

# MARKET WATCH

## Climbing Up The Pay-For-Performance Learning Curve: Where Are The Early Adopters Now?

P4P programs among these early adopters have persisted, but rigorous evaluation is needed to fully assess their impact.

by **Meredith B. Rosenthal, Bruce E. Landon, Katherine Howitt, HyunSook Ryu Song, and Arnold M. Epstein**

**ABSTRACT:** The diffusion of performance-based payment incentives is arguably the most striking change in the U.S. health care system since the managed care era. Because there is little knowledge about best practices, sponsors of payment-incentive programs must learn by doing. We examine the experiences of twenty-seven early adopters and profile the evolution of their pay-for-performance (P4P) strategies as well as perceptions of key lessons learned. Our findings suggest that leading-edge sponsors of P4P have expanded the reach of their efforts, particularly with regard to specialists, and increasingly are focused on outcome and cost-efficiency measures, rather than clinical process measures alone. [*Health Affairs* 26, no. 6 (2007): 1674–1682; 10.1377/hlthaff.26.6.1674]

PERFORMANCE-BASED PAYMENT incentives are now routinely used by both private and public payers in the U.S. health care system.<sup>1</sup> Aspirations that payment incentives will serve as the linchpin of delivery system reform are widely held, as reflected in the recent endorsement by the Institute of Medicine of pay-for-performance (P4P) for Medicare.<sup>2</sup> Although adoption of P4P has been rapid, there remains considerable uncertainty about how best to design and implement it.<sup>3</sup> The few formal evaluations of individual P4P programs published to date have provided some insight into these programs' potential effectiveness, but they paint a limited and static picture.<sup>4</sup>

In 2003 we identified and examined a cross-section of P4P programs using publicly available data.<sup>5</sup> At that time, many of these programs were only recently initiated, and our analysis showed variation in the ways in which they were constructed, on several dimensions, which we hypothesized would lead to differences in their effects. In this study we revisit these "early adopters" to map their progress on the P4P learning curve. Our earlier paper presented a taxonomy of programs; in revisiting the subjects of that earlier work, we sought to examine the following: (1) how P4P strategies have evolved, (2) what lessons the early adopters have learned, and (3) whether there was any qualitative support for our earlier hypoth-

.....  
*Meredith Rosenthal (mrosenth@hsph.harvard.edu) is an associate professor of health economics and policy, Department of Health Policy and Management, Harvard School of Public Health, in Boston, Massachusetts. Bruce Landon is an associate professor, Department of Health Care Policy, at Harvard Medical School, also in Boston. Katherine Howitt is a research assistant in the Department of Health Policy and Management. HyunSook Ryu Song is a consultant for the Partners Center for Connected Health in Boston. Arnold Epstein is a professor and chair of the Department of Health Policy and Management.*

eses about the effects that different types of programs would have.

## Study Data And Methods

■ **Sample and data collection.** Thirty-one sponsors of P4P programs were included in our original study and served as the starting point for our analysis.<sup>6</sup> The thirty-one programs included all operational performance-based provider incentive programs that could be identified in 2003 through keyword searches using Lexis/Nexis and Google. We excluded from follow-up two programs, the CMS/Premier Inc. demonstration and the Local Initiative Rewarding Results program in California, because they were explicitly time-limited. In another case, two of the original sponsors—Blue Cross Blue Shield of Missouri and Anthem Blue Cross and Blue Shield in the Midwest (plans in Indiana, Kentucky, and Ohio)—merged to become part of a single region of WellPoint Inc. Finally, one plan sponsor, HealthGuard of Pennsylvania, ceased operations at the end of 2006. For each of the remaining twenty-seven sponsors, we identified respondents, typically medical directors, directors of quality management, or executive directors in the case of health care coalitions.

Between January and July of 2006 we undertook a series of structured telephone interviews to elicit information about the existence and scope of respondents' current P4P programs (provider types, products); types of measures targeted; amounts at risk; changes in design of the programs and payment algorithms; and what they knew about the early effects of their programs on quality, patients' experiences, costs, and provider satisfaction. With regard to program results, we asked respondents whether they had seen evidence of major quality improvement as a result of the program, had found no improvement, or were uncertain of impact. We also asked whether they had observed any negative consequences from the program, including providers' avoidance of sick or nonadherent patients. Our final characterization of each program was verified with respondents to ensure accuracy.

We report two types of data in this study:

(1) the prevalence of specific program characteristics (for example, percentage of plans paying on measures of outcomes), and (2) opinions of P4P leaders (for example, on the future of P4P). We report frequencies related to program characteristics on an enrollment-weighted basis.<sup>7</sup> Results did not differ qualitatively from the unweighted data. When reporting subjective information, however, we report only the number of respondents reporting a point of view, since we have no basis for attributing respondents' views to their entire organization.

■ **Hypotheses revisited.** We also revisited three key hypotheses that we developed in the earlier study: that payers with greater market share and larger incentives would find P4P more effective; that P4P as currently implemented would lead to concerns about simply paying more to providers who were already of high quality; and that P4P would increase incentives to avoid more-challenging patients.

## Study Results

■ **P4P has taken root.** We interviewed respondents for all twenty-seven sponsors in our final sample and found that all but three were still engaged in P4P in 2006 (Exhibit 1). Of those three programs, one was canceled because of the perception that its market share was too small for the payment incentives to influence the targeted physicians. In another, the provider organization that had been the locus of P4P left the health plan's network. The third was terminated after a three-year pilot, and although it was anticipated that P4P would continue in some form, no new program was in place yet.

The sponsors of the twenty-four programs profiled in Exhibit 1 varied in size from 52,000 to eleven million enrollees and were geographically diverse. As we noted in 2004, seven of the health plans participated in the Integrated Healthcare Association (IHA) P4P initiative in California and collectively represent the largest program in the United States.

■ **Reaching out to new provider types.** Although primary care physicians continue to be the most common provider type subject to

### EXHIBIT 1 Characteristics Of The Sample Of Twenty-Seven Pay-For-Performance (P4P) Program Sponsors, 2006

Sponsor/program	Number of enrollees affected	Providers included	Total payments in 2005	Payment per enrollee (\$)
Aetna (CA)	250,000	PCP, S	\$5.4 million	22
Anthem Blue Cross Blue Shield of NH, CT, ME	– <sup>a</sup>	PCP, H	– <sup>a</sup>	– <sup>a</sup>
Anthem Blue Cross Blue Shield (VA)	2.9 million	PCP, S, H	\$12 million	4
Blue Cross Blue Shield of IL	850,000 <sup>b</sup>	PCP, S, H	\$153 million <sup>c</sup>	180
Blue Cross Blue Shield of MA	Over 2 million	PCP, S, H	\$94 million	47
Blue Cross Blue Shield of MN	– <sup>a</sup>	PCP, S	– <sup>a</sup>	– <sup>a</sup>
Blue Cross Blue Shield of MI	Over 4 million	PCP, S, H	\$45 million	11
(Excellus) Blue Cross Blue Shield of Rochester (and RIPA)	250,000	PCP, S	\$15 million	230
Blue Cross of CA	1.4 million	PCP, S	\$66 million	47
Blue Shield of CA	1.2 million	PCP, S	\$26 million	15
Bridges to Excellence	250,000	PCP, S	\$3.9 million	16
Buyers Health Care Action Group	226,000	PCP, S	– <sup>a,d</sup>	0.5
CIGNA (CA)	290,000	PCP, S	\$4 million	14
Employers' Coalition on Health	52,000	PCP, H	\$106,000	2
Harvard Pilgrim Health Care (MA)	– <sup>a</sup>	PCP, H	– <sup>a</sup>	– <sup>a</sup>
Hawaii Medical Service Association	500,000	PCP, S, H	\$23 million	19
HealthNet	1.2 million	PCP, S	\$19 million	29
HealthPartners (MN)	660,000	PCP, S, H	\$11.2 million	4
Highmark Blue Cross and Blue Shield of PA	2.5 million	PCP, S, H	– <sup>a</sup>	– <sup>a</sup>
Independence Blue Cross (PA)	878,000	PCP, H	\$35 million	40
Independent Health (NY)	252,000	PCP, S	\$3.5 million	3
PacifiCare (CA)	1.3 million	PCP, S	\$14.2 million	1
Western Health Advantage (CA)	65,000	PCP, S	\$408,000	6
WellPoint's Central Region (IN, OH, KY, MO, WI)	11 million	PCP, S, H	– <sup>a</sup>	– <sup>a</sup>

**SOURCE:** Authors' interviews with sponsor representatives, 2006.

**NOTES:** PCP is primary care physician. S is specialist physician. H is hospital.

<sup>a</sup> Not available.

<sup>b</sup> The Blue Cross Blue Shield of Illinois physician P4P program covers the 850,000 enrollees in the health maintenance organization (HMO) product. Including the hospital P4P program, which covers members of all Blue Cross Blue Shield of Illinois products, the number of enrollees affected is 3.4 million.

<sup>c</sup> This number only describes the payouts for the HMO portion of the P4P program.

<sup>d</sup> In 2005, Buyers Health Care Action Group was still planning its new P4P program and defining the parameters for payment, so it did not pay out any money. In 2006 it paid out \$100,000 through its P4P program.

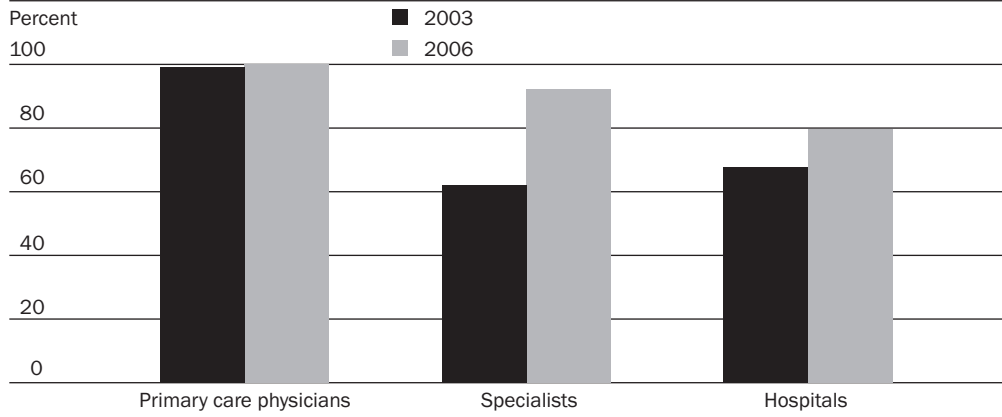
P4P, the inclusion of specialists increased between 2003 and 2006 (Exhibit 2). Cardiologists and general surgeons were the most commonly mentioned specialists for inclusion in P4P programs, but gastroenterologists and orthopedists were also mentioned by several sponsors.

Sponsors cited measurement issues as the largest barrier to the inclusion of specialists. Three respondents mentioned the paucity of nationally accepted quality measures for spe-

cialists as the biggest obstacle, and another explained that attributing patients' receipt of recommended care or outcomes to specific physicians is more difficult with specialists than with primary care physicians.

■ **Evolving measurement.** The performance measures that form the basis for evaluating and rewarding physicians and hospitals have also changed over time. Although many programs began with a focus on widely accepted process measures of quality (for exam-

**EXHIBIT 2**  
**Percentage Of Enrollees Covered By Pay-For-Performance (P4P) Sponsors Covering Primary Care Physicians, Specialists, And Hospitals, 2003 And 2006**



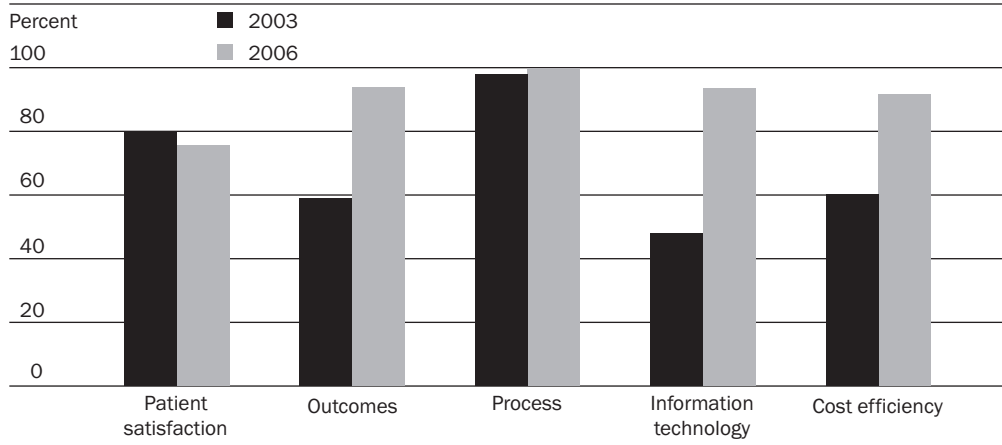
**SOURCE:** Authors' interviews with sponsor representatives, 2003 and 2006.

ple, mammography rates), outcomes, cost efficiency, and information technology (IT) measures have become increasingly common program elements (Exhibit 3).

In 2003, sponsors representing 60 percent of enrollees targeted cost efficiency, but by 2006, sponsors representing 92 percent of enrollees were doing so. Some sponsors included cost-efficiency measures as a prominent aspect of their P4P payment schemes (comparative cost of care accounted for 50 percent of

one plan's physician's P4P score, for example, and 20 percent of another's); however, most incorporated smaller rewards for a narrowly defined measure such as generic prescribing or formulary compliance. Our interviews also revealed a sharp increase in outcome measures: Although sponsors representing 59 percent of enrollees targeted health outcomes in 2003, by 2006 that number had increased to 94 percent. For physicians and medical groups, the most commonly targeted outcome measures were

**EXHIBIT 3**  
**Percentage Of Enrollees Covered By Pay-For-Performance (P4P) Sponsors That Use Specific Measure Sets, 2003 And 2006**



**SOURCE:** Authors' interviews with sponsor representatives, 2003 and 2006.

intermediate outcomes such as HbA1c, LDL cholesterol, and blood pressure control. For hospitals, complication and in-hospital mortality rates were frequently targeted. With some exceptions, the measures incorporated into the early adopters' P4P schemes reveal a focus on chronic illness treatment guidelines and preventive medicine. In particular, all of the sponsors we interviewed incorporated indicators of compliance with diabetes or asthma care guidelines, and most also included process measures aimed at recommended preventive services.

Nearly every surviving program (covering 99 percent of enrollees) has increased the total number of measures that factor into the calculation of performance bonuses or withholds since 2003. In addition to supplementing measures, respondents representing 33 percent of enrollees reported that they had eliminated measures from their programs. For instance, patient satisfaction scores were dropped by three sponsors representing 17 percent of enrollees. Two of those three sponsors reported that they had done so because of a lack of variation in scores across providers; a third sponsor did so because of the expense of collecting patient survey data. Some payers also reported eliminating the following measures because the scores were consistently very high: counseling for tobacco cessation; well-baby and well-child visits; mammography; cervical cancer screening; colorectal cancer screening; and measles, mumps, and rubella (MMR) vaccination.

■ **More money.** The majority of programs (covering 58 percent of enrollees) have augmented the pool of money available for performance-based pay, even after accounting for the fact that more providers are often being drawn into the programs. P4P bonuses were typically about \$1.40 per member per month and ranged from twenty cents to fifteen dollars per member per month.<sup>8</sup> Thus, despite reported increases since 2003, P4P remains a very small

portion of total payments. For example, the average physician bonus in our sample would translate into approximately 2.3 percent of reimbursement.<sup>9</sup>

■ **Technical aspects of reward calculations.** Respondents representing 54 percent of enrollees were using risk adjustment in their performance profiling, while other plans used stratification as a means of accounting for the variation in the populations served by different providers.<sup>10</sup> Several respondents noted that

risk-adjustment formulas were on the agenda for development. For example, the IHA in California is reviewing risk-adjustment strategies and expects to move in this direction in the next several years.<sup>11</sup>

In terms of payment formulas, sponsors representing 23 percent of enrollees have begun to reward improve-

ment in addition to attainment of performance goals, in contrast to our findings three years ago, when none of the P4P programs in our sample targeted improvement.

■ **Evidence of impact.** Among those in our study, programs representing only 7 percent of enrollees have been formally evaluated by independent researchers. Of the three programs that were formally evaluated, one evaluation was still ongoing (with no results currently available), and another found improvements in hospital process and outcome measures (but this evaluation did not include a control group). The third study found statistically significant improvements in diabetes, mammogram rates, and coronary disease measures compared with a control group but was unable to determine whether to attribute the improvements to the incentives themselves or to the education and direction that accompanied the rollout of this particular P4P program.

Several respondents reported that it was still too early to expect changes in performance. It was also widely acknowledged that the dynamic nature of P4P arrangements, cou-

**“Among those in our study, programs representing only 7 percent of enrollees have been formally evaluated by independent researchers.”**

pled with contemporaneous shifts in benefit design, public reporting, and other aspects of the health system, makes the identification of the true impact of P4P challenging, if not impossible. Where some type of evaluation had been undertaken, outcomes were arguably positive. Respondents covering 38 percent of enrollees reported solid gains, with another 42 percent finding mixed results and 20 percent finding no effect. Clinical areas where improvement was documented included diabetes care, cancer screening, and inpatient cardiac care. Causality cannot be determined from our cross-sectional, qualitative data; however, we note that as we had predicted in our previous paper, programs reporting a positive impact from P4P had higher market shares and offered higher rewards as a percentage of providers' pay.<sup>12</sup>

Lacking strong evidence of impact on quality improvement, most respondents reported that their programs were sustained instead by at least one of three motivations. The first is a belief that if P4P is not yet improving quality, it is because they have yet to find the right technical specification and that if they keep tweaking the programs—by adding money, coordinating with other payers' P4P programs, or changing the targeted measures—they will eventually formulate an effective program. The second motivator is a sense that even if P4P is not helping improve quality, paying more for higher quality is simply fairer than paying solely for the quantity of services provided. The third motivator is a desire to use P4P as an intermediary step toward other goals, such as making performance transparent to consumers and purchasers or developing a tiered payment system.

■ **Negative effects.** Our previous analysis had led to hypotheses about the potential for unintended consequences such as impaired access for certain patient groups and payments rewarding primarily groups that had already attained high-quality care. However, we found little support for those two hypotheses: Most sponsors (representing 81 percent of enrollees) had not observed such effects, although none had looked for them rigorously.

Two respondents, however, expressed concerns about patient dumping, despite having no clear evidence of this. Representatives of two other plans also voiced concern that payments had been made to providers who were already high-performing without the realization of quality improvement. Finally, one payer noted that a provider group in its network had dissolved in a dispute over how to distribute performance bonuses.

■ **Views on the future of P4P.** All twenty-four sponsors that were still engaged in P4P in 2006 intended to continue with this approach in the near term. Indeed, fourteen of them anticipated expanding the use of P4P across their provider networks, and four explicitly mentioned plans to increase the share of payments to be allocated based on performance. There were two distinct and opposing camps, however, with regard to views on the long-term role that P4P should play. One group expected that some form of P4P would become a permanent feature of reimbursement systems. The other group took the view that the regulatory model of P4P would cede to a health care market where both consumers and payers would be able to distinguish high-value from low-value providers and that the former would command a price premium. This latter group affirmed a strong belief in the power of publicly reported provider performance data. In fact, three respondents identified transparency of performance data as a more important lever for performance improvement than payment incentives. In contrast, four respondents expressed serious doubts about the public's ability to understand health care quality and efficiency reports.

■ **What is needed to make P4P work.** Respondents identified three major challenges to building and maintaining effective P4P programs: (1) overcoming physician resistance, (2) determining the necessary size of incentive pools to capture providers' attention, and (3) finding the resources necessary to continue funding the programs.

When asked what lessons their P4P experience had taught them, some respondents focused on the importance of promoting pro-

vider involvement as a means of reducing opposition to the programs. Nine sponsors specifically noted that it was critical to involve physicians and hospitals early in the development of P4P and to maintain communication and program transparency. Five also emphasized the need to start small and build trust and capacity for measurement and quality improvement over time.

The second most prevalent lesson noted was in the area of selection of measures. Four respondents stressed the importance of using clinical rather than administrative data, in part to overcome physicians' concerns about the validity of performance measurement. However, most programs still relied on administrative measures. Four sponsors also cited the need to use only nationally accepted measures such as those approved by the National Quality Forum (NQF), both to satisfy participating physicians and to permit coordination with other programs. Two respondents, both affiliated with IHA, mentioned the importance of collaborating with other plans to build common P4P schemes, to increase the market impact on providers.

Five respondents noted that the most serious threat to the sustainability of P4P was the absence of a demonstrable return on investment (that is, evidence of net savings). Although only three sponsors had documented a return on investment, another five stressed that quantifying a return on investment was a future goal. These financial concerns may be reflected in the substantial increase in payers' incorporating efficiency measures into their programs (see Exhibit 3).

## Discussion

Our findings suggest that there has been substantial evolution in the P4P programs established by early adopters. Payers have increased both the size of the financial incentives

and the numbers of measures on which they are based. Programs more frequently incorporate indicators of health outcomes and efficiency. Risk adjustment is slowly becoming more prevalent. In addition, we found evidence of increasing inclusion of specialists in P4P, although we cannot quantify their importance relative to primary care physicians in terms of dollars at risk. Despite the persistence and growth in P4P among the payers we surveyed, evaluation has been sparse among early

adopters, with the initial results showing an inconsistent impact on quality improvement and very little evidence of cost savings, but no important deleterious effects.<sup>13</sup>

■ **Implications.** These findings yield several important implications for policy, practice, and research. First, as early adopters have increased levels of payment and migrated to health outcomes and costs as targets, there is a

need to incorporate an evaluative component in programs' designs. Program evaluation could assess both the impact of the programmatic changes on targeted measures and unintended, adverse consequences. With increasing focus on outcomes and cost, we also need better risk adjustment to address real and perceived concerns about patient differences that might undermine quality incentives.

Second, many early adopters of P4P have made efforts to strengthen their programs by increasing the size of the incentive pool, although such increases remain modest, judging from the dollars paid out. Some sponsors have also begun to reward improvement explicitly, alongside attainment of benchmark levels of performance. Although it is not a foregone conclusion that these changes will make P4P a major force for change, they will in theory improve the incentives for more providers to change their behavior. Further information is needed, however, to understand the level and distribution of rewards on a per measure basis and to determine the practical significance of

**“Several respondents noted that P4P was an inherently more sensible way to pay for care than capitation or volume alone, even if there was no effect on behavior.”**

current schemes.

Finally, expansion in the comprehensiveness of measure sets among early adopters means that there is less latitude for providers to focus on a single population group or condition to maximize P4P payments. Theoretically, this trend should motivate more holistic approaches to quality improvement, which are viewed by many as critical to making real progress in achieving national quality goals.<sup>14</sup> Despite the increasing scope of measurement, the range of measures and conditions covered by the P4P programs we studied remains relatively narrow, largely because of the limits of current measure sets. P4P may become important in those clinical domains for which there is (or will be) sufficient evidence to support meaningful process or outcome measurement. However, it should probably be acknowledged that some areas of medicine—for example, where patients’ preferences greatly affect the appropriate course of treatment—may never be well suited to performance incentives.

The commitment to P4P among these early adopters is apparent in the fact that all but three had retained their programs and the vast majority had extended P4P to more of their networks. Although one might question this faith in the future value of P4P based on the current state of the evidence, it should be noted that prior studies do not fully address the potential of P4P because of the relatively limited nature of the experiments studied.<sup>15</sup> Moreover, several respondents noted that P4P was an inherently more sensible way to pay for care than capitation or volume alone, even if there was no effect on behavior. Of course, these findings might not generalize to other payers, who might be more skeptical of the potential benefits of P4P.

■ **Detrimental forces.** Our findings also highlight developments that could undermine P4P. Many respondents were focused on making the business case for P4P—a concern that is clearly driven by employers that are alarmed at the continued pace of rising health insurance premiums. Although efficiency is an important component of quality, emphasis on reducing the cost of care may ultimately

undermine the credibility of these programs with physicians and other stakeholders.

■ **Strengths and weaknesses of study and data.** Our findings should be considered in light of the strengths and weaknesses of our research design and data. The principal strength of the study is that we were able to track twenty-seven early adopters of P4P three years after we first identified them and could obtain detailed information about the changes they had made in their programs and what key outcomes and lessons they had observed.

The study has two main limitations. First, we relied on a small, nonrandom sample of systematically identified payers. These early adopters may have motivations or other characteristics that limit the generalizability of our findings to other sponsors of P4P. Second, we lacked objective data to measure the effectiveness and unintended consequences of the programs we studied.

AS CONGRESS AND the Centers for Medicare and Medicaid Services (CMS) prepare to incorporate P4P into the Medicare payment system, the early adopters continue to refine their programs. The most notable change is the shift from paying on a few, nationally accepted process measures of care to focus more on health care outcomes and the efficient delivery of care. With the latest published results on P4P showing only modest effects on quality, it will be of critical importance to evaluate whether these changes yield dividends in terms of improved quality and cost control.<sup>16</sup>

.....  
*Financial support for this paper was generously provided by the Program for Health Systems Improvement at Harvard University.*

## NOTES

1. K. Kuhmerker and T. Hartman, "Pay-for-Performance in State Medicaid Programs: A Survey of State Medicaid Directors and Programs" (New York: Commonwealth Fund, 2007); S. Trude, M. Au, and J.B. Christianson, "Health Plan Pay-for-Performance Strategies," *American Journal of Managed Care* 12, no. 9 (2006): 537-542; and M.B. Rosenthal et al., "Pay for Performance in Commercial HMOs," *New England Journal of Medicine* 355, no. 18 (2006): 1895-1902.
2. Institute of Medicine, *Rewarding Provider Performance: Aligning Incentives in Medicare* (Washington: National Academies Press, 2007).
3. Leapfrog Group, "Incentive and Reward Compendium Guide and Glossary," 2005, <http://ir.leapfroggroup.org/compendium> (accessed 31 July 2007); G. Baker and B. Carter, "Provider Pay-for-Performance Programs: 2004 National Study Results" (San Francisco: Medvantage, 2005); and Rosenthal et al., "Pay for Performance."
4. R.A. Dudley, "Pay-for-Performance Research: How to Learn What Clinicians and Policy Makers Need to Know," *Journal of the American Medical Association* 294, no. 14 (2005): 1821-1823; and L.A. Petersen et al., "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145, no. 4 (2006): 265-272.
5. M.B. Rosenthal et al., "Paying for Quality: Providers' Incentives for Quality Improvement," *Health Affairs* 23, no. 2 (2004): 127-141.
6. *Ibid.*
7. We were unable to attain 2003 enrollment data for five plans and 2005 enrollment data for three plans in our sample. We calculated weights for these plans by assigning them an estimated enrollment figure equal to the health maintenance organization (HMO) total enrollment obtained from the 2003 or 2004 InterStudy Competitive Edge databases, respectively.
8. Although we would ideally capture the amount of funds "at risk" or available for bonus payments under P4P, most respondents were unable or unwilling to provide that figure but instead provided us with the amount paid out in the most recent year, 2005.
9. We obtained an average capitation rate per member per month for physicians in 2005 (\$61.69) from the 2007 Capitation Survey (Healthcare Compliance Company). To obtain the average physician bonus as a percentage of average physician pay, we divided the average physician bonus per member per month (obtained from our data) by this physician capitation rate per member per month.
10. The use of process measures includes a form of implicit risk adjustment, since denominators for these measures are defined based on patient characteristics, although there may still be some argument for risk adjustment.
11. Integrated Healthcare Association, "Advancing Quality through Collaboration: The California Pay-for-Performance Program," February 2006, <http://www.iha.org/wp020606.pdf> (accessed 31 July 2007).
12. Using a two-tailed student's t-test, we tested for equivalence between the market shares for plans reporting effective programs and the market shares for plans not reporting effective programs. The *p* value was 0.026. We also tested for equivalence between the incremental revenues possible for providers through P4P in 2004 for plans reporting effective programs and the incremental revenues possible through P4P for providers in 2004 for plans not reporting effective programs. The *p* value for that test was 0.058.
13. Rosenthal et al., "Pay for Performance."
14. IOM, *Crossing the Quality Chasm: A New Health System for the Twenty-first Century* (Washington: National Academies Press, 2001).
15. M.B. Rosenthal and R.G. Frank, "What Is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review* 63, no. 2 (2006): 135-157.
16. P.K. Lindenauer et al., "Public Reporting and Pay for Performance in Hospital Quality Improvement," *New England Journal of Medicine* 356, no. 5 (2007): 486-496; A.M. Epstein, "Pay for Performance at the Tipping Point," *New England Journal of Medicine* 365, no. 5 (2007): 515-517; and G.J. Young et al., "Effects of Paying Physicians Based on Their Relative Performance for Quality," *Journal of General Internal Medicine* 22, no. 6 (2007): 872-876.