



Online article and related content  
current as of July 27, 2008.

## Early Experience With Pay-for-Performance: From Concept to Practice

Meredith B. Rosenthal; Richard G. Frank; Zhonghe Li; et al.

*JAMA*. 2005;294(14):1788-1793 (doi:10.1001/jama.294.14.1788)

<http://jama.ama-assn.org/cgi/content/full/294/14/1788>

Correction

[Contact me if this article is corrected.](#)

Citations

[This article has been cited 133 times.](#)  
[Contact me when this article is cited.](#)

Topic collections

Quality of Care; Quality of Care, Other  
[Contact me when new articles are published in these topic areas.](#)

Related Articles published in  
the same issue

Pay-for-Performance Research: How to Learn What Clinicians and Policy Makers  
Need to Know  
R. Adams Dudley. *JAMA*. 2005;294(14):1821.

Subscribe

<http://jama.com/subscribe>

Permissions

[permissions@ama-assn.org](mailto:permissions@ama-assn.org)

<http://pubs.ama-assn.org/misc/permissions.dtl>

Email Alerts

<http://jamaarchives.com/alerts>

Reprints/E-prints

[reprints@ama-assn.org](mailto:reprints@ama-assn.org)

# Early Experience With Pay-for-Performance From Concept to Practice

Meredith B. Rosenthal, PhD

Richard G. Frank, PhD

Zhonghe Li, MA

Arnold M. Epstein, MD, MA

**T**HE NUMBER OF HEALTH PLANS and purchasers in the United States that have adopted pay-for-performance mechanisms for quality improvement is growing rapidly.<sup>1-3</sup> However, most of these programs are in the early stages of trial, evaluation, and adjustment. Although there is intense interest in and optimism about pay-for-performance programs among many policy makers and payers, there is little published research on pay-for-performance in health care.<sup>4,6</sup> In fact, there are only a few studies demonstrating that pay-for-performance leads to improved quality of care.<sup>7-10</sup>

One area that is particularly controversial is whether to reward providers (ie, hospitals, medical groups, and/or physicians depending on the program) according to attainment of a predetermined level of performance or according to improvement. Paying according to the level of performance is common to the majority of pay-for-performance programs.<sup>1</sup> Critics, however, have worried that physicians or hospitals that have historically performed above the targeted level will have no incentives to improve because they can receive the bonus simply for maintaining the status quo.<sup>1</sup> Moreover, providers whose performance is initially much below the target may have weak incentives to at-

**For editorial comment see p 1821.**

**Context** The adoption of pay-for-performance mechanisms for quality improvement is growing rapidly. Although there is intense interest in and optimism about pay-for-performance programs, there is little published research on pay-for-performance in health care.

**Objective** To evaluate the impact of a prototypical physician pay-for-performance program on quality of care.

**Design, Setting, and Participants** We evaluated a natural experiment with pay-for-performance using administrative reports of physician group quality from a large health plan for an intervention group (California physician groups) and a contemporaneous comparison group (Pacific Northwest physician groups). Quality improvement reports were included from October 2001 through April 2004 issued to approximately 300 large physician organizations.

**Main Outcome Measures** Three process measures of clinical quality: cervical cancer screening, mammography, and hemoglobin A<sub>1c</sub> testing.

**Results** Improvements in clinical quality scores were as follows: for cervical cancer screening, 5.3% for California vs 1.7% for Pacific Northwest; for mammography, 1.9% vs 0.2%; and for hemoglobin A<sub>1c</sub>, 2.1% vs 2.1%. Compared with physician groups in the Pacific Northwest, the California network demonstrated greater quality improvement after the pay-for-performance intervention only in cervical cancer screening (a 3.6% difference in improvement [ $P = .02$ ]). In total, the plan awarded \$3.4 million (27% of the amount set aside) in bonus payments between July 2003 and April 2004, the first year of the program. For all 3 measures, physician groups with baseline performance at or above the performance threshold for receipt of a bonus improved the least but garnered the largest share of the bonus payments.

**Conclusion** Paying clinicians to reach a common, fixed performance target may produce little gain in quality for the money spent and will largely reward those with higher performance at baseline.

*JAMA.* 2005;294:1788-1793

www.jama.com

tempt to improve their performance when the target seems infeasible to reach. On the other hand, paying for improvement may fail to reward the best providers for whom improvement is likely to be substantially more difficult because of ceiling effects.

We evaluated a natural experiment in pay-for-performance conducted within one of the nation's largest health plans, PacifiCare Health Systems. In 2003, PacifiCare began paying its California medical groups bonuses according to meeting or exceeding 10 clinical and service quality targets. We

examined the performance of California medical groups, which were subject to pay-for-performance, and a contemporaneous comparison group in the Pacific Northwest (Oregon and Washington) over time to address 3 specific

**Author Affiliations:** Department of Health Policy and Management, Harvard School of Public Health (Drs Rosenthal and Epstein and Ms Li), Division of General Medicine, Brigham and Women's Hospital (Dr Epstein), and Department of Health Care Policy, Harvard Medical School and National Bureau of Economic Research (Dr Frank), Boston, Mass.

**Corresponding Author:** Meredith B. Rosenthal, PhD, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115 (meredith\_rosenthal@harvard.edu).

questions: What changes in clinical quality of care were associated with the adoption of pay-for-performance? How much did the plan pay out in performance bonuses? How were the rewards distributed across the network relative to quality improvement?

## METHODS

In California, PacifiCare contracts with approximately 300 large multispecialty physician organizations that provide treatment for an average of approximately 10 000 PacifiCare enrollees each, among other patients (PacifiCare represents roughly 15% of the patients treated by the average group), typically under capitation arrangements that cover all professional services.<sup>11</sup> Since 1993, PacifiCare of California has measured the performance of affiliated medical groups on a battery of clinical and patient-reported measures of quality. The information has been reported to medical group leaders to prompt quality improvement and, since 1998, has been made public to encourage selection of high-quality groups by consumers.

The same set of performance measures has been tracked and fed back to a set of 42 medical groups in the Pacific Northwest that serve as PacifiCare's network there. As in California, PacifiCare enrollees in the Pacific Northwest have access to a public report card, which allows them to compare medical group performance on a set of domains that reflect clinical quality, service quality, and affordability. In our analyses, we relied on the Northwest network to serve as a contemporaneous comparison group for the California groups, which were subjected to the pay-for-performance scheme. We assessed the comparability of the physician groups in the Northwest for the purposes of this quasi-experiment by comparing baseline trends in performance because the outcome of interest is a change in performance levels. Although average levels of performance differed between the California and Pacific Northwest networks with regard to the 3 measures compared in

this study, there were no statistically significant differences in the trends between the 2 networks before the quality incentive program (QIP). Thus, the central assumption of the difference-in-differences approach is supported by our data. By comparing quality improvement in the California network with that in the Pacific Northwest where pay-for-performance was not introduced, we were able to determine secular trends in quality to identify the impact of the pay-for-performance program.

### The Pay-for-Performance Program

In early 2002, PacifiCare announced a new QIP for its California network. The program was incorporated into contracts with most groups by July 2002, and these contracts became effective beginning in January 2003. Eligibility for the QIP in the first year was based on having a minimum of 1000 PacifiCare Commercial and 100 Secure Horizons (Medicare Advantage) members. When the first awards were paid in July 2003, 163 California physician groups met these eligibility criteria. The QIP targeted 5 ambulatory care quality indicators and 5 patient-reported measures of service quality (adapted from the Consumer Assessment of Health Plan Survey), as well as a set of hospital quality measures (the groups were rewarded essentially for referring their patients to high-quality hospitals). We focused on 3 of the measures of clinical quality for which complete data were available before and after the QIP in both settings: rates of cervical cancer screening, mammography, and hemoglobin A<sub>1c</sub> (HbA<sub>1c</sub>) testing for diabetic patients. All 3 measures use the Health Plan Employer Data and Information Set (HEDIS) specification.

The performance targets were set at the 75th percentile of 2002 performance by the physician groups and were made known in advance to the participating physician organizations. Because the plan had been feeding back quarterly performance information to physician groups, all participants also had a record of their own perfor-

mance data. Beginning in July 2003, participants received a quarterly bonus of approximately \$0.23 per member per month for each performance target that was met or exceeded. For example, a physician group with 10 000 continuously enrolled plan members (roughly equal to the average group) that reached 1 target would receive approximately \$6900 ( $\$0.23 \times 10\,000$  members  $\times 3$  months) per quarter, or \$27 600 per year for that target. The overall potential for a group with 10 000 PacifiCare patients would thus be about \$270 000 per year for perfect performance. The bonus potential represents about 5% of the professional capitation paid by the plan and about 0.8% of the groups' overall revenue on average. Although bonus payments are calculated and distributed quarterly, performance is assessed according to a rolling year of data (or multiple years as appropriate for measures such as mammography) with a 6-month lag. For example, for payment in July 2003, the HbA<sub>1c</sub> testing measure was based on treatment provided between January 1, 2002, and December 31, 2002; payment in September 2003 was based on treatment provided between April 1, 2002, and March 31, 2003.

The measures and targets for the QIP remained unchanged through the April 2004 payout, after which a new QIP regimen took effect. In the second round of the QIP, some new measures were added, others were altered, and the formula for computing the bonus changed slightly; a second tier of performance was added to induce improvement among the best performers. The 3 measures in this study, however, remained in the QIP with their original specification.

The QIP was undertaken just before an effort by the Integrated Healthcare Association (IHA), a multiple stakeholder coalition, to launch coordinated medical group pay-for-performance across 7 health plans in California, including PacifiCare, using a consistent set of measures. The 7 health plans participating in the IHA effort constitute roughly

60% of the revenue stream to the physician organizations in the network.<sup>1</sup> Although the other 6 plans did not begin distributing financial awards related to the IHA targets until early 2004, the anticipation of additional rewards associated with performance on the same set of 10 clinical and service quality measures might strengthen the incentives to the physician groups to undertake quality improvement, which would be particularly true if there were fixed costs associated with quality improvement or spillover effects from plan enrollees to all patients treated by the physician groups.

### Data Acquisition

We obtained longitudinal data from PacifiCare on the performance of physician groups in its California and Pacific Northwest networks on the quality measures targeted by the QIP. Performance reports issued between October 2001 and April 2004, which covered patient treatment delivered between April 2001 and October 2003, were included in the study. The unit of observation is a physician group-quarter. Only physician groups with data for the entire period are included in the estimation of the effects of the QIP on clinical quality (numbers of groups vary by measure and are noted in the tables); in descriptive analyses of the awards, however, we include all bonus-eligible groups (numbers vary by quarter). Performance scores for clinical quality measures are computed by PacifiCare according to HEDIS and other specifications by using its administrative (encounter) data, which are routinely audited for accuracy. Numerators (individuals in a population group receiving evidence-based treatment) and denominators (individuals who should have received a particular service) were provided to us and are used in the modeling. PacifiCare also provided us with detailed, quarterly reports of bonus payments broken down by target and physician group, from which we calculated the financial impact of the program.

### Analytical Approach

All of our analyses related to changes in clinical quality relied heavily on the timing of the intervention to identify the impact of the financial incentives. Although the QIP officially began in January 2003, it is likely that there were anticipation effects because the details of the program had been known since early 2002. Moreover, because we measured performance by using administrative data, there was a lag between changes in practice and improvement in scores. If a physician group improved adherence to the clinical guidelines associated with a targeted measure in a given quarter, those results appeared in our data 6 months after the quarter ended. Thus, if the physician groups in the California network began implementing practice improvements in July 2002 when their contract for 2003 was signed, performance might be observed to improve as early as the April 2003 report. If, instead, policies and programs to improve quality were not put into place until January 1, 2003, we would expect to see those effects in the performance report produced in October 2003. We tested alternative assumptions about the timing of the response to the QIP in each of our models by defining the post-QIP period as beginning with the April, July, or October 2003 report. Because our findings were not qualitatively sensitive to the assumption about the timing of the response, we report only the results using performance reported in April 2003 as the beginning of the post-QIP period.

We first estimated whether the difference in performance scores for California physician groups after the QIP relative to before was greater than the same difference in the Northwest comparison practices. Covariates in this model included variables that indicate whether the observation is from a California group (the intervention group), whether it occurred in the post-QIP period (ie, between April 2003 and April 2004), and the interaction of these variables. For ease of interpretation, we report means of predicted values for the

intervention and comparison groups, before and after the QIP, along with bootstrapped SEs for the differences and for the difference-in-differences.

The difference-in-differences model was estimated with generalized estimating equations (GEEs) to account for the repeated-measures feature of the data in the context of a non-Gaussian outcome. We assumed that, consistent with the underlying nature of the performance data, which describe rates of adherence to guidelines, the error terms were binomially distributed with a logit transformation. The correlation structure was modeled as first-degree autoregressive, which allows for correlation in the error term of adjacent observations; results were qualitatively insensitive to less restrictive assumptions about the correlation structure.

To examine the financial impact of the QIP, we report the total potential dollars that could have been distributed in each quarter and the total, average, and maximum payouts. To give a better sense of the distribution of bonus payments, we also report the number of groups in each quarter that received any bonus and the number that reached at least half of the targets.

To examine differential improvement of practices above and below the common target, another set of models was estimated using only data for the California groups because of the small number of Northwest groups. For each of the 3 targeted measures, we compared the performance of each group at baseline by using performance data released in October 2002 to the QIP target and created 3 categories: groups at or above the target, groups below but within 10% of the target, and groups more than 10% below the target. These cutoffs divided the network into segments of roughly equal numbers of physician groups.

We also tested the sensitivity of our results to using 20% as the cutoff between the middle and lowest groups. We then estimated a model with the post-QIP dummy variable, dummy variables for the second and third of the 3 groups just described, and interactions

between the post-QIP variable and the 2 performance-based subgroup dummy variables. We report predicted values for each of the groups from these models before and after the QIP, as well as the difference (the percentage point improvement). For comparison with the estimates of post-QIP improvement of groups at different distances from the common target, we computed the total bonus dollars the groups received in the first year of the QIP in relation to each quality domain. Because bonus allocations are computed as a function of the number of PacifiCare members served, we also report membership for each category of physician groups.

All analyses were conducted using SAS version 9.1.2 (SAS Institute Inc, Cary, NC) and  $P < .05$  was set a priori as statistically significant.

## RESULTS

TABLE 1 reports the population average predicted values from the GEE models for cervical cancer screening, mammography, and HbA<sub>1c</sub> in California and the Pacific Northwest before and after the QIP. Although improvement occurred in California on all 3 measures after the QIP, improvements also occurred in the Pacific Northwest. Among the difference-in-differences, only the 3.6% difference for cervical cancer screening improvement between California and the Pacific Northwest was significant ( $P = .02$ ).

During the first year of the program, which included quarterly payouts between July 2003 and April 2004, PacifiCare offered approximately \$12.9 million in potential quality bonuses (TABLE 2). In total, the plan awarded \$3.4 million (27% of the amount set aside) in bonus payments. The mean quarterly bonus payment to each medical group during the first year increased from \$4986 in July 2003 to \$5437 in April 2004.

Of 163 eligible physician groups, 97 (60%) received a distribution of funds from the program related to at least 1 physician group quality performance target in the first quarter of the QIP. In the last payout based on the original set

of targets (April 2004), 129 of 172 (75%) groups reached at least 1 physician group quality target. It was uncommon for a physician group to reach more than half of the 10 quality targets; only 14 groups achieved this rate of success, even in the final quarter before the targets were raised.

In the stratified analyses examining performance improvements within the California network as a function of initial proximity to the target, a clear pattern emerged (TABLE 3). In this analysis, we designated group 1 to be physician groups with baseline performance at or above the target; group 2 includes those below but within 10% of the target; and group 3 includes physician groups that are more than 10% below the target. For all 3 quality domains, group 1 improved the least,

whereas group 3 improved the most. For example, group 1 improved mammography rates by only 0.7%, whereas group 3 improved 6.6% ( $P = .07$ ). Pairwise differences between group 1 and group 2 and group 1 and group 3 were statistically significant for cervical cancer screening ( $P = .03$ ;  $P = .02$ ). For HbA<sub>1c</sub> testing, the difference between groups 1 and 3 was also statistically significant ( $P = .001$ ). Results were qualitatively similar using 20% as the threshold distance from the target at baseline to divide groups 2 and 3 (data not shown).

The bonus awards paid out to physician groups in group 1 for cervical cancer screening, mammography, and HbA<sub>1c</sub> testing totaled \$436 618, \$383 370, and \$360 155, respectively. Payouts to group 2 were about one

**Table 1.** Improvement in Clinical Quality Scores for Quality Incentive Program (QIP) Measures\*

	Pre-QIP, %	Post-QIP, %†	Difference (Post – Pre), % (SE)	P Value
Cervical cancer screening				
California (n = 134)	39.2	44.5	5.3 (1.6)	<.001
Pacific Northwest (n = 33)	55.4	57.1	1.7 (0.9)	.03
Difference	-16.2	-12.6	3.6 (1.8)	.02
Mammography				
California (n = 134)	66.1	68.0	1.9 (1.1)	.04
Pacific Northwest (n = 32)	72.4	72.6	0.2 (1.1)	.43
Difference	-6.3	-4.6	1.7 (1.5)	.13
Hemoglobin A <sub>1c</sub> testing				
California (n = 134)	62.0	64.1	2.1 (1.0)	.02
Pacific Northwest (n = 31)	80.0	82.1	2.1 (3.3)	.20
Difference	-18.0	-18.0	0.0 (3.5)	.50

\*Data from authors' analysis of PacifiCare physician group performance reports for 2001-2004. Predicted values obtained from generalized estimating equation models of performance.

†For the purposes of this analysis, we defined the post-QIP period as beginning with the data reported for the first quarter of 2003.

**Table 2.** Financial Awards Offered and Made in the First Year of the Quality Incentive Program (QIP)\*

	July 2003	October 2003	January 2004	April 2004	Total QIP Year 1
Maximum possible network award, \$	3 251 092	3 334 328	3 379 169	3 253 095	12 850 505
Total amount paid, \$	812 772	833 582	844 792	935 079	3 426 226
Mean group payment	4986	5083	5183	5437	20 702
Maximum group payment	62 767	73 129	72 572	86 872	305 702
No. of eligible groups	163	164	163	172	172
No. of groups to reach any medical group quality target	97	110	118	129	129
No. of groups to reach $\geq 5$ medical group quality targets	13	15	14	14	15

\*Data from authors' analysis of PacifiCare QIP reports.

quarter to one third as much as those to group 1, whereas payouts to group 3 were far less. In all, across the 3 quality domains examined here, 75% of bonuses accrued to group 1 (calculations not shown), whereas only 5% accrued to group 3. By comparison, group 1 treated an average of just under 50% of the plan members served by groups in the analysis (ie, eligible organizations with complete data for the 10 quarters of the analysis), whereas both group 2 and group 3 averaged approximately 25% of members each across the 3 measures.

### COMMENT

Our analysis suggests that, although quality of care improved for all 3 targeted quality measures, only for cervical cancer screening was the improvement greater in California than in comparable Pacific Northwest physician groups not subject to such incentives.

In the first year of its QIP, the plan paid \$3.4 million of a potential bonus pool of \$12.9 million. Three quarters of the 172 physician groups eligible at some point during the year for the program received some funds from the bonus pool. We also observed that few groups reached a majority of targets, consistent with the low correla-

tion in performance across clinical areas that has been observed in other studies.<sup>12</sup>

Physician groups whose performance was initially lowest improved the most, whereas physician groups that had previously achieved the targeted level of performance improved the least. Unlike quality improvement, which followed an inverse relationship to baseline performance, bonus dollars were garnered in direct proportion to baseline performance. Physician groups whose performance was above the bonus threshold at baseline captured 75% of bonus payments on average across the 3 quality domains we examined, despite their limited improvement.

Our findings give rise to a number of speculations about the effects of pay-for-performance. First, groups with baseline performance already above the targeted threshold appeared to understand that they needed only to maintain the status quo to receive the bonus payments. More surprising, perhaps, is that low-performing groups improved as much as they did, given that their short-run chances of receiving a bonus were likely to be low. One possibility is that the groups viewed the QIP as a larger signal of a changing environment in which they would face increasing pressure to improve their care systems and de-

cidated to begin moving in that direction. Paying explicitly for quality improvement might alter the incentives for high-performing and low-performing groups, distribute bonus dollars more toward the latter group, and possibly increase the overall impact of pay-for-performance. It would also at least in part address fairness concerns that some low-performing groups face insurmountable barriers to achieving the benchmark levels of performance because of limited resources or a patient population of low socioeconomic status. Some payers, however, object to the notion of rewarding improvement rather than achievement because it effectively condones low levels of performance. Paying for improvement fails to reward and even penalizes providers that have already achieved high levels of health care quality at the time a pay-for-performance program is initiated. It is possible to reward both performance and improvement and thus fulfill multiple objectives.

One possible reason that the QIP failed to yield a greater response is that the financial rewards for quality were too low to motivate substantial departures from the underlying trend in quality improvement. Per enrollee, the maximum annual bonus was a relatively modest \$27, or about 5% of the

**Table 3.** Quality Improvement After the Quality Incentive Program (QIP) and Bonus Payments to California Groups With High, Middle, or Low Baseline Performance\*

Quality Domain†	Total PacifiCare Members	Pre-QIP Rate, %	Post-QIP Rate, %	Improvement (Post-Pre), % (SE)	P Value	Bonuses Paid in Year 1, \$‡
Cervical cancer screening						
Group 1 (n = 50)	597 091	53.6	56.0	2.5 (0.8)	.001	436 618
Group 2 (n = 32)	287 610	40.8	48.1	7.4 (2.4)	.001	127 632
Group 3 (n = 52)	305 041	23.0	34.1	11.1 (3.9)	.002	26 859
Mammography						
Group 1 (n = 43)	557 119	72.3	73.0	0.7 (0.9)	.22	383 370
Group 2 (n = 50)	384 852	64.9	67.2	2.3 (1.0)	.01	88 787
Group 3 (n = 40)	244 270	52.6	59.1	6.6 (4.1)	.05	987
Hemoglobin A <sub>1c</sub> testing						
Group 1 (n = 46)	547 687	75.4	77.1	1.8 (1.2)	.07	360 155
Group 2 (n = 26)	231 157	62.2	64.8	2.7 (2.3)	.12	101 619
Group 3 (n = 56)	395 450	39.4	49.2	9.8 (2.7)	<.001	53 218

\*Data from authors' analysis of PacifiCare physician group performance reports from 2001-2004. Rates and differences are means of predicted values obtained from generalized estimating equation models of performance. Degrees of freedom in the statistical model depend on both the measure denominator size and numbers of medical groups. Because we have repeated measures over time and differential enrollment across medical groups, the power of the statistical analysis depends also on the within-group correlation over time and the enrollment patterns.

†Group 1 includes practices that were at or above the target at the launch of the incentive program (based on the last profile in 2002); group 2 is within 10% below the target; group 3 is more than 10% below the target.

‡Bonus dollars are those paid in relation to the specific target (eg, cervical cancer screening).

professional capitation amount. Moreover, PacifiCare accounts for only about 15% of the average group's revenue.

Finally, because we examined effects within 5 quarters of the program's initiation, our findings may reflect that more substantial quality improvement takes time. To alter the underlying rate of improvement, physician groups may need to make investments in infrastructure and human resources, and these investments may be staged to take advantage of the cash flow from several quarters of bonus payments.

In many ways, PacifiCare is an ideal laboratory for studying pay-for-performance. For more than a decade, it has been profiling and feeding back comparative performance data on the quality of care delivered by physician groups in California. PacifiCare has also undertaken public reporting of relative rankings of physician groups on a subset of the performance measures since 1998 to leverage consumer choice and professional pride as motivation for quality improvement. Thus, we were able to examine the incremental impact of pay-for-performance after confidential profiling and public reporting had been in place for 5 years and to take analytic advantage of well-documented trends in performance. In many other settings, new pay-for-performance initiatives represent the first time that quality-of-care data are being systematically collected and, in some cases, publicly reported, making it difficult, if not impossible, to isolate the contribution of the payment incentives.

The uniqueness of PacifiCare's history and the California health maintenance organization market, which continues to be largely organized around capitated, multispecialty physician organizations, limits the generalizability of these results to other settings. Similar payment incentives offered in settings in which individual physicians are more likely to be compensated by fee for service might have a greater impact because correcting problems of underuse, which is what all of the quality measures we examined reflect, will also increase base compensation.

Our findings should be viewed in light of several inherent limitations of the study design. First, for identification of the effect of pay-for-performance, we relied on the assumption that absent the QIP, trends (or differences in trends) in quality improvement in California would have resembled those in the Pacific Northwest network. Although this assumption is generally supported by the similarity of pre-QIP trends between the 2 networks, it is not directly testable.

In addition, our estimates of the differential improvement after the QIP of high- and low-performing providers are influenced by regression to the mean and ceiling effects of unknown magnitude. These issues confound the causal interpretation of the differences among groups that differed in baseline performance. They do not, however, change that the majority of bonus funds were paid to groups that did not

demonstrate significant, measurable improvement.

PacifiCare's QIP, like most current pay-for-performance programs, should be viewed as a first step in the direction of aligning payment incentives with health system quality goals. Realization of the full potential of pay-for-performance to reduce the persistent gap between evidence-based and actual practice will require that payers adapt their incentive strategies as evidence to support best practices accumulates. The principal lesson we derive from this experience is that incentive design matters. The accumulating evidence from the continuing experimentation with pay-for-performance in the market will highlight these initial findings and other potential design lessons.

**Author Contributions:** Dr Rosenthal had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study concept and design:** Rosenthal, Frank, Epstein.

**Acquisition of data:** Rosenthal, Li, Epstein.

**Analysis and interpretation of data:** Rosenthal, Frank, Li, Epstein.

**Drafting of the manuscript:** Rosenthal, Frank, Epstein.

**Critical revision of the manuscript for important intellectual content:** Rosenthal, Frank, Li, Epstein.

**Statistical analysis:** Rosenthal, Frank, Li.

**Obtained funding:** Rosenthal, Frank, Epstein.

**Administrative, technical, or material support:** Epstein.

**Study supervision:** Rosenthal.

**Financial Disclosures:** None reported.

**Funding/Support:** Financial support for this research was provided by The Commonwealth Fund.

**Role of the Sponsor:** The Commonwealth Fund played no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

**Acknowledgment:** We are grateful to Sam Ho, MD, PacifiCare's corporate medical director, and numerous others at PacifiCare for providing access to the data and technical assistance.

## REFERENCES

- Rosenthal MB, Fernandopulle R, Song HR, Landon B. Paying for quality: providers' incentives for quality improvement. *Health Aff (Millwood)*. 2004;23:127-141.
- Baker G, Carter B. Introduction to case studies in health plan pay-for-performance. In: *The Evolution of Pay for Performance Models for Rewarding Providers*. Washington, DC: Atlantic Information Services; 2004.
- Epstein AM, Lee TH, Hammel MB. Paying physicians for high-quality care. *N Engl J Med*. 2004;350:406-410.
- Berwick DM, DeParle NA, Eddy DM, et al. Paying for performance: Medicare should lead. *Health Aff (Millwood)*. 2003;22:8-10.
- Dudley RA, Frolich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS. *Strategies to Support Quality-Based Purchasing: A Review of the Evidence*. Rockville, Md: Agency for Healthcare Research and Quality; 2004. Technical Review 10, AHRQ No. 04-PO24.
- Rosenthal MB, Frank RG. What is the empirical basis for paying for quality in health care? *Med Care Res Rev*. In press.
- Amundson G, Solberg LI, Reed M, et al. Paying for quality improvement: compliance with tobacco cessation guidelines. *Jt Comm J Qual Saf*. 2003;29:59-65.
- Fairbrother G, Hanson KL, Friedman S, Butts GC. The impact of physician bonus on enhanced fees and feedback on childhood immunization coverage rates. *Am J Public Health*. 1999;89:171-175.
- Kouides RW, Bennett NM, Lewis B, Cappuccio JD. Performance-based physician reimbursement and influenza immunization rates in the elderly. *Am J Prev Med*. 1998;14:89-95.
- Roski J, Jeddelloh R, An L, et al. The impact of financial incentives and a patient registry on preventive care quality: increasing provider adherence to evidence-based smoking cessation practice guidelines. *Prev Med*. 2003;36:291-299.
- Rosenthal MB, Frank RG, Buchanan J, Epstein AM. Scale and scope of capitated physician organizations in California. *Health Aff (Millwood)*. 2001;20:109-119.
- Schneider EC, Zaslavsky AM, Landon BE, et al. National quality monitoring of Medicare health plans: the relationship between enrollees' reports and the quality of clinical care. *Med Care*. 2001;39:1313-1325.