

© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2009.01072.x
RESEARCH BRIEF

Research and Methods Briefs

Does the Frequency of Pay-for-Performance Payment Matter?—Experience from a Randomized Trial

Sukyung Chung, Latha Palaniappan, Eric Wong, Haya Rubin, and Harold Luft

Objective. To examine the effects of incentive payment frequency on quality measures in a physician-specific pay-for-performance (P4P) experiment.

Study Setting. A multispecialty physician group practice.

Study Design. In 2007, all primary care physicians ($n = 179$) were randomized into two study arms differing by the frequency of incentive payment, either four quarterly bonus checks or a single year-end bonus (maximum of U.S.\$5,000/year for both arms).

Data Collection/Extraction Methods. Data were extracted from electronic health records. Quality measure scores between the two arms over four quarters were compared.

Principal Findings. There was no difference between the two arms in average quality measure score or in total bonus amount earned.

Conclusions. Physicians' responses to a P4P program with a small maximum bonus do not differ by frequency of bonus payment.

Key Words. Pay-for-performance, quality of care, primary care practice, physician incentives, randomized trial

A rapidly growing number of health plans and payers in the United States are adopting some type of pay-for-performance (P4P) mechanisms to reward quality of care in addition to the traditional payment structure that rewards volume of care provided. P4P programs are designed to offer financial incentives to a group of (and occasionally individual) physicians based on their performance on clinical and service quality measures. Implementation of P4P programs, however, cannot readily be informed by empirical evidence, as findings of existing studies are often inconsistent and details on program design and implementation are generally not well documented (Petersen et al.

2006). Organizational-level incentives from payers are generally not distributed to individual physicians or are shared equally (Hillman et al. 1998, 1999; Kouides et al. 1998; Amundson et al. 2003; Roski et al. 2003). Some recent studies examined financial rewards based on individual physician's performance (Beaulieu and Horrigan 2005; Doran et al. 2006; Levin-Scherz, DeVita, and Timbie 2006; Gilmore et al. 2007), suggesting that even a small amount of bonuses could improve performances on incentivized measures. In P4P programs in published studies, incentives were typically provided annually. One study assessed quarterly bonuses, which were given to medical groups rather than individual physicians (Rosenthal et al. 2005).

Reinforcement theory from psychology literature suggests that changing behavior by incentives is easiest when the linkage between behavior and incentive (or positive reinforcer) is clearest, and the reinforcers are placed in routine (Luthans and Kreitner 1985). Therefore, bonuses based on individual performance and provided promptly might be more effective than group-level and delayed bonuses. On the other hand, theories in organizational behavior suggest that the *amount* of payment contingent on performance, apart from any system effect, would be an important determinant of the effectiveness of incentives (Lawler 1968). Empirical evidence in nonhealth areas suggests that increasing the anticipated bonus *amount* may improve performance (Locke et al. 1980; Khan and Sherer 1990). To our knowledge, however, there is no empirical evidence on whether the *frequency* of payment by itself, given a fixed maximum achievable bonus with a periodic reporting, would have an impact on performance.

Our study examined this question by taking advantage of a rare opportunity of conducting a randomized experiment within a P4P program. The notion we test here is that a check attached to the periodic quality measure report will garner a physician's attention and impact a physician's performance more than a quality measure report without immediate financial feedback. As a physician-specific P4P program for primary care physicians (PCPs) was being implemented in a large group practice, the physicians were randomized into one of the two study arms differing only in the frequency of

Address correspondence to Sukyung Chung, Ph.D., Palo Alto Medical Foundation Research Institute, 795 El Camino Real, Palo Alto, CA 94301; e-mail: chungs@pamfri.org. Latha Palaniappan, M.D., M.S., and Eric Wong, M.S., are with the Palo Alto Medical Foundation Research Institute, Palo Alto, CA. Haya Rubin, M.D., Ph.D., is with the Research and Evaluation, Palo Alto, CA. Harold S. Luft, Ph.D., is with the Palo Alto Medical Foundation Research Institute, and Philip R. Lee Institute for Health Policy Studies, University of California, San Francisco, CA.

payment—quarterly versus year end. Physicians in both arms continued receiving quarterly reminders on quality reporting, and all were newly eligible for the performance-based individual incentive payments. The overall favorable effect of the physician-specific P4P program on the incentivized quality measures is reported elsewhere (Chung et al. 2009); here we focus on whether the frequency of payment alone makes difference in physicians' response to the incentive program.

Our study adds to the literature on physician P4P in two important ways. First, we provide the first empirical evidence on a specific method of implementing physician-specific performance-based payment in a practice setting. If increasing frequency of bonus payment, which incurs additional cost, does not make an improvement, a less frequent payment option would be more cost-effective. Second, in comparing the two methods, we use a randomized controlled design that allows for assessing the causal relationship between incentive frequency and performance improvement.

METHODS

Study Setting

The study was conducted at Palo Alto Medical Clinic (PAMC) of the Palo Alto Medical Foundation (PAMF). The PAMF is a not-for-profit health care organization in Northern California, contracting in 2007 with three multi-specialty physician groups. The PAMC currently provides care in five clinics operating in Fremont, Los Altos, Palo Alto, Redwood City, and Redwood Shores. Approximately 13 percent of the general population in the underlying geographic area is enrolled with PAMC. The data of this study came from the electronic health records (EHR) that have been in use at the study sites since 2000.

Design of the Physician Incentive Program

In 2007, PAMC implemented a *physician-level* P4P incentive program where the amount of the incentive was determined by individual physician performance. Physicians at PAMC are normally paid based on relative value-based units of service. All the PCPs at PAMC in family medicine, internal medicine, or pediatrics participated in the incentive program, and all the patients of the participating PCPs were considered for the performance evaluation, regardless of insurance plan.

The program at PAMC had the two arms differing only by the frequency of incentive payment. Participating physicians were randomized to one of these two arms: one arm received four quarterly bonuses and another received a single year-end bonus. Bonuses were calculated each quarter, paid to the quarterly group and accumulated for the year-end group. The incentive payments were included with the regular paychecks. The maximum achievable bonus was U.S.\$5,000/year or U.S.\$1,250/quarter, representing about 2 percent of the annual PCP salary.

The incentive program targets and incentives were developed via a consensus process by representatives of the participating physician departments. This included definitions of measures, eligible and qualifying patients for each measure, and formulae for incentive calculation. See Chung et al. (2009) for additional details on the program and quality measures.

PCPs at PAMC had been receiving reports, with quarterly updates, on performance on a variety of quality measures since 2003. With this reporting system, physicians were alerted by e-mail with an electronic link to a detailed quality score workbook describing their scores for each quality measure, peer physicians' scores (individually identified), and rank relative to other physicians in the department. With the implementation of the incentive program in 2007, the plan was for the quarterly score report to be sent on the 24th day of the month following the evaluation quarter and the paycheck to be delivered 2 weeks after the quarterly report. In the first quarter, however, there was a 2-month administrative delay in both the calculations and check. The quarterly group, therefore, did not receive the first quarter's bonus until July 2007 and received the second quarter's bonus 1 month after the first quarter's bonus. The other reports and bonuses were properly timed as planned.

Quality Measures

Nine of 15 measures implemented in the incentive program were selected from measures routinely monitored and reported to the physicians for several years. These included three outcome measures for diabetes control (blood pressure $\leq 130/80$ mmHg, HbA1C $< 7\%$, and LDL < 100 mg/dL) and six process measures (prescription of asthma controller, cervical cancer screening, Chlamydia screening, colon cancer screening, whether the height and weight were measured and recorded, and documentation of tobacco use history).

The other six measures, all specific to pediatric patients, were newly adopted for the 2007 program. As physicians began to use the new measures,

many raised questions about specific definitions of these metrics, and some metrics were modified during the year. We therefore excluded these pediatrics measures and the assessments of all pediatricians in the present analyses to ensure meaningful comparisons of scores over the four quarters.

Scoring and Incentive Scheme

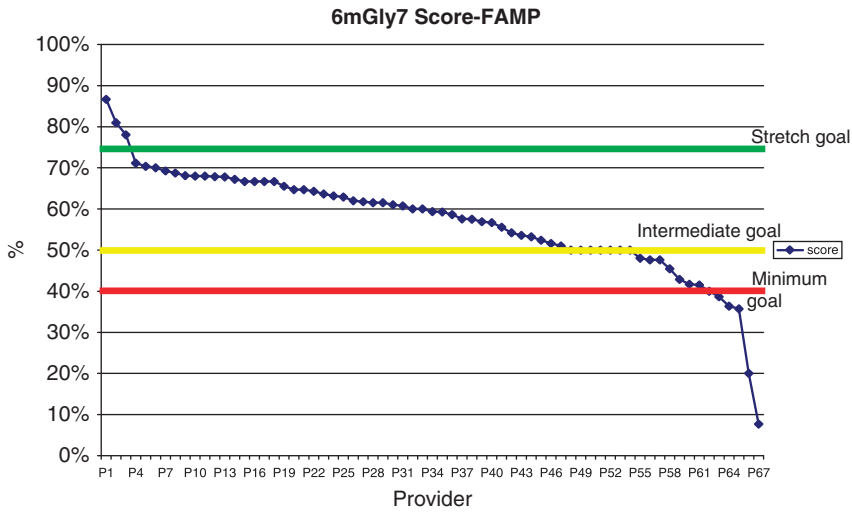
Physicians received either four quarterly bonuses or single year-end bonuses; both were based on individual composite points calculated quarterly by an algorithm developed by the incentive program leadership. For each quality measure, the percent score (i.e., numerator/denominator \times 100, where numerator was the number of patients receiving recommended care and denominator was the number of patients eligible for the recommended care) was calculated. Thresholds for performance ranging from minimally acceptable (1) to stretch (3) points were set based on previous scores (reflecting all the physicians in each department) for that measure. Only measures for which a physician had six or more patients in the quarter were considered qualifying. (The purpose of physicians setting the minimum criterion for the denominator in the incentive design was to prevent the percent score from being dominated by a few cases; we followed their approach.) The bonus for each quarter was calculated as (points earned for the qualifying measures/maximum achievable points) \times U.S.\$1,250, where the maximum achievable points = 3 \times number of qualifying measures.

Figure 1 is an example of percent scores, thresholds, and achieved points for the “diabetes HbA1C control” measure among family medicine providers, as it appears in the provider workbook screen. All the physicians received instructions on how to use and interpret the quality score workbook that allows them to look at their own patients who were included in their numerators and denominators.

Statistical Analysis

We used the percent score (0–100) of each qualifying measure as a main dependent variable. (This eliminates the “boundary effects” potentially arising from scores close to the threshold levels that affected bonuses.) We tested for a difference in the percent score in each quarter or in the trend of improvement over the four quarters between quarterly and year-end payment groups. We also examined the quarterly trend in bonus amount (either paid to the quarterly group or accumulated for the year-end group) between the two arms; this measure reflects the impact of thresholds on bonus payments. We included

Figure 1: Provider Workbook: An Example of the “Diabetes HbA1C Control” Measure for the Family Medicine at Quarter 1, 2007



site-fixed effects and interaction terms between site and intervention arm to examine whether the intervention effect differed across clinic sites.

In all analyses, the first quarter score was the referent point. The unit of analysis was physician for each quarter. For each physician, observations from all nine measures were assessed. Within-physician correlations were taken into account using a physician random effects model. Statistical significance was considered at the $p < .05$ level. All statistical analysis was performed using *STATA* 10.0 (College Station, TX).

RESULTS

Summary Statistics

Among the 179 physicians initially randomized, 167 physicians were included in the program for all four quarters; 12 did not participate in the program for the whole years for various reasons (some left the medical group, others were on medical leave or sabbatical leave, and still others were working part time and did not have enough qualifying patients). After excluding physicians in the pediatrics department ($n = 43$) who had very few eligible

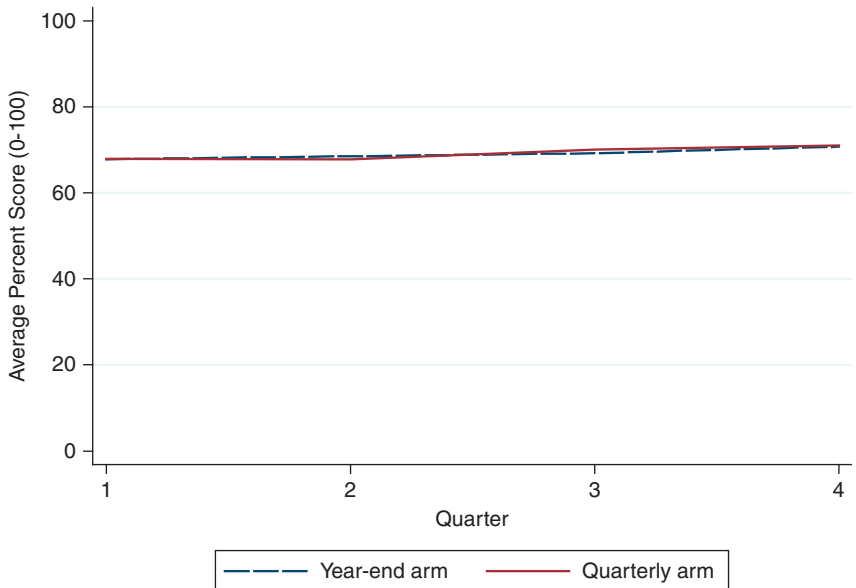
patients for the adult-focused measures, 124 physicians were included in the present study.

Among the 124 physicians, nearly all ($n = 120$) had one or more qualifying measures all four quarters. Slightly less than half (44 percent) were in the quarterly paid group. Reflecting the random assignment, there was no difference in the average prebaseline performance scores between the two arms.

The Effect of Frequency of Incentive Payment

The frequency of payment—quarterly or year-end—did not affect the average quality score over the four quarters (result table not presented). The average quality score in each quarter, by study arm, is plotted in Figure 2. In the plot,

Figure 2: Average Percent Score over the Four Quarters of 2007—Comparison between Quarterly and Year-End Arms

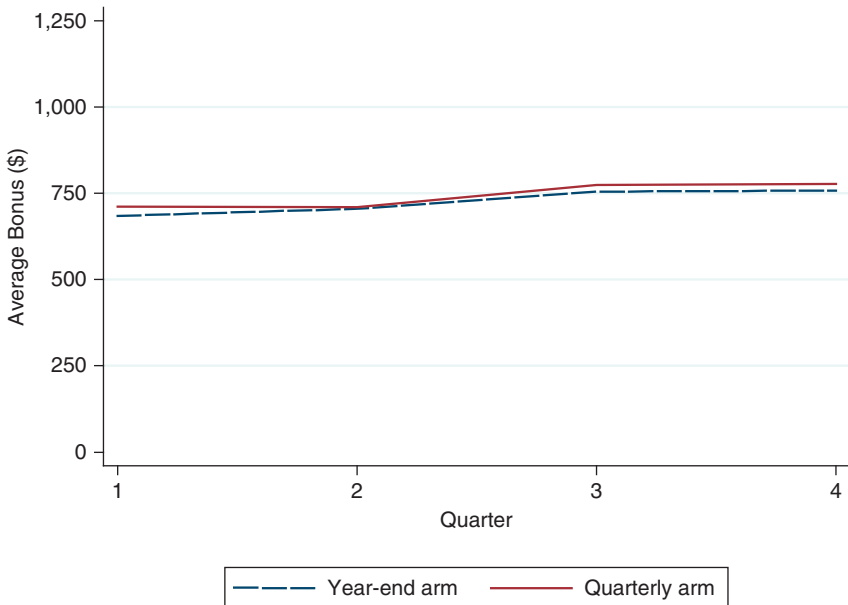


Note. No statistical difference in the average percent score was detected between the two arms, after controlling for indicators of quarter and measure. Similarly, there was no difference in trend in percent score over the four quarters between the two arms.

the y axis is the average of the scores of the nine measures, weighing each measure equally. While it appears there was a slight increasing trend in the third and fourth quarter scores as compared to the first and second quarter scores, which may be associated with the delay in the payment and reporting, the trend was not statistically significant.

Similarly, the average bonus amount did not differ by the frequency of payment, as seen in Figure 3. Total bonus amount received varied substantially across physicians, ranging from U.S.\$425 to U.S.\$4,484 (average U.S.\$2,868, standard deviation U.S.\$724). Trends over the four quarters in bonus amount (either received for quarterly arm or accumulated for year-end arm) between the two arms did not differ. There was no site-specific variation in the effect of payment frequency on either quality score or bonus amount.

Figure 3: Average Quarterly Bonus Amount over the Four Quarters of 2007—Comparison between Quarterly and Year-End Arms



Note. No statistical difference in the average bonus amount was detected between the two arms. Similarly, there was no difference in trend in percent score over the four quarters between the two arms.

DISCUSSION

Despite of the promise and growing popularity of P4P as a method to compensate physicians, our current understanding on specifics of how to design and implement the program in real practice settings is lacking. The present study addresses frequency of payment, which is a potentially important factor to consider in implementing P4P in a clinical setting. In our P4P experiment, where incentives were given to individual physicians for their performance on a variety of quality measures, however, we found no differential improvement in overall quality measure scores based on the frequency of payment.

The main limitation of our study is that the impact of quarterly *payment* cannot be isolated from the impact of quarterly *reporting*. An alternative study design to assess the impact of *reporting* concurrently would use a third arm that only received year-end reporting along with bonus payment, but it was considered unethical to withhold the quarterly reports as part of a trial. Although the quarterly report had been sent to the physicians for several years, practice directors reported to the investigators that physicians in both arms suddenly began raising questions regarding the quarterly report after the implementation of the physician-level incentive program. While the purpose of the quarterly report was to remind physicians about quality monitoring, it may have become more effective in conjunction with the bonus program. If true, that is, the “bonus” payment changed physician’s response to quality reporting, that should be interpreted as an incremental effect of bonus payment. The question the present study sought to answer is whether the incremental effect differed by the frequency of payment, given other factors being equal, including continuation of the existing quarterly reporting.

Findings of our study should be interpreted within the context of the setting: physicians in this large medical group have been exposed for several years to reporting of the measures that were the focus of the bonus program. With information technology tools already in place, physicians in both study arms could easily identify their own eligible patients for each measure, not only to check the validity of the measures, but for which patients improvements might be necessary. They could also compare their performance with that of other physicians. The effect of frequency of payment might have been different in another setting.

The maximum bonus offered in this P4P program was roughly 2.5 percent of the average physician’s annual pay, and the average bonus (U.S.\$2,868) was 1.4 percent of the average physician’s annual pay. The magnitude of bonuses used in other studies examining physician-specific P4P

incentives for quality improvement varies widely. Larger bonuses seem to be more effective in changing physician's practice: a U.K. study showed significant improvement in measured quality with exceptionally generous bonuses (average of 35 percent of physician income; Doran et al. 2006). Thus, one reason our study did not show any effect of frequency of payment may be because the achievable bonus amount was too small, regardless of the frequency of payment.

We did not formally analyze the costs of implementing quarterly versus annual bonuses. Because the P4P experiment was built on the existing performance evaluation and reporting system, incremental costs for the preparation of the bonus calculation were small. Quarterly reporting is clearly more expensive than annual, but the ongoing feedback is perceived to be valuable; the incremental cost of sending three additional checks per physician per year is small.

In conclusion, the frequency of payment itself, with no difference in the maximum bonus amount or in the frequency of reporting, may not substantially affect physicians' response to a P4P program. Future work should further investigate the effect of varying the frequency of reporting under the same financial incentive scheme, as well as and the effects of varying bonus amounts.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This manuscript is derived from work supported under a contract with The Agency for Healthcare Research and Quality (contract #HHSA290200600023I). We appreciate inputs from Laurel Trujillo, MD, and Tomas Moran, MS, who provided data and contextual information about the P4P program, and Sally Kraft, MD, and Dan Dohan, PhD, research collaborators at the earlier stage of the program implementation.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Amundson, G., L. I. Solberg, M. Reed, E. M. Martini, and R. Carlson. 2003. "Paying for Quality Improvement: Compliance with Tobacco Cessation Guidelines." *Joint Commission Journal on Quality and Safety* 29 (2): 59–65.

- Beaulieu, N. D., and D. R. Horrigan. 2005. "Putting Smart Money to Work for Quality Improvement." *Health Services Research* 40 (5, Part 1): 1318–34.
- Chung, S., L. Palaniappan, L. Trujillo, H. Rubin, and H. S. Luft. 2009. "Effect of Physician-Specific Pay-for-Performance Incentives in a Large Group Practice." *American Journal of Managed Care*, in press.
- Doran, T., C. Fullwood, H. Gravelle, D. Reeves, E. Kontopantelis, U. Hiroeh, and M. Roland. 2006. "Pay-for-Performance Programs in Family Practices in the United Kingdom." *New England Journal of Medicine* 355 (4): 375–84.
- Gilmore, A. S., Y. Zhao, N. Kang, K. L. Ryskina, A. P. Legorreta, D. A. Taira, and R. S. Chung. 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, Part 1): 2140–59; discussion 294–323.
- Hillman, A. L., K. Ripley, N. Goldfarb, I. Nuamah, J. Weiner, and E. Lusk. 1998. "Physician Financial Incentives and Feedback: Failure to Increase Cancer Screening in Medicaid Managed Care." *American Journal of Public Health* 88 (11): 1699–701.
- Hillman, A. L., K. Ripley, N. Goldfarb, J. Weiner, I. Nuamah, and E. Lusk. 1999. "The Use of Physician Financial Incentives and Feedback to Improve Pediatric Preventive Care in Medicaid Managed Care." *Pediatrics* 104 (4, part 1): 931–5.
- Khan, L. M., and P. D. Sherer. 1990. "Contingent Pay and Managerial Performance." *Industrial and Labor Relations Review* 43 (3): 107–20.
- Kouides, R. W., N. M. Bennett, B. Lewis, J. D. Cappuccio, W. H. Barker, and F. M. LaForce. 1998. "Performance-Based Physician Reimbursement and Influenza Immunization Rates in the Elderly. The Primary-Care Physicians of Monroe County." *American Journal of Preventive Medicine* 14 (2): 89–95.
- Lawler, E. 1968. "Equity Theory as a Predictor of Productivity and Work Quality." *Psychological Bulletin* 70 (6): 259–63.
- Levin-Scherz, J., N. DeVita, and J. Timbie. 2006. "Impact of Pay-for-Performance Contracts and Network Registry on Diabetes and Asthma HEDIS Measures in an Integrated Delivery Network." *Medical Care Research and Review* 63 (1 suppl): 14S–28S.
- Locke, E. A., D. B. Feren, V. M. McCaleb, K. N. Shaw, and A. T. Denny. 1980. "The Relative Effectiveness of Four Methods of Motivating Employee Performance." In *Changes in Working Life*, edited by K. D. Duncan, M. M. Gruneberg, and D. Wallis, pp. 383–8. New York: Wiley.
- Luthans, F., and R. Kreitner. 1985. *Organizational Behavior Modification and Beyond: An Operant and Social Learning Approach*. Glenview, IL: Scott, Foresman.
- Petersen, L. A., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does P4P Improve the Quality of Health Care?" *Annals of Internal Medicine* 145 (4): 265–72.
- Rosenthal, M. B., R. G. Frank, K. Li, and A. M. Epstein. 2005. "Early Experience with Pay-For-Performance: From Concept to Practice." *Journal of the American Medical Association* 294 (14): 1788–93.

Roski, J., R. Jeddloh, L. An, H. Lando, P. Hannan, C. Hall, and S. H. Zhu. 2003. "The Impact of Financial Incentives and a Patient Registry on Preventive Care Quality: Increasing Provider Adherence to Evidence-Based Smoking Cessation Practice Guidelines." *Preventive Medicine* 36 (3): 291–9.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.