Innovative Health Reform Models: Pay-for-Performance Initiatives

Seth W. Glickman, MD, MBA; and Eric D. Peterson, MD, MPH

Abstract

Pay-for-performance (P4P) programs have the potential to improve overall quality of care by narrowing gaps between what national care guidelines recommend and those treatments actually delivered in routine community practice. P4P is also viewed as a tool to promote more efficient use of healthcare resources while improving patient outcomes. P4P provides financial incentives for quality of service instead of quantity of service. Despite the promise of healthcare quality, concerns have been raised that P4P may have potential unintended consequences for patients, physicians, and hospitals. The shortcomings of many traditional P4P programs have fueled the emergence of new and innovative models of payment reform. P4P and newer models that link reimbursement with quality and efficiency show promise to improve patient outcomes and lower costs, but multiple approaches are needed to ensure that future initiatives provide value for key stakeholders, including patients, providers, and payers.

(Pam J Manag Care. 2009;15:S300-S305)

What Is Pay for Performance?

The concept of linking financial incentives to the quality of healthcare provided has been termed pay for performance, or P4P. P4P has received significant national attention as a potential means of improving overall quality of care by narrowing gaps between what national care guidelines recommend and those treatments actually delivered in routine community practice. In an era of rapidly rising healthcare costs, P4P is also viewed as a tool to promote more efficient use of healthcare resources while improving patient outcomes. Over the past decade, commercial and federal payers have implemented a vast array of P4P initiatives. The structure of these programs is remarkably diverse, spanning a variety of payment models, intended targets (eg, hospitals, physician groups, individual providers), and clinical conditions. While a number of programs have shown promise, additional work is necessary to determine whether they achieve their intended long-term effects. The purpose of the article is to provide a brief overview of current P4P initiatives, discuss the evidence regarding their effectiveness, and provide insight into newer, innovative payment models that have emerged.

P4P Programs

More than half of commercial health plans in the United States currently use P4P incentives in their provider contracts. Many of these programs involve joint efforts among employers, health management organizations, pharmaceutical companies, physician groups, academia, as well as for-profit and not-for-profit organizations (Table). Complementing efforts by the private sector, the Centers for Medicare & Medicaid Services (CMS) has sponsored P4P demonstration projects in a variety of clinical settings, including physician practices, acute care hospitals, dialysis facilities, nursing homes, as well as programs to increase the adoption of information technology and disease management. The largest demonstration project to date is the Hospital Quality Improvement Demonstration (HQID) Project, which offers financial incentives to hospitals based on the inpatient quality of care for 5 clinical conditions—acute myocardial infarction, heart failure, pneumonia, coronary artery bypass surgery, and hip and knee replacement. In addition to these US programs, P4P has gained significant traction overseas. For example, the National Health Service (NHS) in the United Kingdom has invested massive resources in P4P initiatives. The NHS's Quality and...
Innovative Health Reform Models: Pay-for-Performance Initiatives

Outcomes Framework, which provides financial incentives to primary care physicians for 146 quality indicators related to chronic disease and patient experience, has distributed over £2 billion to providers since 2004.

**Table. Examples of Some of the Largest Pay-for-Performance Initiatives to Date**

<table>
<thead>
<tr>
<th>Program</th>
<th>Participant</th>
<th>Sponsor</th>
<th>Target</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital Quality Improvement Demonstration Project</td>
<td>230 Acute care hospitals in the United States</td>
<td>Center for Medicare &amp; Medicaid Services</td>
<td>Process measures for heart failure, acute myocardial infarction, pneumonia, hip replacement, coronary artery bypass grafting surgery</td>
<td>Modest improvements in process performance, no identifiable impact on outcomes</td>
</tr>
<tr>
<td>Quality and Outcomes Framework/General Practitioner Pay-for-Performance contract</td>
<td>42 Family practices in England</td>
<td>National Health Service</td>
<td>146 Indicators related to chronic disease and patient experience</td>
<td>Short-term improvements in care which slowed once performance targets were reached</td>
</tr>
<tr>
<td>Integrated Healthcare Association Pay-for-Performance Program</td>
<td>225 Physician groups in California</td>
<td>8 Health plans in California representing 10.5 million patients</td>
<td>Multiple, including clinical process measures, patient experiences of care, adoption of information technology</td>
<td>Modest improvements in targeted areas of care</td>
</tr>
<tr>
<td>Bridges to Excellence</td>
<td>Multiple provider groups operating in 13 states</td>
<td>Collaborative effort among large employers, including General Electric and Verizon Communications</td>
<td>Includes excellence programs in diabetes, cardiac, spine, and depression</td>
<td>Cost savings in diabetes care, achievement of performance thresholds in diabetes and cardiac care</td>
</tr>
<tr>
<td>Hill Physicians Medical Group</td>
<td>2200 Physicians in North Carolina</td>
<td>Hill Physicians Medical Group serving 332,000 patients in 7 HMOs</td>
<td>Resource utilization, clinical performance (cancer, diabetes, low back pain, immunization), patient experience, up to 15% of compensation to quality performance</td>
<td>Improvement in threshold diabetes care by 42% and cholesterol levels by 32%</td>
</tr>
<tr>
<td>Hawaii Medical Service Association Practitioner and Hospital Quality and Service Recognition Programs</td>
<td>More than 2500 physicians in its preferred provider plan, 17 hospitals</td>
<td>Blue Cross Blue Shield of Hawaii</td>
<td>Patient safety, adherence to evidence-based guidelines, patient satisfaction</td>
<td>Significant improvements in adherence to clinical measures in a number of areas, including cancer screening, immunization, and heart failure</td>
</tr>
</tbody>
</table>

HMO indicates health maintenance organization.

Potential Benefits of P4P

Compensation models that link financial incentives to performance have been widely implemented in other industries and are a powerful lever to influence behavior. P4P is seen as a way to create a “business case” for quality by better aligning payment with quality of service instead of quantity of service.13,14 This helps address the issue that hospitals and physicians are not necessarily rewarded for delivering high-quality care. For example, hospitals that produce better health outcomes may paradoxically face lower margins through phenomena such as diagnosis-related group switching and a reduction in unplanned rehospitalizations for the chronically ill.19 P4P also holds promise because many of the traditional approaches to improving quality, such as physician education, provider certification, and consumerism, have failed, largely due to the fragmented nature of our healthcare delivery system.20 Linking financial incentives to quality is also viewed by many as a more palatable approach than traditional managed care models where financial incentives are provided to physicians to limit referrals and see more patients per day.21

Concerns About P4P

Despite the promise of healthcare quality, concerns have
been raised that P4P may have potential unintended consequences for patients, physicians, and hospitals. For example, most P4P programs reward providers based on evaluation of a limited number of process performance metrics. If hospitals and physicians become too focused on these metrics, they may lose sight of the global goals of healthcare (analogous to students studying just what is on the test). Some have argued that forcing hospitals and providers to follow select process patterns could stifle innovation and the ability or willingness of organizations to develop creative solutions to improving quality. Others have worried that the large fixed costs required to support P4P data collection and quality improvement programs could deviate important resources away from patient care and have unintended consequences. In a similar manner, the financial incentives in P4P could paradoxically exacerbate healthcare disparities—either by financially penalizing hospitals that treat underserved populations or by prompting caregivers to avoid sick and “high-risk” patients from their practice.

Evidence Regarding the Effectiveness of P4P

While many have speculated on both the potential positives and negatives of P4P programs, the evidence regarding its impacts has been mixed. A systematic review conducted in 2006 identified 17 studies on P4P published between 1980 and 2005. These studies focused primarily on programs targeting preventive care services. The studies tended to be inconclusive due in part to their small sample size, specialized setting, and short-term follow-up. More recently, however, there have been investigations on the impact of large P4P initiatives sponsored by the US federal government and the United Kingdom. These studies have generally found that P4P programs were associated with modest improvements in process of care measures, yet none of these programs had an impact on patient outcomes or efficiency of care. For example, Lindenauer et al found that hospitals engaged in both public reporting and CMS’ HQID P4P program achieved modestly greater improvements in quality than did hospitals engaged only in public reporting. The estimated incremental effect on a composite quality score for cardiac and pneumonia care ranged from approximately 2.8% to 4.3%. Glickman et al evaluated CMS’ HQID program in a cohort of 500 hospitals participating in a voluntary quality improvement initiative for acute myocardial infarction and found that financial incentives were not associated with a significant incremental improvement in quality or reduction in mortality. Evidence from the large P4P program in the United Kingdom for asthma, diabetes, and coronary heart disease suggests that family practitioners achieved high levels of achievement in the first year of the program. Over the ensuing 3 years, however, improvements in the quality of care slowed and actually declined for conditions not linked to financial incentives. Unfortunately, few large-scale, randomized controlled trials have evaluated the effectiveness of P4P interventions.

Issues Involved in the Methodology for P4P Measurement and Provider Ratings

A key challenge to implementing P4P programs is selecting valid and reliable measures of quality and performance. The largest P4P programs to date have focused primarily on processes of care (ie, adherence to evidence-based treatment guidelines). For example, the 2009 CMS Physician’s Quality Reporting Initiative is tracking 153 quality measures spanning multiple therapeutic areas. A number of professional organizations, including the National Quality Forum, the American Heart Association, and the American College of Cardiology developed consensus methodology for the selection and creation of performance measures. Performance measures must be valid (eg, have a robust evidence base supporting their use), accurate and reliable, easy to interpret, and allow for reliable comparisons among providers. Other important considerations are the clinical relevance of the outcome, adherence to the process measure, and variability in baseline adherence to the measure (ie, is there already a ceiling effect in performance?).

Despite the popularity of using process measures to gauge quality, such measures present significant challenges. Selecting performance measures that do not meet evidence-based criteria may negate their intended effects of improved efficiency and patient outcomes. For example, a recent study of quality measures for acute myocardial infarction in 1351 hospitals found that in a resource-constrained environment, hospitals that focus on “administrative” process measures (eg, smoking cessation counseling or discharge instructions) at the expense of clinical interventions (eg, aspirin or angioplasty) have worse patient outcomes. Another issue is the possibility of obtaining stable estimates of performance for small hospitals and physician groups or whether additional techniques are needed to account for small and unequal denominators. A recent study by O’Brien et al demonstrated that high-volume hospitals had better performance on average, but were significantly less likely to be identified as “top” hospitals (ie, top decile). There are also challenges in finding ways to combine performance measures to create a valid aggregate measure of hospital or physician performance. Composite scores that combine several performance measures into a single ranking are commonly used to assess hospital performance. Yet,
Existing methods used to create composite scores are highly variable in their weighting of process versus outcomes metrics, which can, in turn, lead to highly divergent provider rankings. Composite scores are typically weighted by the total number of treatment opportunities, although evidence suggests that weighting based on how hospitals organize care or the range for possible improvement in scores may provide more useful information. Finally, although performance measures are not typically risk-adjusted, a recent study by Mehta et al. found that accounting for hospital case mix (such as age, sex, race, insurance status, and medical comorbidities) can dramatically affect a hospital’s ranking and financial benefits in P4P programs.

In addition to process measures, P4P programs can also reward patient outcomes (eg, inpatient mortality, 30-day mortality, 30-day hospital readmission), patient perceptions of care, measures of efficiency (eg, ratio of observed to expected costs), and hospital and physician group structural characteristics (eg, adoption of information technology). Outcomes performance assessment, however, can be challenging due to the potential confounding influence of patient case mix on provider outcomes, as well as the instability of hospital outcome estimates due to low event rates.

**New Care Models With Financial Incentives**

The shortcomings of many traditional P4P programs, including their limited impact on provider behavior, value, and health outcomes, have fueled the emergence of new and innovative models of payment reform. Primary care capitation models that include large performance and efficiency bonuses are currently being piloted in Massachusetts. Other groups are moving toward episode-based payments, which reimburse providers on the basis of expected costs for clinically defined episodes of care. The Geisinger Health System recently implemented a bundled payment for coronary artery bypass grafting (ProvenCare) that covers all care 30 days before and 90 days after the procedure, including any complications, hospital readmissions, and follow-up care. This model has been associated with a significant reduction in hospital readmissions and hospital charges, and has been expanded to a number of other clinical conditions, including angioplasty and hip replacement surgery. In California, P4P programs have shifted toward efficiency measurements using bundled payment for acute surgical and medical interventions.

One of the most promising payment models to emerge is the accountable care organization (ACO). This concept is rooted in the observation that there are dramatic differences in Medicare spending, by both region and hospital. Regions with lower per beneficiary spending in Medicare achieve equal or better quality and health outcomes than their counterparts. The goal of the ACO model is to create organizational accountability for quality, improve coordination of care for Medicare beneficiaries, and reward innovations that simultaneously improve healthcare quality and reduce costs.

In 2005, Medicare implemented the Physician Group Practice (PGP) demonstration program in large group practices that serve at least 5000 Medicare patients. The program allows group practices to share the cost savings they achieve in caring for their patients if they simultaneously meet quality improvement targets. The cost savings are calculated by comparing actual spending to a target. The target is determined by the PGP’s base year per capita expenditures trended forwarded by a per capita growth rate from the same area. Medicare savings in excess of 2% are shared with CMS depending on whether process performance targets are achieved. In the first 2 years of the program, there was overall improvement in quality-of-care targets and lower risk-adjusted expenditure growth rates for several of the participating groups. A recent study demonstrated that more than 75% of Medicare beneficiaries nationally receive care in ACO-eligible networks (>5000 beneficiaries) and that widespread adoption of ACOs would lead to considerable savings in Medicare spending. Yet, there are important challenges to widespread adoption of ACOs. These include fragmentation of provider efforts by competing P4P programs, applicability of ACOs to patients living in rural and underserved areas, and concerns that a primary focus in efficiency may undermine the credibility of these programs among physicians and other stakeholders.

Another promising option is to provide financial incentives for patients to modify their behavior. Targeting unhealthy patient behaviors such as smoking, poor diet, and physical inactivity may yield important public health benefits and reduce overall healthcare costs. This is because unhealthy patient behaviors may account for as much as 40% of premature deaths in the United States, whereas inefficiencies in healthcare delivery account for only 10%. In a recent study of 878 employees at a large company, the group that received a $750 financial incentive for smoking cessation had a significantly high rate of smoking cessation up to 18 months after enrollment compared with those in the group who received no incentives (9.4% vs 3.6%, respectively). P4P for patients merits additional study, and future programs will need to draw on insights from behavioral economics in order to optimize the structure of these programs.

**Future Directions**

P4P and newer models of payment reform hold tremendous promise to improve healthcare quality and reduce costs.
Multiple approaches are needed to ensure that future initiatives provide value for key stakeholders, including patients, providers, and payers. In general, the goal should be to foster innovative approaches to improving quality patient outcomes, promote accountability for quality on the part of both patients and providers, and improve the value of services purchased by federal and private payers.

It is important to continue to build consensus about the selection of evidence-based performance measures, including measures that reward technical aspects of care, outcomes, efficiency, and patient-centered care. Given the traditionally long lag time between the translation of new evidence into clinical practice, it will be critical to find ways to rapidly incorporate validated process measures into everyday use. Greater focus on outcome and efficiency measures appears warranted. Further refinement of risk-adjustment methods is needed to standardize outcome and cost-efficiency metrics and allow for valid benchmarking across providers.

The creation of a more compelling “business case” and value proposition for quality on the part of key stakeholders is essential to creating sustainable quality improvement efforts. Adoption of well-designed patient outcome and efficiency measures would encourage providers and hospitals to develop innovative solutions for improving value. In contrast, promulgation of too many measures that focus on technical aspects of care may distract providers and lead to fragmented care. As we transition toward greater use of outcome and efficiency measures, additional investments in comparative effectiveness and health services research will be needed to help payers and providers identify and implement higher-quality, more cost-effective treatment approaches. These efforts would be facilitated by greater investments in large quality improvement registries and patient information systems.

P4P programs will need to be sensitive to hospitals and provider groups that care for vulnerable patient populations, including the underinsured, racial and ethnic minorities, and patients living in rural areas. In addition, further efforts are needed to better understand hospital structural characteristics that facilitate high-quality healthcare delivery, including the role of information technology, management, culture, and organizational structure.

Conclusion

P4P and newer models that link reimbursement with quality and efficiency show promise to improve patient outcomes and lower costs. Future success hinges on collaboration among key stakeholders including patients, physicians, payers, and policymakers.

Author Affiliations: From the Duke Clinical Research Institute and the Department of Medicine, Duke University Medical Center, Durham, NC (SWG, EDP), and the Department of Emergency Medicine, University of North Carolina School of Medicine, Chapel Hill (SWG).

Funding Source: Financial support for this work was provided by GlaxoSmithKline.

Author Disclosure: The authors (SWG, EDP) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (SWG, EDP); analysis and interpretation of data (EDP); drafting of the manuscript (SWG); and critical revision of the manuscript for important intellectual content (SWG, EDP).

Address correspondence to: Eric D. Peterson, MD, MPH, Outcomes Research and Assessment Group, Duke Clinical Research Institute, PO Box 17969, Durham, NC 27715. E-mail: peter016@mc.duke.edu.

REFERENCES


