Other P4P systems require a plurality of primary care visits to determine which provider is assigned performance accountability. Managed care systems or medical home systems that require patients to select an accountable PCP at the time of enrollment avoid this problem, at least on the primary care level. See Chapter 7 for more on patient attribution to physicians or organizations.

Ideally, programs should measure the performance of the individual physicians who provide care to particular patients. Alternatively, programs can measure the performance of the physician group. The group has the advantages of larger sample size and greater statistical reliability for performance measurement, and it may also have organizational mechanisms to provide feedback to individual physicians. Further, rewarding groups of providers—including support staff—emphasizes interdependence and team delivery of health care (Young & Conrad, 2007). Measuring the physician group also reduces concerns about determining accountability among multiple physicians who may be treating a patient because some or all of those physicians may practice in the same group.

For these reasons, P4P programs often focus on physician groups for patient assignment and performance accountability measures (Christianson et al., 2006; Rosenthal et al., 2006). For example, the

Medicare Physician Group Practice Demonstration requires a plurality of visits for assignment, but it assigns patients to groups, not to individual physicians (Kautter et al., 2007). Physician groups may be traditional integrated group practices, other physician organizations such as independent practice associations, or virtual groups (e.g., hospital medical staff or all physicians practicing in a geographic area) established for the explicit purpose of performance measurement.

- **Integrated delivery systems and other combinations of providers.** Payers may evaluate integrated delivery systems (IDSs), physician-hospital organizations, or other organizational forms that combine provider types on both professional and institutional components of performance. Measurement at the level of the IDS allows payers to attribute larger bundles of care, such as episodes, to the provider units they are profiling. Moreover, measurement at this level recognizes and incentivizes the coordination of care across multiple provider types.

**Disease management companies and other third-party care management organizations.** In some P4P models, payers may hold a third party, outside of provider organizations, responsible for performance (e.g., the quality or efficiency of care). For example, the Medicare Health Support Pilot Program holds third-party disease management organizations (e.g., for-profit organizations that payers hire to monitor patients' chronic conditions) accountable for aspects of the care provided to Medicare beneficiaries enrolled in the traditional FFS program. Third-party organizations have certain advantages over provider organizations in achieving performance objectives. They can exploit economies of scale in developing and implementing specialized disease management programs; because they serve large populations for relatively little cost, third-party organizations perform well on cost-effectiveness and cost efficiency. Also, unlike health care providers, third-party organizations do not face the disincentive of foregone revenues when they reduce their clients' use of health care services. However, because third-party organizations do not provide care directly, they must establish mechanisms to gain the cooperation of and influence the behavior of patients and physicians.

**Health plans.** Health insurance plans are a natural unit for performance measurement because they are responsible for arranging all care for covered conditions and services for enrolled members. Because members are enrolled in health plans, the health plans are clearly responsible for their care.

Many individuals have a choice of multiple health plans. The availability of comparative information about the quality and efficiency of health plans may aid individuals in choosing health plans. Employers, governments, or other health plan sponsors can establish incentives at the health plan level and then let plans be responsible for transmitting the incentives to downstream provider organizations.

The health plan is an aggregated level of measurement that is far removed from the individual physicians who treat patients. Given heterogeneity among physicians in a plan network, an individual who enrolls in a plan that has a certain rated performance may receive care that deviates substantially from this average, depending on the particular physician who supplies treatment. Hence, performance measurement at the health plan level does not obviate the need for measurement at disaggregated levels (e.g., at the level of the individual physician).

### **Data Sources for Measuring Performance**

Depending on the domain or specific indicator, a variety of data sources may be used to measure performance in P4P programs. The central clinical quality-of-care domain is typically measured by one or more of three data sources: administrative claims, medical records, and patient surveys. Claims data are useful for some types of quality measures that are consistently and reliably recorded in those data and that are used primarily for billing by provider organizations reimbursed through FFS. For example, one measure that payers commonly derive from claims data is whether patients with diabetes have had annual HbA1c testing. An important benefit of claims data is that no additional data collection burden is placed on provider organizations and physicians because the data have already been submitted to payers for billing purposes. This benefit does not exist for providers who are reimbursed by capitation, however, because they often do not submit claims for individual visits, hospital admissions, or other types of medical services. Claims may also have reporting lags (e.g., when pharmacy data are held by contracted pharmacy benefit managers and the data are not easily available to health plans or payers) (Young & Conrad, 2007). Another limitation of claims data is that they contain a restricted range of clinical information.

Medical records are generally superior to claims for determining more clinically detailed quality measures. The high cost of manual medical records data collection is often viewed as a barrier, however. Widespread adoption of electronic medical records (EMRs) might mitigate this concern, although

the dissemination and use of EMRs remains limited in medical practice. The lack of standardization across medical records is also an issue in judging performance.

Patient surveys are useful for some types of data, such as patient satisfaction or patient experience of care measures that cannot be collected from other sources. However, provider organizations and physicians may be concerned about whether patients are able to report accurately on technical aspects of medical care. Also, patient surveys can be expensive and a burden to patients, and they can suffer from low response rates and nonresponse bias.

### **Mandatory Versus Voluntary Participation in Pay for Performance**

Voluntary programs are easier to implement than mandatory programs, and provider organizations and physicians are less likely to resist the implementation of voluntary programs than required programs. Moreover, providers who expect to do well are more likely to participate in voluntary programs than providers who do not expect to do well. The lack of participation by poor performers may limit the ability of voluntary programs to improve overall system performance.

In programs that offer bonuses for good performance, voluntary participation may lead to the same results as a mandatory program because provider organizations and physicians may choose to participate based on the likelihood of earning a bonus. Thus, if a program is voluntary, the sponsor must offer incentives that lead at least some provider organizations and physicians to want to participate because they expect that the rewards they can achieve under the program will exceed the costs of participating. P4P programs that are less favorable to providers (e.g., those that involve penalties for poor performance or downside financial risk) may need to be mandatory. A strategy that payers may use is to start with a voluntary program to demonstrate feasibility and work out operational problems and then gradually increase the penalties for nonparticipation or eventually mandate participation as a condition of eligibility for receiving any reimbursement.

### **Incentive Schemes to Reward Performance**

Given a measurement of performance, an incentive scheme to reward good performance (or penalize bad performance) is the second crucial ingredient of a P4P program. This section discusses the elements of P4P incentive schemes. We consider how to fund incentive payments and to structure financial

incentives. As well as discussing direct incentives to provider organizations and physicians, we also address financial and nonfinancial incentives that programs can offer patients for using high-performing providers.

### **Funding of Performance Payments**

P4P systems must identify a source of funding for the performance incentives. According to the Institute of Medicine Board on Health Care Services (2007), three possibilities are existing payments, generated savings, and new money

**Existing payments.** Redistributing existing payments is attractive to payers because they do not have to add money to the system. One justification for using existing payments to reward quality is that payers already expect high-quality care, and they should not have to add new money to payments to expect provider organizations and physicians to supply high-quality care. However, this approach inevitably means that low-quality providers will receive lower payments than before.

**Generated savings.** Generated savings are those produced by high performance. Payers often claim that improving quality of care (e.g., reducing medical errors and complications of care) will generate savings. Generated savings are also attractive to payers because they do not require new money and because savings are a prerequisite for any performance payments. Basing performance payments on generated savings favors efficiency improvements and quality enhancements that generate savings.

**New money.** Provider organizations and physicians may justifiably argue that performance payments should be funded out of new money if improving and reporting performance requires new investments and higher costs on their part. For example, improving and reporting performance may require providers to invest in expensive IT equipment and training and to hire additional support personnel (e.g., nurse case managers and IT support workers). However, adding new money to the system raises the question of how cost-effective the program is. The question that arises is whether the performance gains that will result from the system are worth its extra cost.

### **Performance Benchmarks**

P4P programs must establish benchmarks against which performance is judged and that will trigger performance payments. The benchmarks that programs choose can significantly affect the amount of P4P performance payments and the extent to which P4P schemes reward high quality or improvements in quality. The choice of a benchmark is thus a critical decision that each P4P program should tailor to its goals (Werner & Dudley, 2009). P4P programs

have three possible benchmarks for rewarding performance—absolute performance, improved performance, and relative performance—discussed below.

**Absolute performance (target attainment).** Some performance indicators may have natural benchmarks. For example, for clinical process-of-care indicators and practice guidelines, payers expect that every patient satisfying the relevant eligibility criteria should receive the indicated service (and/or not receive an obsolete or contraindicated service). The natural benchmark and goal for such indicators is 100 percent compliance or performance. In real-world situations, 100 percent compliance is unlikely because of patient refusal and other factors, but payers may establish a high absolute threshold or target for rewarding performance (e.g., 90 percent). A target provides a clear, simple, direct standard of expected performance.

Target attainment tends to reward existing high performance, not necessarily improvement. Provider organizations and physicians that exceed the target at baseline can enjoy performance payments without improving their performance, although they must maintain it. Conversely, providers with low performance may see high targets as unattainable and may not attempt to improve. Therefore, if the goal of P4P is to improve overall system performance, not merely to reward current high performers, absolute thresholds have drawbacks (Rosenthal et al., 2005).

**Improved performance.** Explicitly rewarding improved performance focuses P4P on improving overall system performance, rather than just rewarding existing high performers. Both low and high performers are rewarded only if they improve compared with their past performance. However, if payers only reward improvement, then low-performing provider organizations and physicians may find it easier than their high-performing counterparts to earn performance payments because improving from a low rather than a high starting point is easier. Giving greater rewards to low-performing provider organizations and physicians, even if they are improving, may lack face validity and appear inequitable to high performers. One way of ameliorating these concerns would be to phase out rewards based on improvement after some period of time (Institute of Medicine [IOM] Board on Health Care Services, 2007), under the logic that providers should be able to transition to a high absolute level of performance within a limited period of time.

**Relative performance.** A third approach is to reward relative performance. In this approach, payers identify a comparison group for the participating provider organizations and physicians. An advantage of rewarding relative

performance is that the comparison group defines the performance benchmark, relieving the program designers of the need to choose a particular reward threshold and adjust it over time. If the general level of performance improves over time, the performance benchmark automatically adjusts upward. Moreover, payers can define regional or local comparison groups, customizing the benchmark to local conditions and baseline performance. Payers can risk adjust comparisons among groups to standardize for differences in group composition in at least two ways.

First, payers can use a usual care comparison group. One variant of relative incentives defines a comparison group of provider organizations and physicians who are not participating in the P4P program. For example, several of Medicare's FFS P4P demonstrations (discussed in detail in Chapter 9) compare the performance of participating providers to that of nonparticipating providers as representatives of the usual standard of care. An advantage of a usual-care comparison group is that the payer can potentially reward all participating provider organizations and physicians, if they all exceed the performance of their (nonparticipating) usual-care comparison group. In a relative ranking approach (discussed next), payers reward only the top performers among the participants. If P4P participation is voluntary, payers may benefit from creating the potential for all participating providers to earn a reward. A usual-care performance standard presents a more feasible improvement target for low-performing providers than high absolute and relative performance criteria do, but such a standard does not reward below-average performance. Identifying a nonparticipating comparison group may not be feasible in all situations. If a program is extended to all providers to maximize its impact, no nonparticipating providers will exist. Then the only feasible comparison group is other participating providers.

Second, payers can use an approach based on a relative ranking of provider organizations. Sometimes called the "tournament" approach or "yardstick competition," this variant ranks participating provider organizations and physicians and rewards only those in the top ranks. Medicare's Premier Hospital Quality Incentive Demonstration uses this method, rewarding only hospitals in the top two deciles of quality performance (2 percent payment bonus for the top decile and 1 percent bonus for the second decile). This approach does not consider absolute performance. Thus, high-ranked provider organizations and physicians will be rewarded, even if their absolute performance is poor, and low-ranked providers will not be rewarded, even if their absolute performance is good. The tournament approach provides

the greatest incentive for improvement to the provider organizations and physicians who are near the threshold that defines the top-ranked providers. Provider organizations and physicians who are already top performers or poor performers have less incentive to improve. Top performers can simply maintain their current relative performance, and poor performers may have difficulty substantially improving their current rankings. Penalties for poor performance can be added to spur quality improvement among relatively poor performers.

This is a competitive approach that will not foster collaboration among providers. Competitive ranking requires provider organizations and physicians to outperform others in order to earn a P4P bonus payment. This can stimulate higher levels of quality improvement because no one knows in advance how high performance needs to be to earn the bonus payments. However, sample size can sometimes be an issue in differentiating providers' ranks because random variation may affect the measured performance results and hence the levels of P4P bonus payments. For example, one study found that smaller hospitals had a greater risk for misclassification in rankings than larger hospitals when this type of target-setting was simulated (Davidson et al., 2007). Also, relative ranks may not distinguish substantively different performance, and ties (identical scores) may be problematic.

**Combined benchmark approaches.** The various approaches to establishing performance benchmarks are not mutually exclusive. The different incentives and distribution of rewards established by alternative benchmarks may be used in combination. For example, Medicare's Premier Hospital Quality Incentive Demonstration combines all the following four elements: a target attainment award for hospitals exceeding median performance; a top performer award for the top-ranked 20 percent of hospitals; an improvement award for hospitals that attain targets and are in the top 20 percent of improvement; and a threshold penalty for hospitals scoring below the ninth decile of performance (Centers for Medicare & Medicaid Services [CMS], 2009b). In Medicare's Physician Group Practice Demonstration, participating physician group practices can satisfy quality performance standards by exceeding either (1) an absolute threshold (75 percent compliance on process quality measures), (2) an improvement threshold (reducing the gap between baseline performance and 100 percent attainment by 10 percent or more), or (3) an external relative target (established with reference to the performance of Medicare private health plans) (Kautter et al., 2007).

**Graduated or tiered rewards.** Rewards based on achieving a single performance benchmark using any of these approaches—absolute, improved, or relative performance—have the disadvantage of not giving incentives for improvement along the entire spectrum of performance. This limitation can be addressed by a system of graduated or tiered rewards that increase as the level of performance rises. For example, in the absolute approach, a payer may give a reward to provider organizations or physicians for exceeding 70 percent compliance on a process quality measure, a larger reward for exceeding 80 percent compliance, and a still larger reward for exceeding 90 percent compliance. Improvement rewards can also be graduated, with payers giving larger rewards for greater improvements. A graduated relative reward system might give a reward to provider organizations and physicians in the top 50 percent, a larger reward for those in the top 25 percent, and a penalty for those in the bottom 10 percent. All rewards systems are likely to have a minimum performance threshold below which no rewards are given.

**Continuous rewards (percentage of patients receiving recommended care).** An alternative approach is not to rely on specific thresholds of performance at all but to pay provider organizations and physicians more for each appropriately managed patient, episode, or recommended service. For example, a PCP could be paid more for each patient in her panel who had diabetes and had received clinically recommended eye and foot examinations. Under this model, physicians at any level of performance will always do better by achieving recommended care processes for more patients.

**Rebasing benchmarks.** Over time, the general level of performance may improve. Approaches that use absolute thresholds or improvement from baseline to reward performance should eventually rebase to a higher level of expected performance. Payers must find a balance between not rebasing too often, which gives provider organizations and physicians too little reward for performing well or improving performance, and rebasing too infrequently, which gives provider organizations and physicians too little incentive to continue improving performance. In the context of cost efficiency, the payer can financially capture the initial efficiency gain by rebasing (i.e., by lowering) provider payment rates, and thereby give the provider an incentive to achieve further efficiency improvements from the new, higher-efficiency baseline.

## Implementing Financial Incentives

P4P programs may implement financial incentives in a wide variety of ways. Several typical approaches to distributing incentive payments are discussed below.

**Bonus or withhold.** One common approach to distributing reward payments is through a bonus pool, which is disbursed at the end of the measurement period (e.g., annually) and is contingent on performance. A bonus pool can be funded either by using new money or by withholding a portion of regular payments throughout the year. For physicians, payers might withhold 5 or 10 percent of physicians' fees or employers might withhold a small percentage of premiums paid to health plans. The Excellus/Rochester (New York) Individual Practice Association Rewarding Results demonstration project returned to individual physicians 50 to 150 percent of a 10 percent withhold based on relative performance. The Blue Cross of California Preferred Provider Organization Rewarding Results demonstration made available a bonus of up to \$5,000 to physicians, based on their performance on selected clinical indicators (Young et al., 2007). In the context of health plans, some employers have put a percentage of health plan premiums at risk, with payments contingent on performance on administrative services measures (e.g., percentage of claims processed accurately), clinical quality, member access to services, and data reporting (Bailit & Kokenyesi, 2002). About 2 percent of the premium is typically put at risk.

**Penalties.** Payers may reduce payments to provider organizations and physicians who do not achieve an acceptable level or improvement of performance. For example, in year 3 of Medicare's Premier Hospital Quality Incentive Demonstration, participating hospitals faced a 1 percent payment reduction if they scored below the 9th decile baseline quality level and a 2 percent reduction if they scored below the 10th decile baseline level (CMS, 2009b).

**Fee schedule adjustment.** In FFS environments, payers may adjust fee schedule payments up or down, depending on performance, by adjusting the fee schedule conversion factor that translates fee schedule relative value units per service into dollar payments. For example, a PCP might be paid 105 percent of an insurer's base fee schedule if he or she ranked in the top 25 percent of network PCPs on performance measures. The Blue Cross Blue Shield of Michigan Rewarding Results demonstration allows participating hospitals to earn up to a 4 percent diagnosis-related group fee enhancement

for meeting absolute thresholds of performance on selected quality measures (Young et al., 2007).

**Per-member payment.** In capitated environments, or plans in which patients are enrolled with PCPs, a health plan might pay providers an additional or incremental per member per month or per member per year payment that is contingent on measured performance. For example, the Bridges to Excellence Rewarding Results demonstration pays a per patient per year bonus of \$100 for diabetes care and \$160 for cardiac care based on National Committee for Quality Assurance performance recognition (Young et al., 2007).

**Differential payment update.** Payers can reward provider organizations and physicians that perform well with a update factor to their payments that is higher than those given to provider organizations and physicians that perform poorly. For example, under the Medicare Reporting Hospital Quality Data for Annual Payment Update program, hospitals that did not report designated quality measures received a 0.4 percent reduction (later raised to a 2 percent reduction) in their annual payment update (CMS, 2009a).

**Payment for provision of a service.** A payer can establish payment, or enhanced payment, for services that further the goals of the P4P program. For example, if raising the rate of mammography screening is a quality goal of a P4P program, then the payer can increase the provider payment for mammography. Payers could also institute payments for activities involving coordinating and managing patient care. These might include completing an annual patient health-risk assessment and action plan or performing patient education activities.

**Payment for participation or payment for reporting.** Programs might pay provider organizations and physicians to engage in performance-enhancing activities, such as developing quality improvement action plans, attending continuing education programs, or implementing computerized physician order entry. Alternately, payers might pay provider organizations and physicians for reporting performance measures, as in Medicare's Physician Quality Reporting Initiative, which pays successfully reporting physicians 2 percent of their Medicare covered allowed charges.

**Lack of payment for poor performance.** Payers can deny payment for services that appear to be ineffective, harmful, or inefficient. Notably, payers may deny payment for preventable medical errors or their sequelae, including performing surgery on the wrong patient or body part, leaving a foreign object in a patient during surgery, or wrongly prescribing or incorrectly

administering drugs. Since October 1, 2008, Medicare no longer pays for extra costs associated with eight preventable occurrences, including transfusion with the wrong blood type, pressure ulcers, and certain hospital-acquired infections.

**Shared savings.** Payers can give providers incentives to improve efficiency and generate savings by allowing them to share in the realized savings. For example, in Medicare's Physician Group Practice Demonstration, Medicare retains 20 percent of annual measured savings and shares up to 80 percent with participating provider groups, depending on the quality performance (Kautter et al., 2007).

**Quality grants or loans.** A provider could apply to a payer for a grant to implement quality-enhancing infrastructure changes, such as an EMR or patient registry. Payers could commit to invest or lend capital to high-performing providers to build their delivery systems.

**Single versus multiple reward pools.** Payers can set up multiple reward pools to reward performance in the services supplied by each different type of provider organization or physician. For example, one reward pool might focus on hospital services, a second on PCP services, and a third on specialist physician services. With multiple pools, payers can attribute accountability more easily, but a smaller number of pools or linked pools increase the incentives for coordination of care and overall efficiency. For example, given the primary role of physicians in hospitalization, payers might partly fund physician performance payments out of the hospital pool if one of the performance goals is to keep enrollees out of the hospital. In the long run, as provider organizations that can take responsibility for entire episodes of care evolve, consolidating multiple pools into a single pool can establish better incentives for overall efficiency.

### **Magnitude and Risk of Financial Incentives**

Several important characteristics of financial incentive schemes will affect provider response to them. Among the more important are the magnitude of the incentive and the financial risk to which programs subject the provider organizations and physicians. Payer design choices affect these characteristics. Payment frequency is also an important implementation issue.

**Magnitude of incentives.** Payers must decide on the magnitude of performance incentives that they will offer to providers. The necessary incentive will depend on the cost to the provider of the intervention that the payer is rewarding. Prescribing more generic drugs may be relatively costless, but coordinating a patient's care through a nurse case manager is not. The

incentive per provider depends on the total payout and the proportion of providers who will receive the incentive. Extending incentive payments to more providers will involve a higher proportion of providers in the incentive scheme but will lower the incentive payment per provider, holding total payments constant. If a payer has a small market share, then to represent a meaningful incentive to provider organizations and physicians it may have to offer a larger incentive per member than payers with larger market shares would need to offer.

Most P4P systems have started out with incentives of limited size, although the United Kingdom's program is an exception. Reasons for limiting the size of incentives include concerns about the validity and reliability of quality measurement and data collection, the controversy created by payment disparities between providers, and provider market power to resist P4P programs. Many P4P systems in the United States provide incentives of below 5 percent of providers' total FFS incomes, although this amount may grow over time (Dudley & Rosenthal, 2006).

In this context, studies and reviews of the scientific literature on P4P have reported only limited evidence of its impact (Mullen et al., 2010; Petersen et al., 2006; Rosenthal & Frank, 2006; Sorbero et al., 2006). However, given the limited size of the incentives implemented in P4P to date, one can ask whether evaluation results showing no impact or limited impact of P4P are a fair test of this new approach to provider payment. Indeed, recent evidence has shown more positive effects of P4P, although studies have found that the effect size remains modest in most cases and the largest effects are often for provider organizations and physicians that have started at lower levels of performance (Campbell et al., 2007; Felt-Lisk et al., 2007; Gilmore et al., 2007; Glickman et al., 2007; Golden & Sloan, 2008; Grossbart, 2006; Lindenauer et al., 2007). One review suggests that incentives of about 5 percent of total physician earnings are large enough to attract "meaningful attention" from physicians (Young et al., 2007).

Nonfinancial factors may either enhance or dilute the effects of financial payments under P4P; they may certainly affect the size of the incentive payments that are needed to improve performance. Incentive payments that payers make to organizations such as hospitals or physician groups may have diluted (or enhanced) effects in relation to the individual physicians working in those organizations (Christianson et al., 2006; Young & Conrad, 2007). The organizations may or may not transmit the incentive payments directly to the physicians. Conversely, some physicians in group practices may free-ride

on the efforts of their colleagues. Organizations and payers may also support P4P programs in complementary ways, with investments in electronic health records, public reporting of performance, patient incentives for adherence to care, education of boards of directors, feedback reports to providers, and staff support for case management and care coordination (IOM Board on Health Care Services, 2007). Senior staff may work actively to promote an organizational culture that fosters quality improvement and collaboration among staff. Large incentive payments may generate more quality-maximizing behavior but may also break down the norms of clinical teamwork that are needed to improve quality. Large incentive payments may also lead to gaming or manipulation of measurement systems that could defeat the purpose of P4P.

**Risk to providers.** An important aspect of a P4P financial incentive program is the financial risk to which programs subject provider organizations and physicians. Different designs of P4P programs may greatly affect the amount and type of risk that participating providers face. We discuss different aspects of provider risk below.

- **Upside versus downside risk (“carrots versus sticks”).** Shared-savings incentives involve only upside bonus risk. If provider organizations and physicians generate savings, then they benefit by sharing in those savings. If provider organizations and physicians do not realize savings, the status quo ante is maintained, so provider organizations and physicians face no downside risk. A withhold, however, involves downside risk because the payer will not return the withhold to the provider unless the provider has met performance objectives. If participation in a P4P program is voluntary, positive incentives will be necessary to induce providers to participate.
- **Limitations on risk. Provider risks in P4P systems are typically capped.** For example, withholds are limited to 5 or 10 percent of provider payments, which is the largest amount that provider organizations and physicians can lose because of poor performance. Upside risk is also typically limited. For example, in Medicare’s Physician Group Practice Demonstration, the maximum performance payment that participating providers may earn is 5 percent of the target expenditure amount.
- **Additional versus foregone revenues.** Process quality measures may involve the provision of additional services, which are separately reimbursed under FFS payment. Because provider organizations’ and physicians’ costs of meeting the performance objective are entirely or

largely covered, the risk that they incur by meeting the performance objective is low. The necessary incremental P4P incentive may be small.

Other quality interventions reduce needed services, thus reducing provider organizations' or physicians' revenues under FFS payment and thereby creating a foregone revenue risk. For example, better ambulatory management of care may reduce hospital admissions, which would lower inpatient revenues for IDss. In this case, the P4P incentive may need to be larger to offset the foregone revenues (e.g., the provider organization could share in generated savings). Alternatively, the provider organization may realize substantial cost savings because of reduced utilization, or it may be operating at capacity and can replace lost utilization from the queue of patients waiting to use its services. To avoid the disincentive of foregone revenues, a payer may give the performance incentive to an entity that does not forego revenue, such as a physician group without an affiliated hospital or a third party such as an independent disease management organization.

- **Business risk of performance-enhancing investments.** Improving performance typically requires providers to make investments in systems and processes to improve and report their performance. Provider organizations and physicians incur business risk in making these investments because there is usually no guarantee that investments will lead to performance payments. The larger the required investments and the greater their perceived risk, the less likely provider organizations and physicians are to make them. One aspect of risk is the certainty of reward. Absolute thresholds or improvement targets have greater certainty than relative rewards, which depend on the performance of other provider organizations and physicians.

An approach that payers can take to reduce the business investment risk is to pay an upfront fee, either a lump sum “grant” or a periodic per-member payment, that finances a provider organization's or physician's performance-enhancing investments. This ameliorates the provider organization's or physician's cash flow concerns, given the lag between the necessary investments and the realization of performance payments. The greatest reduction in provider risk occurs if the upfront fee does not depend on ultimate performance. Alternatively, the upfront fee can be used only as an advance on ultimate performance payments. In this case,

the provider is at risk for the fee and ultimately for the investments it supports.

**Payment frequency.** The frequency of P4P payments may also be an issue (Young et al., 2005). Annual payments are common, but more frequent payments may provide more visibility for P4P programs and have more impact on provider behavior. However, more frequent payments will necessarily be smaller and thus may dilute a behavioral response. More frequent payments may also raise administrative burden and cost.

### **Nonfinancial Incentives**

P4P programs may also use nonfinancial incentives. Nonfinancial incentives may require less investment on the part of payers and may be less threatening to providers, whose income is not directly affected. Nonfinancial incentives include performance profiling, public recognition, technical assistance, practice sanctions, reduced administrative requirements, and automatic assignment of patients (Llanos & Rothstein, 2007).

In performance profiling, payers provide confidential feedback to providers on their performance. Public recognition, discussed in greater detail in the next section, publicizes provider performance and recognizes high-performing provider organizations and physicians. Technical assistance might occur when the payer provides help to providers in improving, for example, their achievement rates related to process-of-care criteria. A practice sanction might involve an insurer's excluding provider organizations and physicians from the provider network until they meet a threshold level of quality or efficiency performance. Reduced administrative requirements could involve quality audits of provider organizations and physicians every other year instead of annually if they meet specified performance thresholds.

In the Medicaid context, enrollees may be automatically assigned to health plans, provider organizations, or physicians (Kuhmerker & Hartman, 2007). In other contexts, payers may also automatically assign patients who fail to choose their own plans, provider organizations, or physicians. The payer or sponsor managing the enrollment process (e.g., state government or employer) can direct more patients to higher-performing plans, provider organizations, or physicians by automatically assigning more patients to them. Providers not achieving a minimum level of performance may not be assigned any patients. Assuming they are not already at capacity, automatic assignment provides an incentive for plans or providers to perform better because they will receive

more enrollees or patients without incurring marketing or other acquisition costs to enroll them.

### **Incentives for Patients to Use High-Performing Providers**

Steering patients to high-performing provider organizations and physicians creates an indirect but potentially powerful incentive for providers to improve their performance. Even a small proportion of patients changing their providers based on these incentives could represent significant revenue risk for provider organizations and physicians. From a patient's point of view, these incentives maintain freedom of provider choice but create consciousness of quality and cost differences among provider organizations and physicians. For example, a patient can continue to patronize a high-cost provider organization or physician if he or she chooses to do so, but the patient will have to pay more for this choice and must weigh the perceived quality or other advantages of the provider against the higher cost. The ability to use financial incentives may be limited for some populations, such as Medicaid enrollees, who may not be able to afford significant out of pocket payments.

**Nonfinancial incentives.** Nonfinancial incentives may be used prior to, in lieu of, or together with financial incentives. They provide a means of introducing the concept of performance measurement with less controversy than with financial incentives. Nonfinancial incentives create reputational effects that may drive referrals and patient choice.

- **Public reporting or report cards.** The mildest form of patient incentive is public reporting of quality and cost efficiency information for provider organizations and physicians. These reports are sometimes known as report cards. Public reporting arms patients with information that may help them choose provider organizations and physicians, although there is no clear reason why patients should prefer to go to more efficient provider organizations or physicians unless these providers offer lower patient out-of-pocket costs. Provider organizations and physicians may feel peer pressure, or pressure from their own internal norms of professionalism, competence, or competition, to improve their performance scores. Public reporting may also be a first stage in the introduction of patient incentives, to vet the performance measures and work out the kinks in the system.
- **Designation of high-performing provider organizations/Centers of Excellence.** The next step may be to designate certain provider organizations as superior in some way, perhaps on both cost and quality.

For example, payers may designate certain provider organizations as Centers of Excellence, giving those groups an imprimatur of quality. This designation is designed to steer patients to these groups even in the absence of financial incentives to use these providers. Provider organizations may be willing to accept a discounted payment from the payer to achieve this designation.

**Financial Incentives.** To create stronger patient incentives to use high-performing provider organizations, payers may introduce financial incentives. Payers may implement financial incentives and categorize provider performance in several ways.

- **Differential premiums.** Health plans may require their members to choose a health care system from which to receive their care or a PCP to direct and authorize their care. At the point of annual enrollment, lower health plan premiums may be charged to members who choose higher-performing health care systems.
- **Differential cost sharing.** Payers may impose financial incentives at the point of service rather than at the annual premium stage. The basic procedure is to create differential cost sharing based on the measured performance of provider organizations so that patients pay less to use higher-performing providers. Payers may charge lower copayments, coinsurance, and/or deductibles for higher-performing provider organizations and physicians.
- **Provider tiering.** Payers may classify provider organizations and physicians into tiers based on cost and quality performance. The FFS rates (e.g., discount off Medicare payment rates) that the provider offers the payer may measure cost performance. Alternatively, cost performance could be measured by case-mix-adjusted episode or per-patient-per-month expenditures that the provider organization or physician incurs for episodes or patients attributed to it. These latter approaches measure performance in controlling use as well as price charged per service. Payers may employ standard approaches to measuring process, outcome, or structural quality. Payers charge patients lower premiums or cost sharing for using providers in the higher-performance tiers.
- **Centers of Excellence.** In this approach, payers do not rank all provider organizations into tiers. Rather, payers award a smaller number of provider organizations the special designation of Centers of Excellence for specified procedures or episodes of care, such as expensive organ

transplants, or heart or orthopedic surgeries, based on their charge to the payer and the quality of care that they provide for these episodes. Patients may be required to use these centers or receive lower cost sharing or premium reductions for receiving care for these procedures at the designated centers. The goal of this approach is to consolidate volume into these centers and to use the resulting higher volume to drive quality improvement and cost reductions.

- **Other methods.** Other ways of structuring consumer incentives also exist. For example, one of the Medicaid health plans in the Local Initiative Rewarding Results P4P demonstration tried providing low-income parents with gift certificates as incentives to bring children in consistently for well-child visits, but the effort was unsuccessful in that case. Only about 3 percent of the parents sent in cards to document well-child visits to receive their gift certificates (Felt-Lisk et al., 2007).

**Health plan provider network designation.** For certain types of health plans, plan sponsors create networks of providers. In health maintenance organizations (HMOs), enrollees generally have no coverage for out-of-network providers. In preferred provider organizations (PPOs), enrollees have some out-of-network coverage but face higher cost sharing if they use out-of-network providers. Health plans may base their provider network selection largely on the cost and quality performance of provider organizations and physicians. Cost is typically the payment rate that the provider organization or physician is willing to accept from the health plan, and quality may be based on simple credentialing or more sophisticated quality indicators. Network-based health plans provide a means of translating provider performance into differential premiums or cost sharing for patients.

## **Limitations of Pay for Performance**

Although P4P is currently a powerful movement in health care, payers and policymakers should recognize its limitations. This section discusses those limitations and approaches to dealing with them. The next section identifies some alternatives and complements to P4P.

### **Lack of Valid, Reliable, and Important Performance Indicators**

Measuring performance in health care can be quite difficult. Quality of care, for instance, is influenced by many physician, patient, health care system, and environmental factors. Determining the marginal contribution of a

provider organization or physician to a given process or outcome is often challenging (Hahn, 2006). Also, many areas of medical practice suffer from large uncertainties about the best approaches. The relationship between many health care processes and outcomes is difficult to discern. The number of times that payers can observe recommended processes and especially outcomes for individual physicians is often small, leading to concern about the statistical reliability of performance measurement.

Health care expenditures, often used to measure efficiency, are subject to enormous variation because of patient case mix, which is unrelated to efficiency. It is challenging to hold constant or adjust for this underlying patient health status variation so that payers can distinguish differences of a few percentage points in the efficiency of provider organizations. Large sample sizes—a minimum of 10,000 to 15,000 patients per profiled provider unit—and powerful risk adjustment methodologies that provider organizations cannot manipulate are necessary but often not available (Kautter et al., 2007; Nicholson et al., 2008).

In these circumstances, reliably isolating, measuring, and attributing the incremental contributions of individual provider organizations, physicians, or even health plans to quality or efficiency is difficult. Available performance indicators are often driven by the data (e.g., administrative billing data) that are available at reasonable cost and that have usually been collected for purposes other than measuring performance. Administrative data have limitations, however; for example, they may lack the clinical detail necessary to measure quality of care adequately. The result is that payers may base P4P programs on available performance indicators rather than important or optimally measured performance indicators.

Provider organizations are understandably concerned about having their performance judged on measures that may not be valid, reliable, or important. Approaches to improving the value of a P4P program include focusing on areas that have a high degree of consensus about appropriate medical practice, that amass accurate data and sample sizes sufficient to measure performance reliably, that represent important areas of medical practice in terms of quality or cost, and adjusting for as many noncontrollable factors as possible. Using a transparent and relatively simple performance assessment and reward system can also promote understanding and acceptance by provider organizations and physicians (Folsom et al., 2008).

### **Lack of Comprehensive Performance Indicators**

Even if performance can be gauged accurately in some areas, comprehensive performance measurement may not be possible or may be too costly to obtain. If such assessments are not comprehensive, provider organizations and physicians may focus on improving their performance in the areas that can be measured and neglect areas that are not examined or rewarded. Performance could actually deteriorate in unmeasured areas, and this unintended consequence may be more important to ultimate outcomes than measured areas. One solution to this problem is to rotate measures among multiple areas across performance periods (e.g., across years), so that provider organizations and physicians cannot consistently do well by focusing on only one performance domain or narrow set of indicators.

### **Prescriptiveness or Lack of Flexibility of Performance Measures**

Process measures may use considerable detail to specify how patients should be treated in specific circumstances. If the goal of a P4P program is to promote the adoption of certain evidence-based care processes, this level of detail may be an advantage, especially when exceptions processes are available to excuse justified noncompliance with recommended care. Alternatively, the performance measurement approach may be overly prescriptive and may intrude on provider organizations' and physicians' autonomy, flexibility, and ability to use professional judgment to decide the best course of care in particular situations (Epstein et al., 2004). Measuring and rewarding the ultimate outcomes of interest rather than detailed intermediate care processes allows provider organizations and physicians to have the autonomy and flexibility to determine the best means to achieve ultimate outcomes.

### **Lack of Cost-Effectiveness**

Implementing and administering a P4P program may be quite costly. P4P programs may impose large costs on provider organizations. Simply reporting performance measures may be quite expensive for provider organizations, especially for solo practices, smaller groups, or institutions with limited resources and when reporting requires large fixed investments. Such providers may need to purchase and implement complex information systems and collect and validate expensive data. The investments required to improve performance may also be costly, depending on the performance measure. Provider organizations may have to hire additional staff to manage patient care (e.g., to document which patients are not receiving recommended care and to

convince them to receive it), purchase information technology systems, and allocate portions of individual physicians' time to complying with performance indicators.

Payers also have burdens for administering P4P programs. They must define performance measures, collect and process the necessary data, evaluate performance, disseminate results, and implement incentives. Payers must involve, educate, assist, and adjudicate appeals from provider organizations and physicians. Organizations and physicians are likely to demand higher reimbursement from payers to defray their costs of reporting performance measures and otherwise participating in a P4P program.

The value of the performance gains that one can reasonably expect from a P4P system may not clearly justify the large costs that the system imposes on both providers and payers. The business case for P4P may be especially hard to make when the financial gains from improved performance are likely to mostly accrue in the future—as may be true with better management of chronic disease—but the costs are immediate. The imbalance of short-run costs and long-run savings is especially difficult to justify in settings, such as in employer-based health insurance, that have a high turnover of plan members; the principal reason is that the gains are not likely to accrue to the same health plan or even to the same employer as the one that incurs the initial cost.

Given these considerations, sponsors of P4P programs should evaluate all performance indicators for cost-effectiveness. For example, insurers with a high rate of patient turnover may wish to focus on measures that have short-term payback. The Medicare program, or the government more generally, has more reason to establish P4P efforts that can invest in improving longer-term performance.

**Unintended consequences.** P4P incentive payments may have unintended consequences that could be detrimental in several ways. One concern is that physicians may begin to avoid taking on more “difficult” patients so that they can avoid scoring poorly on quality or efficiency (Epstein et al., 2004). However, difficult patients—such as those with multiple chronic diseases or low socioeconomic status—may need high-quality, coordinated health care more than other patients.

Performance measures may be risk adjusted for the characteristics of patients to reduce or eliminate providers' disincentive to accept high-risk patients; that is, programs might disburse higher payments to providers for taking on more difficult patients (IOM Board on Health Care Services, 2007). Risk adjustment is complex and controversial, however. Adjusting

quality indicators for patient characteristics may implicitly create a lower benchmark standard for the care of high-risk patients, who often have lower socioeconomic status, are minorities, and have worse health than other patients. Moreover, whether (or to what extent) provider organizations and physicians trust the currently available risk adjustment systems to protect them from the potential for negative performance assessments in P4P programs is not clear (Dudley & Rosenthal, 2006). If provider organizations and physicians believe that P4P payments put a substantial portion of their compensation at financial risk, they may demand as a condition of participating a risk premium (higher payments) from payers to compensate them for this risk (Nicholson et al., 2008).

Another potential unintended consequence of P4P would be exacerbated disparities in care (IOM Board on Health Care Services, 2007; Karve et al., 2008). For example, provider organizations in high-income communities might be able to fund IT and quality improvement systems at a higher level than organizations in low-income communities, thereby earning a larger share of P4P bonus payments. A study of P4P in Medicaid health plans found that provider organizations often reported lacking the office staff and systems needed to respond to the quality improvement incentives, and they did not have the financial resources needed to hire more staff and install better information systems (Felt-Lisk et al., 2007).

In general, P4P incentives have the potential to either narrow or widen disparities in health care (Chien et al., 2007). P4P programs can be designed to reduce disparities in care, if that is established as an explicit goal; however, very few P4P programs to date have been designed explicitly either to limit their impact on disparities or to actively reduce them. One program that has is the CMS Cancer Prevention and Treatment Demonstration for Ethnic and Racial Minorities, a randomized controlled trial that uses patient navigators to reduce racial and ethnic disparities in cancer screening and treatment (Mitchell et al., 2008). Demonstration sites receive monthly capitation payments to provide navigation services for beneficiaries enrolled in the intervention arm of the study.

Payers may also be vulnerable to unintended consequences. In the United Kingdom, doctors initially met more targets than projected, resulting in much larger payouts than the government had expected (Epstein, 2006). This produced a larger deficit for the National Health Service than the government had anticipated. Advocates of P4P programs may believe that they will reduce costs, but an exclusive focus on improving quality, especially service underuse as indicated by process-of-care measures, may or not may not reduce long-run

costs. In the short run, under FFS payment, payer costs are likely to rise as additional services are provided (Hahn, 2006).

### **Difficulty of Patient Attribution**

In situations in which many provider organizations or physicians treat individual patients without coordinated care, attributing care to individual provider organizations or physicians may be difficult; similarly, ascertaining which provider organizations or physicians are responsible for the observed processes or outcomes is challenging. Analysts may use various attribution rules, but none of these may be ideal (see Chapter 7 for more on patient attribution to physicians or organizations). An alternative approach to using attribution rules is to institute a voluntary or mandatory system in which patients choose a provider organization or physician (e.g., a primary care gatekeeper or “medical home”) that is assigned overall responsibility for managing the patient’s care. However, making such assignments mandatory may conflict with patient freedom of choice. Inherent conflicts may exist between complete patient freedom of provider choice and provider responsibility in P4P programs. Organizational changes that clarify provider responsibility—changes that to some degree may compromise patient freedom of choice—may be necessary precursors to effective P4P programs.

### **Multiple Payers with Inconsistent Programs**

Most provider organizations and physicians treat enrollees who, together, are covered by multiple private insurers and government programs. If each payer implements its own P4P program with different performance measures, reporting requirements, and incentive schemes, the costs to provider organizations and physicians of participation or compliance will be much higher than with a single, coordinated P4P program. Working with multiple programs is likely to result in confusion and to dilute the impact of P4P (Hahn, 2006). The obvious solution is for payers to coordinate their programs, which they have done in some cases, such as the Integrated Healthcare Initiative in California. However, coordinating programs may be costly and difficult, especially among competing private insurers, when P4P programs are an aspect of competitive advantage or coordination may be subject to antitrust restrictions.

## **Alternatives and Complements to Pay for Performance**

Payers and policymakers can consider using several alternative approaches instead of or in combination with P4P to further the goals of improving the quality and efficiency of medical care. The main ones are provider reimbursement; professionalism/provider education; quality regulation and accreditation; malpractice insurance; market competition, reputation, and public reporting; and incentives to patients or enrollees.

### **Provider Reimbursement**

The basic system that payers use to reimburse provider organizations and physicians embodies incentives for quality and efficiency. FFS and capitation are the two canonical reimbursement systems. FFS rewards the provision of extra services; thus, it tends to perform well on access but poorly on cost efficiency. Capitation rewards withholding services; therefore, it tends to perform well on cost efficiency but poorly on access. A payer that is particularly concerned about its members' access to care may find using FFS provider reimbursement more effective than adopting a P4P program that rewards good performance on access measures. Similarly, a payer particularly concerned about cost efficiency and controlling costs might want to use capitated reimbursement rather than pay for performing well on cost efficiency indicators.

P4P programs can be added to the underlying reimbursement system to reinforce its incentives or to provide incentives for performance that the reimbursement system does not. In the latter circumstance, if a P4P program is going to be incremental to FFS provider reimbursement, the program logically should include a focus on cost efficiency. A P4P program that is added to capitated reimbursement would logically incorporate a focus on access.

In terms of incentives for quality, if higher quality is associated with the provision of more services, FFS promotes higher quality. FFS contains no incentives to achieve the ultimate outcome (i.e., good health) and may be inimical to it if achieving good health involves providing fewer services (e.g., avoiding medical mistakes that require additional treatment). Capitation provides incentives to avoid mistakes and invest in cost-effective quality enhancements that reduce long-run costs, but a basic incentive of capitation is to provide fewer services, especially expensive ones, which may be inconsistent with high-quality care.

Given these considerations, a P4P program that has the goal of improving the quality of care may be a useful supplement to either FFS or capitated

provider reimbursement. The form of the P4P program should complement the underlying reimbursement system. For example, an extra per member per month payment for good quality performance is feasible in a capitated environment, while a higher fee schedule conversion factor for high-quality provider organizations and physicians could only be used in a FFS environment.

### **Professionalism/Provider Education**

Payers may rely on providers' sense of professionalism to promote good care. That is, provider organizations and physicians may provide good care because that is "the right thing to do," because they are being paid to care for patients and want to do their job well. For physicians, the power of professional training, ethics, and norms of patient-centered behavior are important factors that may limit the impact of financial incentives from P4P (Golden & Sloan, 2008). Intrinsic motivation—the internal reward of performing a task for its own sake—may be as important for some providers as extrinsic motivation from P4P incentive payments, peer and community recognition, and other external factors.

The statistics that show that the current health care system provides recommended care inconsistently (McGlynn et al., 2003) have somewhat undermined this professionalism argument and have abetted the rise of P4P programs with explicit financial incentives for high-quality care. Professionalism may not be enough to ensure high-quality care, but it is an important adjunct to financial incentives. Provider profiling, feedback, and education fit in with the professionalism approach. Programs may attain performance improvements, the argument goes, by educating provider organizations and physicians about their performance and relying on their sense of professionalism to improve, even without public reporting and financial incentives.

### **Quality Regulation and Accreditation**

An alternative to offering financial incentives is regulating quality. Payers and accreditation bodies may regulate provider organizations and physicians with minimum quality standards. For example, government programs such as Medicare and Medicaid and private accreditation organizations such as the Joint Commission review the credentials, eligibility, and suitability of provider organizations and individual physicians to provide care paid for by their programs or member organizations. State agencies may license

organizations and physicians to allow them to operate legally. These licensure, program approval, and accreditation assessments may review structural and other quality indicators as well as legal compliance, malpractice actions, and other factors. If provider organizations or physicians do not satisfy these organizations' quality and other standards, payers can deny payment for provided care and regulating bodies can prohibit provider organizations and physicians from practicing medicine.

Quality regulation can ensure that all provider organizations and physicians meet a minimum threshold of measured quality, but achieving improvements in quality above the minimum threshold may be cumbersome and expensive. If something is so critical to quality that a payer feels that all organizations and physicians that it pays must have it, then mandating it as a condition of eligibility for the payer's reimbursement can be effective. For aspects of quality in which a higher rate of performance is desirable but 100 percent compliance is infeasible, prohibitively expensive, or not critical, it makes more sense for a P4P program to offer incentives for compliance rather than to mandate it.

### **Malpractice Insurance**

Legal actions against provider organizations and physicians by patients who believe they have suffered adverse outcomes of care create an incentive for providers to avoid medical mistakes and furnish, if not high-quality care, at least the usual standard of care. Physicians and provider organizations purchase malpractice insurance against lawsuits, but if their policy premiums are experience rated, those who suffer more adverse malpractice judgments will pay higher insurance premiums.

However, the legal system is a limited mechanism for maintaining and improving the quality of medical care. Only a small portion of adverse medical events result in malpractice lawsuits, and many malpractice claims are unrelated to physician negligence (Weiler et al., 1993). P4P programs can provide a much more comprehensive and systematic measurement of the quality performance of provider organizations and physicians than the occasional malpractice lawsuit. Another problem with relying on the legal system is that malpractice lawsuits are very expensive (e.g., lawyer's fees) and time consuming. A P4P program may be a much more cost-effective and timely means of improving quality than the legal process. The legal process does have the advantage of compensating some victims of poor medical care, however, which is not a feature of P4P programs.

### **Market Competition, Reputation, and Public Reporting**

In typical markets, competition among sellers (provider organizations and physicians) and seller reputation are important forces for maintaining quality. If sellers do not maintain their quality and reputation for quality, buyers (patients) will buy from other sellers or lower the price that they are willing to pay to low-quality sellers. Competition on quality can be enhanced through credible third-party measurement and reporting of seller quality (Nicholson et al., 2008).

The health care market is different from typical markets in several important ways. One is the presence of insurance, which means that payment for medical services is mostly made by the insurer, not out of pocket by the patient. A second is that quality is difficult to measure and judge in health care, especially for many patients. These factors create a strong role for the payer to ensure and promote quality in health care markets. Competition on and reputation for quality are important in health care, just as they are in other industries. One way in which the health care payer can promote quality is by measuring and publicly reporting the quality performance of provider organizations and physicians. Moreover, the payer's role in purchasing care means that it can also create payment incentives around the quality of care, as in a P4P program.

### **Incentives to Patients or Enrollees**

P4P tends to focus on provider organizations and physicians rather than patients, although, as mentioned earlier, payers may use provider performance measurement to create incentives for patients to patronize high-performing providers. An alternative or complement to provider P4P is patient or enrollee P4P. Rather than giving provider organizations and physicians an incentive for the proportion of their patients complying with a process quality measure, payers could give patients direct incentives to comply, either through a direct payment or lower insurance cost-sharing for the service in question (so-called value-based insurance design). For example, payers could give patients with diabetes an incentive payment to keep their blood sugar under control, or they could reduce these patients' cost sharing for annual eye examinations. Some companies give employees incentives for lifestyle changes (e.g., lower health insurance premiums for nonsmokers, bonuses for losing weight or participating in fitness programs). Enabling patients to benefit from P4P payments is an explicit acknowledgement that they are part of what economists

call the health care production function. However, only a limited number of programs include patients in P4P incentive payments today.

Patient incentives can complement provider incentives. Patient incentives can address concerns that provider organizations and physicians have sometimes expressed about being held accountable for quality performance for patients who do not adhere to prescribed tests and treatments. For example, physicians may prescribe angiotensin-converting enzyme inhibitors or angiotensin II receptor blockers for patients with heart failure, but physicians cannot force the patients to fill the prescriptions or take the medications as prescribed. Similarly, a study of P4P in Medicaid health plans found that low-income parents often lacked time and transportation needed to bring children in consistently for well-child visits that were included in P4P assessments (Felt-Lisk et al., 2007).

Clearly, direct patient incentives are not appropriate for some of the more technical aspects of medical care that are not under patient control (e.g., avoiding surgical mistakes). Patient and provider incentives may be more complements than substitutes, but for performance measures that are ultimately under patient control, such as lifestyle, it is an empirical question whether provider or patient incentives are more effective in improving performance. On the efficiency side, consumer-directed health plans put consumers in charge of managing their own health care and focus financial incentives on the consumer. Demand side (consumer, patient) efficiency incentives are an important complement to supply side (provider organization, physician) efficiency incentives.

## Concluding Comments

P4P has substantial conceptual appeal. It seems logical that payment should be related to demonstrated performance on the objectives established by the payer. However, P4P is a general framework for payment, not a specific method that can be applied in every situation. As this chapter illustrates, a very large variety of performance measurement and payment schemes can fall under the rubric of P4P. The success or failure of P4P in particular applications depends on how payers evaluate performance and structure incentives. As is often true, the devil is in the details.

Because P4P is a general conceptual framework, considerable experimentation and evaluation is likely to continue for some time. We are unlikely to conclude that P4P universally fails or succeeds. Over time, payers

and policy makers will discover and disseminate the successful elements of P4P and discard the unsuccessful elements. Payers and policy makers will incorporate the successful elements of P4P into other big conceptual frameworks, such as managed care.

A major limitation of P4P is that implementing it well in practice is often difficult. Achieving a valid, reliable, and comprehensive measurement of performance in an area as complex as medical care is extremely challenging. Structuring financial incentives to achieve the intended goals while avoiding unintended consequences can also be difficult. The theory of optimal incentive contracts shows that when available performance measures are “noisy” (imprecise in their relation to the outcomes of ultimate interest) and “distorted” (improving the measure does not necessarily improve the outcome of ultimate interest), the proportion of compensation that should be based on them is lower (Baker, 2002). Thus, payers should be cautious about tying a large proportion of physician and other provider reimbursement to incomplete and flawed performance measures. P4P may prove most useful in specific, narrow applications in which an accurate assessment of performance can be obtained.

Because of the diversity of P4P programs and their contexts and environments, evaluating and generalizing individual programs is hard, too. Whether evaluation results from one P4P program will apply to other programs is rarely, if ever, clear (Hahn, 2006). Rigorous evaluation evidence to support P4P’s impact on quality of care and other performance metrics remains limited (IOM Board on Health Care Services, 2007; Christianson et al., 2008; Damberg et al., 2009; Epstein, 2007).

P4P is not a panacea for improving health care (Sorbero et al., 2006). We need to consider it as part of a set of complementary and substitutable strategies to achieve payer objectives, such as those discussed in this chapter. P4P is not necessarily the best strategy, or even appropriate, in all situations. Nicholson et al. (2008) discuss circumstances under which P4P is more or less useful. An important contribution of the P4P movement, however, is payers’ increased emphasis on holding provider organizations and physicians accountable for the value of their health care, rather than simply paying for the volume of care. This orientation, the P4P framework, has the potential to eventually contribute significantly to improving the quality and efficiency of health care.

## References

- Bailit, M., & Kokenyesi, C. (2002). *Financial performance incentives for quality: The state of the art* (National Health Care Purchasing Institute Executive Brief). Retrieved January 13, 2011, from <http://www.bailit-health.com/articles/NHCPI-brief1.pdf>
- Baker, G. (2002). Distortion and risk in optimal incentive contracts. *The Journal of Human Resources*, 37(4), 728-751.
- Campbell, S., Reeves, D., Kontopantelis, E., Middleton, E., Sibbald, B., & Roland, M. (2007). Quality of primary care in England with the introduction of pay for performance. *New England Journal of Medicine*, 357(2), 181–190.
- Centers for Medicare & Medicaid Services. (2009a, May 1). *Hospital quality initiatives: Reporting hospital quality data for annual payment update*. Retrieved June 18, 2009, from [http://www.cms.hhs.gov/HospitalQualityInits/08\\_HospitalRHQDAPU.asp](http://www.cms.hhs.gov/HospitalQualityInits/08_HospitalRHQDAPU.asp)
- Centers for Medicare & Medicaid Services. (2009b, July). *Premier Hospital Quality Incentive Demonstration: Fact sheet*. Retrieved February 10, 2010, from <http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalPremierFactSheet200907.pdf>
- Chien, A. T., Chin, M. H., Davis, A. M., & Casalino, L. P. (2007). Pay for performance, public reporting, and racial disparities in health care: How are programs being designed? *Medical Care Research and Review*, 64(5 Suppl), 283S–304S.
- Christianson, J. B., Knutson, D. J., & Mazze, R. S. (2006). Physician pay-for-performance. Implementation and research issues. *Journal of General Internal Medicine*, 21(Suppl 2), S9–S13.
- Christianson, J. B., Leatherman, S., & Sutherland, K. (2008). Lessons from evaluations of purchaser pay-for-performance programs: a review of the evidence. *Medical Care Research and Review*, 65(6 Suppl), 5S–35S.
- Damberg, C. L., Raube, K., Teleki, S. S., & Dela Cruz, E. (2009). Taking stock of pay-for-performance: A candid assessment from the front lines. *Health Affairs (Millwood)*, 28(2), 517–525.
- Davidson, G., Moscovice, I., & Remus, D. (2007). Hospital size, uncertainty, and pay-for-performance. *Health Care Financing Review*, 29(1), 45–57.

- Doran, T., Fullwood, C., Gravelle, H., Reeves, D., Kontopantelis, E., Hiroeh, U., et al. (2006). Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine*, 355(4), 375–384.
- Dudley, R. A., & Rosenthal, M. B. (2006). Pay for performance: A decision guide for purchasers (final contract report). Retrieved January 12, 2011, from <http://www.ahrq.gov/qual/p4pguide.htm>
- Eddy, D. M. (1998). Performance measurement: Problems and solutions. *Health Affairs (Millwood)*, 17(4), 7–25.
- Epstein, A. M. (2006). Paying for performance in the United States and abroad. *New England Journal of Medicine*, 355(4), 406–408.
- Epstein, A. M. (2007). Pay for performance at the tipping point. *New England Journal of Medicine*, 356(5), 515–517.
- Epstein, A. M., Lee, T. H., & Hamel, M. B. (2004). Paying physicians for high-quality care. *New England Journal of Medicine*, 350(4), 406–410.
- Felt-Lisk, S., Gimm, G., & Peterson, S. (2007). Making pay-for-performance work in Medicaid. *Health Affairs (Millwood)*, 26(4), w516–w527.
- Folsom, A., Demchak, C., & Arnold, S. B. (2007). *Rewarding results pay-for-performance: Lessons for Medicare*. Washington, DC: AcademyHealth, Changes in Health Care Financing and Organization Program of the Robert Wood Johnson Foundation. Retrieved January 12, 2011, from <http://www.hcfo.org/files/hcfo/HCFOMonograph0308.pdf>
- Gilmore, A. S., Zhao, Y., Kang, N., Ryskina, K. L., Legorreta, A. P., Taira, D. A., et al. (2007). Patient outcomes and evidence-based medicine in a preferred provider organization setting: A six-year evaluation of a physician pay-for-performance program. *Health Services Research*, 42(6 Pt 1), 2140–2159; discussion 2294–2323.
- Glickman, S. W., Ou, F. S., DeLong, E. R., Roe, M. T., Lytle, B. L., Mulgund, J., et al. (2007). Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA*, 297(21), 2373–2380.
- Golden, B., & Sloan, F. (2008). Physician pay for performance: Alternative perspectives. In F. Sloan & H. Kasper (Eds.), *Incentives and choice in health care*. Cambridge, MA: MIT Press.
- Grossbart, S. R. (2006). What's the return? Assessing the effect of “pay-for-performance” initiatives on the quality of care delivery. *Medical Care Research and Review*, 63(1 Suppl), 29S–48S.

- Hahn, J. (2006). *Pay-for-performance in health care*. Congressional Research Service Report for Congress, Order Code RL33713. Retrieved January 13, 2011, from <http://www.allhealth.org/briefingmaterials/CRSReportingforCongress-Pay-for-PerformanceinHealthCare-501.pdf>
- Hussey, P. S., de Vries, H., Romley, J., Wang, H. C., Chen, S. S., Shekelle, P. G., et al. (2009). A systematic review of health care efficiency measures. *Health Services Research, 44*(3), 784–805.
- Institute of Medicine, Board on Health Care Services. (2007). *Rewarding provider performance: Aligning incentives in Medicare*. Washington, DC: National Academies Press.
- Karve, A. M., Ou, F. S., Lytle, B. L., & Peterson, E. D. (2008). Potential unintended financial consequences of pay-for-performance on the quality of care for minority patients. *American Heart Journal, 155*(3), 571–576.
- Kautter, J., Pope, G. C., Trisolini, M., & Grund, S. (2007). Medicare Physician Group Practice Demonstration design: Quality and efficiency pay-for-performance. *Health Care Financing Review, 29*(1), 15–29.
- Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (1999). *To err is human: Building a safer health system*. Washington, DC: Institute of Medicine, National Academies Press.
- Kuhmerker, K., & Hartman, T. (2007). *Pay-for-performance in state Medicaid programs: A survey of state Medicaid directors and programs*. Retrieved January 13, 2011, from [http://www.commonwealthfund.org/usr\\_doc/Kuhmerker\\_P4PstateMedicaidprogs\\_1018.pdf?section=4039](http://www.commonwealthfund.org/usr_doc/Kuhmerker_P4PstateMedicaidprogs_1018.pdf?section=4039)
- Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., et al. (2007). Public reporting and pay for performance in hospital quality improvement. *New England Journal of Medicine, 356*(5), 486–496.
- Llanos, K., & Rothstein, J. (2007). *Physician pay-for-performance in Medicaid: A guide for states*. Report prepared for the Commonwealth Fund and the Robert Wood Johnson Foundation. Hamilton, NJ: Center for Health Care Strategies, Inc.
- McDermott, S., & Williams, T. (2006). Advancing quality through collaboration: The California Pay for Performance Program: A report on the first five years and a strategic plan for the next five years. Retrieved January 13, 2011, from [http://www.iha.org/pdfs\\_documents/p4p\\_california/P4PWhitePaper1\\_February2009.pdf](http://www.iha.org/pdfs_documents/p4p_california/P4PWhitePaper1_February2009.pdf)

- McGlynn, E. A., Asch, S. M., Adams, J., Keesey, J., Hicks, J., DeCristofaro, A., et al. (2003). The quality of health care delivered to adults in the United States. *New England Journal of Medicine*, 348(26), 2635–2645.
- Mitchell, J. B., Holden, D. J., & Hoover, S. (2008). Evaluation of the Cancer Prevention and Treatment Demonstration for Racial and Ethnic Minorities: Report to Congress. Available from [http://www.cms.hhs.gov/reports/downloads/Mitchell\\_CPTD.pdf](http://www.cms.hhs.gov/reports/downloads/Mitchell_CPTD.pdf)
- Mullen, K. J., Frank, R. G., & Rosenthal, M. B. (2010). Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *RAND Journal of Economics*, 41(1), 64–91.
- Nicholson, S., Pauly, M. V., Wu, A. Y., Murray, J. F., Teutsch, S. M., & Berger, M. L. (2008). Getting real performance out of pay-for-performance. *Milbank Quarterly*, 86(3), 435–457.
- Petersen, L. A., Woodard, L. D., Urech, T., Daw, C., & Sookanan, S. (2006). Does pay-for-performance improve the quality of health care? *Annals of Internal Medicine*, 145(4), 265–272.
- Rosenthal, M. B., & Frank, R. G. (2006). What is the empirical basis for paying for quality in health care? *Medical Care Research and Review*, 63(2), 135–157.
- Rosenthal, M. B., Frank, R. G., Li, Z., & Epstein, A. M. (2005). Early experience with pay-for-performance: from concept to practice. *JAMA*, 294(14), 1788–1793.
- Rosenthal, M. B., Landon, B. E., Normand, S. L., Frank, R. G., & Epstein, A. M. (2006). Pay for performance in commercial HMOs. *New England Journal of Medicine*, 355(18), 1895–1902.
- Sorbero, M., Damberg, C. L., Shaw, R., Teleki, S., Lovejoy, S., DeCristofaro, A., et al. (2006). Assessment of pay-for-performance options for Medicare physician services: Final report. Available from <http://aspe.hhs.gov/health/reports/06/physician/report.pdf>
- Weiler, P. C., Hiatt, H. H., Newhouse, J. P., Johnson, W. G., Brennan, T. A., & Leape, L. L. (1993). A measure of malpractice: Medical injury, malpractice litigation, and patient compensation. Cambridge, MA: Harvard University Press.

- Wennberg, J. E., Fisher, E. S., Stukel, T. A., & Sharp, S. M. (2004). Use of Medicare claims data to monitor provider-specific performance among patients with severe chronic illness. *Health Affairs (Millwood), Supplemental Web Exclusives*, VAR5–VAR18.
- Werner, R. M., & Dudley, R. A. (2009). Making the “pay” matter in pay-for-performance: implications for payment strategies. *Health Affairs (Millwood)*, 28(5), 1498–1508.
- Young, G. J., Burgess, J. F., Jr., & White, B. (2007). Pioneering pay-for-quality: Lessons from the rewarding results demonstrations. *Health Care Financing Review*, 29(1), 59–70.
- Young, G. J., & Conrad, D. A. (2007). Practical issues in the design and implementation of pay-for-quality programs. *Journal of Healthcare Management*, 52(1), 10–18; discussion 18–19.
- Young, G. J., White, B., Burgess, J. F., Jr., Berlowitz, D., Meterko, M., Guldin, M. R., et al. (2005). Conceptual issues in the design and implementation of pay-for-quality programs. *American Journal of Medical Quality*, 20(3), 144–150.