Pioneering Pay-for-Quality: Lessons from the Rewarding Results Demonstrations

Gary J. Young, J.D., Ph.D., James F. Burgess, Jr., Ph.D., and Bert White, M.B.A., D.Min.

This article reports six overarching lessons learned from seven pioneering initiatives in the pay-for-quality (P4Q) movement. These lessons relate to the specific design and implementation challenges sponsors of P4Q programs can expect. The lessons are: (1) P4Q can prioritize providers' quality goals, (2) provider engagement is difficult, (3) P4Q escalates concerns for data accuracy and validity, (4) P4Q increases the need for population-based information technology and infrastructure, (5) tradeoffs exist between stimulating investment in quality infrastructure and diluting the power of incentives for individuals, and (6) significant challenges exist in documenting a positive return on investment.

INTRODUCTION

In the last several years, the U.S. health care industry has embraced the concept of P4Q as a key approach to improving quality of care. The basic concept entails giving providers, usually physicians or hospitals, financial incentives to achieve assigned quality goals. Over 100 programs are now in place through health plans and employer groups (Belden, 2006). Several States either have adopted or are considering adopting a P4Q strategy for their Medicaid Programs, and Congress has mandated Medicare to have a P4Q program for hospitals by 2009. The adoption of a P4Q strategy for Medicare, the largest health insurance program in the U.S., sends a clear signal throughout the industry that using financial incentives to promote quality has become a central element of U.S. health care policy.

As the P4Q strategy spreads, program sponsors need information that can guide them in designing and implementing P4Q programs. While some academic studies are now appearing in the literature (Lindenauer et al., 2007; Rosenthal et al., 2006), these generally tend to focus on documenting program effects on quality and do not speak to the challenges that program sponsors face in implementing such programs. In this article, we report on several overarching lessons that we gleaned during our role as the national evaluator for the Rewarding Results (RR) demonstration. RR, which consisted of seven demonstration sites, was designed to test different ways of designing and implementing programs for linking financial incentives with quality measures. While we have reported selected findings from our evaluation elsewhere, in this article we seek to synthesize our evaluation results into a set of key lessons that can provide guidance to program sponsors in their efforts to design and implement P4Q programs in the future. Our previous reports of selected findings from the evaluation are cited throughout the article as they apply to the key lessons.

The authors are with the Department of Veterans Affairs and Boston University School of Public Health. The research in this article was supported by the Agency for Healthcare Research and Quality (AHRQ) under Grant Number 5 R01 HS013591 and by the Robert Wood Johnson Foundation (RWJF) under Grant Number 04851. The statements expressed in this article are those of the authors and do not necessarily reflect the views or policies of the Department of Veterans Affairs, Boston University School of Public Health, AHRQ, RWJF, or the Centers for Medicare & Medicaid Services (CMS).

BACKGROUND

In 2002, three private foundations—RWJF, California HealthCare Foundation, and the Commonwealth Fund—solicited sponsored projects to provide early, hands-on expertise in the design, implementation, and evaluation of provider reactions to P4Q initiatives. The 7 selected demonstration projects were chosen from 151 applications submitted by health plans, employers, unions, State agencies, and collaborative efforts among employers, providers, and plans. Each demonstration made providers eligible for financial and non-financial rewards based on the achievement of specific quality goals linked to clinical quality. Then, AHRQ funded us to do the national evaluation of the RR program.

As can be seen in Table 1, the RR demonstration projects offered varied approaches to designing and implementing P4Q programs. The sites covered several geographic areas of the country with strong representation of California, where there has been considerable interest in the concept. Across the seven programs, several types of insurance arrangements also were represented, namely health maintenance organization, preferred provider organization, and Medicaid. The programs varied as to the unit of accountability—the type of provider or organizational arrangement eligible to receive the incentive money. In some programs, individual physicians were eligible, whereas in others it was physician organizations (POs) (i.e., group practices and independent practice associations). In either case, the focus was generally on primary care physicians. Hospitals did serve as the unit of accountability for one program, Blue Cross Blue Shield® (BCBS®) of Michigan. Most of the selected quality measures targeted clinical processes undertaken to prevent disease or manage chronic illnesses. Additionally, the programs had different incentive structures consisting generally of either bonuses, enhanced fee schedules, or withhold arrangements. Also, the programs linked incentives to one or more reward categories in addition to clinical quality (Table 2).

METHODS

We arrived at the overarching lessons presented in this article by synthesizing information from several sources. One was the data we collected in our role as the RR national evaluator. These data include responses from over 1,500 physicians to a survey questionnaire we fielded in several of the demonstration sites that focused on respondents' general attitudes toward P4Q and specific experiences with P4Q programs (Young et al., 2007b). Another was telephone interviews we conducted with over 50 senior leaders of POs to learn how their organizations were responding to quality-related financial incentives (Bokhour et al., 2006). Most of these interviews were conducted with senior leaders whose organizations were involved with the demonstration sites in Massachusetts and California. We also conducted site visits to selected POs and hospitals (Sautter et al., 2007). In addition, we interviewed representatives of the health plans and employer groups that sponsored the P4Q programs in the demonstration sites. Details of our data collection methods and analytic procedures are presented in the previously referenced articles.

Aside from the data we collected, we also drew from the work of other researchers who studied these seven demonstration sites. While we served as the RR national evaluator, each program sponsor had a local evaluator as well who conducted analyses of particular interest to the sponsor. We drew from these local evaluators'

Table 1
Rewarding Results Demonstrations

Demonstration Project	Location	Unit of Accountability	Selected Clinical Measures	Reward Structure
Blue Cross Blue Shield® of Michigan All Products	Michigan	Hospitals	JCAHO AMI CHF Pneumonia Surgical Infection Prevention	Absolute Threshold, up to 4% DRG Fee Enhancement
Blue Cross® of California Preferred Provider Organization	San Francisco, CA	Individual Physicians	HEDIS® CHF Diabetes Depression	Bonus up to $5,000
Bridges to Excellence Employer Coalition	Albany/Schenectady, NY Boston, MA Cincinnati, OH Louisville, KY	Individual Physicians and Physician Organizations	NCQA Performance Recognition for Diabetes and Cardiac Care	Bonus Per Patient Per Year up to $100 Diabetes; $160 Cardiac
Excellus/Rochester Individual Practice Association Health Maintenance Organization	Rochester, NY	Individual Physicians	HEDIS® (Adjusted with Community Standard) Diabetes Asthma	Withhold 50 to 150% Return Based on Relative Performance
Integrated Healthcare Association Health Maintenance Organization	California	Physician Organizations	HEDIS® Asthma Diabetes Women's Health	Plan-Specific Proprietary Reward Arrangements
Local Initiative Rewarding Results Medi-Cal Medicaid Providers	California	Individual Physicians and Physician Organizations	HEDIS® Well-Child Well-Adolescent Preventive Visits	Plan-Specific Reward/Bonus Per Service Tiering and Threshold
Massachusetts Health Quality Partners Health Maintenance Organization and Point of Service	Massachusetts	Physician Organizations	HEDIS® Asthma Diabetes Women's Health	Plan-Specific Proprietary Reward Arrangements

NOTES: JCAHO is Joint Commission on Accreditation of Healthcare Organizatons. AMI is acute myocardial infarction. CHF is chronic heart failure. DRG is diagnosis-related group. HEDIS® is Health Plan Employer Data and Information Set. HMO is health maintenance organization. NCQA is National Committee for Quality Assurance. Additional information is available on request from the authors and at Internet address: http://sph.bu.edu/p4p.

SOURCE: Young, G.J., Boston University School of Public Health, 2007.

Table 2

Reward Category and Corresponding Weights for Rewarding Results Demonstrations

Reward Category	Blue Cross Blue Shield® of Michigan	Blue Cross® of California	Bridges to Excellence[1]	Excellus/ Rochester Individual Practice Association	Integrated Health Association[2]	Local Initiative Rewarding Results	Massachusetts Health Quality Partners[2]
			Percent				
Clinical Quality Measurement	60	40	100	40	50	100	Plan-Specific Proprietary Reward Arrangements
Patient Satisfaction	—	—	—	20	30	—	Plan-Specific Proprietary Reward Arrangements
Information Technology	—	18	100	—	20	—	—
Patient Safety	30	—	—	—	—	—	—
Utilization Cost Effectiveness	—	—	—	40	—	—	—
Generic Prescribing	—	20	—	—	—	—	—
Access to Care	—	22	—	—	—	—	—
Community Project	10	—	—	—	—	—	—

[1] Offers distinct, focused incentive programs in the clinical quality measurement and information technology reward categories.

[2] These programs are consortiums among a group of health plans and other stakeholders, the Integrated Health Association consortium recommends weights for a category for specific initiatives, but Massachusetts Health Quality Partners does not.

SOURCE: Young, G.J., Boston University School of Public Health, 2007.

published and unpublished written material and interviewed them several times during the course of the demonstrations. We also attended several meetings that brought all the evaluators together to discuss and share results of their research.

LESSONS

These lessons (Table 3) do not speak specifically to the most critical question concerning P4Q; that is, does this approach lead to actual improvements in quality of care. We do not address this question in the form of a lesson because the demonstrations were not designed to provide definitive evidence on the impact of P4Q on clinical quality. Evaluations of the impact of the incentive programs have been conducted at several of the demonstrations sites (Felt-Lisk, Gimm, and Peterson, 2007; Young et al., 2007a), and these evaluations have produced evidence pointing to positive, albeit modest, improvements in quality. However, conclusions about the overall impact of financial incentives across the seven demonstration sites are significantly constrained by methodological limitations such as inadequate baseline data or a lack of suitable control groups. Accordingly, our focus in this article is to highlight the challenges and opportunities that program sponsors and providers confronted as pioneers in the P4Q movement.

Prioritize Providers' Quality Goals

The spread of P4Q in the health care industry has occurred in the face of a vigorous debate about whether and to what degree providers will respond to financial incentives for improving quality of care. While, as noted, we cannot offer definitive

Table 3

Six Key Overarching Lessons from Rewarding Results Demonstrations

P4Q Can Prioritize Providers' Quality Goals
Provider Engagement in P4Q is Difficult to Achieve
P4Q Escalates Concerns for the Accuracy and Validity of Quality Measurement
A Lack of Quality Improvement Infrastructure is a Major Barrier to Achieving P4Q Goals
Selecting a Unit of Accountability for a P4Q Program Entails Potentially Important Tradeoffs Between Stimulating Investment in Quality Infrastructure and Diluting the Power of Incentives for Individual Providers
Program Sponsors Face Significant Challenges in Documenting a Positive Return on Investment

NOTE: P4Q is pay-for-quality.

SOURCE: Gary J. Young, Ph.D., J.D., Boston University School of Public Health, 2007.

evidence of actual quality improvement (QI), the findings from the demonstrations strongly indicate that many providers gave priority to the quality targets for which they had incentives. Across the sites, many providers reportedly responded to the incentives by shifting their attention to the assigned quality targets. Their attention to the quality targets often led to investments in QI infrastructure, such as disease registries, point-of-service alert sheets in patient charts, and additional staff to conduct patient outreach activities.

For example, at the BCBS® of Michigan demonstration, which focused on hospitals, approximately 75 percent (50/66) of hospital leaders reported that they responded to incentives by enhancing their QI infrastructure. During site visits to 10 hospitals participating in this demonstration, we saw many examples of how hospitals responded to the financial incentives by adding staff or investing in information technology (Sautter et al., 2007). Similarly, at the Integrated Healthcare Association (IHA) demonstration, POs were observed to be very responsive to quality-related incentives, particularly through investments in information technology (Williams et al., 2006).

Indeed, across the sites, providers exhibited a willingness to pursue quality-related incentives even though the majority of physicians we surveyed had concerns about the adequacy of the incentive money (Young et al., 2007b). While the available evidence from the demonstrations cannot be used to estimate a precise relationship between magnitude of financial incentive and provider performance, some evidence from the demonstrations suggest that incentives amounting to approximately 5 percent of total physician earnings are adequate for gaining some degree of meaningful attention from physicians (Mehrotra et al., 2007; deBrantes, 2005). Of course, program sponsors who possess substantial market share, and thus account for a significant portion of providers' earnings, will likely have a distinct advantage in attracting the attention of providers.

Aside from money, provider responsiveness to P4Q programs appeared to reflect, at least in part, a feeling of professional responsibility for quality of care. Among providers interviewed at one demonstration site, there was a strong consensus that financial incentives tied to improving quality were far preferable to incentives tied to reducing utilization or increasing productivity (Bokhour et al., 2006). Further, the surveys we conducted indicated that most physicians are comfortable with the concept of P4Q, though they had definite concerns about the ways programs were designed and implemented. The surveys also indicated that in the early phases of the demonstrations, providers were not

overly concerned that the incentive-linked quality measures would distract them from performing other important, but not explicitly rewarded, activities (Young et al., 2007b).

Nevertheless, provider attitudes and behaviors regarding the incentive programs did not always prove optimistic for the future of P4Q. For instance, some providers found P4Q to be a one-size fits all proposition that can stand in the way of meaningful QI. In addition, while policymakers would of course like to see P4Q programs raise the quality of care for all patients, whether or not they are enrolled in a health plan with a financial incentive program, we found that some providers limited their attention only to those enrollees for whom incentives were available. Indeed, a number of POs reported that efforts to limit administrative costs associated with pursuing a quality-related incentive required that they focus only on covered patients.

Provider Engagement

While evidence from the demonstrations indicates shifting attention to incentive-linked quality targets, program sponsors also faced difficult challenges to engage many providers fully in P4Q programs. That is, providers turned their attention to the rewarded measures, but often with less than full appreciation of the critical features of the incentive programs. For program education, sponsors largely relied on standard methods of communication with providers—mass mailings, provider relations contacts, and newsletters; however, such methods clearly fell short. Our own surveys indicate that physicians, on average, had a low level of understanding of basic features of the P4Q programs in which they participated, such as incentive structures, payout formulas, and measurement specifications (Young et al., 2007a; Meterko et al., 2006). This was true of all physicians we surveyed, regardless of whether they or the organizations with which they were affiliated (i.e., physician or hospital) were the unit of accountability. Further, approximately 10 percent of the physicians surveyed indicated that they did not know any of the quality measures for which they were eligible to receive financial incentives. Other researchers reported similar findings from their independent efforts to assess provider awareness of and engagement with P4Q in the demonstration sites. One conclusion from the evaluation of the Bridges to Excellence demonstration was that physicians found the rules regarding patient eligibility and rewards complicated and difficult to understand (Thomson Medstat, 2007). In one site that rewarded individual physicians, 75 percent of those eligible received a bonus payment in the first year, yet very few knew if they had received it or made a connection between the payment and their performance on the program's quality measures (Teleki et al., 2006). Anecdotally, some physicians in this site were so unaware of the financial rewards available to them that they tossed out mail that included their bonus checks (Curtis, 2006).

The apparent difficulties program sponsors faced in achieving provider engagement in P4Q is not surprising given the many barriers that exist. Critical barriers to provider engagement in P4Q are the relative absence of organizational and physician leadership to create and foster engagement, and absence of systems support (i.e., data systems and personnel to generate feedback reports at the physician level) to facilitate monitoring and behavior change (Teleki et al., 2006). An additional barrier is that providers often do not read communications from health plans. To overcome some of these barriers, program

sponsors did try to engage providers by using more creative methods of outreach that, in one case, even entailed hiring a public relations firm. Some sponsors were able to enlist the support of physician leaders to speak about P4Q at professional events (Damberg et al., 2005). As a strategy to capture provider attention, one Medi-Cal health plan in San Francisco sent out two checks to physicians, one with an amount for what they actually earned in the incentive program and one voided check for what they could have earned had they reached all the quality targets for which they were eligible (Highsmith and Rothstein, 2006). Accordingly, program sponsors should use multiple methods of communication to convey initial program information.

However, even if program sponsors are successful in educating providers about the basic elements of a P4Q program, they still may lack their full commitment to the program. The evaluators for the Rochester Individual Practice Association (RIPA) demonstration theorized based on their own observations that commitment occurs largely as a developmental sequence across several stages that are similar to Kubler-Ross' stages of dying: denial that a program even exists; anger at being the subject of quality evaluation; desire to bargain the details of the program, and, finally, acceptance and commitment (Beckman et al., 2006). Because engagement is a likely critical success factor, it needs to be evaluated prior to performance measurement and regularly throughout the program. Midcourse corrections in communication and engagement should be anticipated.

Accuracy and Validity of Quality Measurement

All initiatives to measure provider performance raise concerns about the reliability and validity of the selected measures. These concerns were also evident across the demonstration sites. However, some initial concerns appeared to diminish as providers learned that the incentives would be linked to quality measures from well-established measurement sets such as the National Committee for Quality Assurance's Health Employer Data and Information Set (HEDIS®). Program sponsors relied on standardized measurement sets for two reasons: (1) the measures had been vetted through the clinical community for many years and had wide acceptance among providers, and (2) the measures were readily available, and in most cases did not impose additional significant data collection burden on themselves or providers. The measurement sets that sponsors adopted consisted primarily of measures of clinical process and not outcomes. Some providers did raise concerns that some of these process measures were more about documentation than actual QI.

In general, program sponsors collaborated with providers to select particular measures from these measurement sets for the programs. Such collaborations, which also helped to ease provider concerns about measurement issues, were particularly strong at Massachusetts Health Quality Partners (MHQP) and IHA, both of which are consortia of health plans, providers, and consumer groups. At the RIPA demonstration program sponsors modified the specifications of several HEDIS®-based diabetes measures in response to the recommendations of a RIPA community-based advisory committee. Further, RIPA developed an appeals process and reserve fund to address measurement concerns and rectify errors that resulted in inaccurate incentive payments (Curtin et al., 2006).

Nevertheless, providers still had significant concerns about data accuracy. In general, program sponsors wanted to rely on

electronic claims data to facilitate the collection and exchange of data. However, at one California demonstration, nearly all sampled physicians expressed suspicion about the ability of quality measures to reflect quality accurately unless the data were obtained from medical chart review (Teleki et al., 2006). A common concern was the correct assignment of patient eligibility for a given quality indicator. In recognition of such concerns, program sponsors expanded opportunities for technical assistance by instituting, for example, billing forums to train providers to avoid miscoding of claims (Highsmith and Rothstein, 2006).

Providers also expressed concerns about measurement based on small numbers of patients. Program sponsors took a number of steps to counter these concerns. Some sponsors established minimum numbers of patients for eligibility and others created composite measures. MHQP and IHA aggregated patient data across multiple payers. Another strategy assigned physicians the average score for their specialty for each measure where their patient count did not meet the required 30 patient minimum.

Data accuracy always will need to be a work in progress that is part of a continuous QI process. Central to this process is the provision of patient lists to providers that relate to quality measures for which they are accountable (Francis et al., 2006). Providers also should understand the measurement and reporting methods, and have a fair appeal process when there is dispute over data.

QI Infrastructure

Although many providers responded to quality-related incentives by investing in QI infrastructure, a significant number of providers also believed quite strongly that they lacked important resources for achieving the quality goals of the programs. In the surveys we conducted, physicians revealed some degree of uncertainty as to whether they had the necessary resources to achieve the quality measures for which they had incentives. At the same time, many providers apparently thought the financial incentives were not large enough to offset the costs of making needed investments in quality infrastructure (Young et al., 2007b).

For example, providers at the Medi-Cal site lamented that they lacked infrastructure to create monthly contact lists of enrollees who were due for well child visits (Highsmith and Rothstein, 2006). At another site in California, physicians commented that to be more successful at achieving assigned quality targets, they need better capabilities to monitor their own performance during the year (Teleki et al., 2006). While a majority of the physicians interviewed had received some feedback about their performance, the data were sometimes inaccurate. Such concerns appeared to be most acute at smaller POs that lacked internal resources to collect and improve clinical quality data (Williams et al., 2006). Some program sponsors did step in to assist providers by using their own staff or other resources to help track patients, monitor performance, and develop clinical protocols (Highsmith and Rothstein, 2006; Beckman et al., 2006; Damberg et al., 2005). Based on surveys and interviews with providers, however, such efforts were not sufficient to close what appeared to be a substantial shortfall in resources required for some providers to achieve quality targets.

Providers' concerns about lack of resources did not appear to be unfounded. Considerable evidence from the demonstrations indicated that resource availability influenced how well providers performed

on quality-related measures. At the BCBS® of Michigan demonstration, a hospital's ability to respond to the quality-related incentives appeared to depend on existing quality infrastructure, including staff availability, planning capacity, and information technology (Reiter et al., 2006; Sautter et al., 2007). At the RIPA site, initial incremental improvements in quality appeared to be unsustainable over time because in part the participating physicians, who were largely organized as solo practitioners and members of small groups, lacked the financial means to invest in infrastructure (Meterko et al., 2006; Young et al., 2007a). These physicians may have worked harder by spending more time with patients to convince them to have a particular preventive test performed, but this seemed to translate into one-time gains in quality measures rather than consistent increases in performance. These observations are consistent with the findings of researchers and QI organizations that have found that higher clinical performance and greater use of information technology go hand-in-hand (Casalino et al., 2003).

Clearly, without infrastructure, it is difficult for providers to make substantial and sustainable improvements in quality. This problem may be particularly acute for providers whose baseline performance is at the low end of the distribution for a quality measure. One recommendation is that incentives for such providers could be based in part on improvement instead of attainment of absolute thresholds.

Accountability

A lack of quality infrastructure was ostensibly a barrier for many providers to achieve quality goals of P4Q program. Sponsors of future P4Q programs will likely need to consider how to spur greater investment in quality infrastructure among participating providers. In this vein, findings from the demonstration sites point to possible tradeoffs that program sponsors may need to strike between selecting organizations versus physicians as the unit of accountability.

In terms of spurring quality-related investment, program sponsors may do best by selecting hospitals or POs as the unit of accountability. As one would expect, provider investment in QI infrastructure appeared considerably greater in those demonstration sites where either hospitals or POs served as the unit of accountability relative to sites that focused on individual physicians. Hospitals and POs are simply more likely to have the financial means to invest in new technology and personnel. In the previously noted survey results on physician resource availability, the lowest scores, on average, were from physicians who were participating in sites where individual physicians were the unit of accountability (Young et al., 2007b; Meterko et al., 2006). Provider interviews and site visits also revealed clear differences in investment activity between sites that rewarded organizations and sites that rewarded individuals. Moreover, for physicians in solo or small practices, money is not the only barrier to enhancing their use of quality infrastructure. Because many of these physicians have had little exposure to electronic medical records and related technologies, they typically do not possess a strong conceptual understanding of how to use advanced medical decision support. In one of the California sites, individual physicians seemed to lack an orientation to data-driven, patient population QIs (Teleki et al., 2006).

However, incentives directed to POs rather than individual providers may also tend to limit direct physician involvement. Our interviews and site visits revealed that when hospitals and POs are the unit

of accountability, these organizations may do little to engage individual physicians and other providers in the programs (Sautter et al., 2007; Bokhour et al., 2006). Thus, at the BCBS® of Michigan site, most of the hospitals that we site visited had not had any detailed discussions with medical staff members about the hospital incentive program even though the hospitals did undertake initiatives that directly or indirectly sought to modify physician behaviors in ways that were consistent with the quality measures. While a few hospitals did engage in some limited gainsharing arrangements with certain clinical departments, most commonly when the physician staff was highly connected to the hospital, none took the step of distributing incentive money to medical staff members based on their own performance. In sites that targeted incentives to POs, we found much more variation in terms of whether and how organizations engaged individual physicians in the program. Still, it was apparent that many of these organizations did not directly engage their physician members through shared incentive arrangements.

The implications for the future of P4Q in pursuing quality-related incentives without engaging individual physicians in the programs are extremely complex. Some organizations may be able to build the necessary infrastructure to achieve quality measures without directly interfacing with physicians about the incentives and sharing incentive money. Some of the POs that took this approach reportedly retained all incentive money at the organization-level to invest in that infrastructure. Still, such an approach obviously diminishes the power that incentives have for those who work on the frontlines of the health care industry, namely physicians. Thus, while program sponsors may be able to spur investment on the part of providers by selecting hospital or physicians as the unit of accountability, they need to be mindful of the diminished incentives for physicians and the long-term effects this may have on their morale and willingness to step up their quality of care.

Program Sponsors Face Challenges

To compute an accurate return on investment for a P4Q program, it is necessary to track all financial costs and returns, including opportunity costs resulting from activities not undertaken given the development of the program. While the BCBS® of Michigan, Bridges to Excellence, and RIPA demonstrations have reported a positive return on investment (Nahra et al., 2006; deBrantes, 2005; Curtin et al., 2007), program sponsors face significant challenges in computing valid return on investments (Wheeler et al., 2007). In particular, evaluation periods and data collection cycles often do not line up well and at this time program sponsors lack sufficient experience with P4Q to simulate effects. Like most business investments, much of the cost of P4Q is frontloaded while benefits may not accrue for many years. However, many payers are not guaranteed these benefits because of enrollee turnover from year-to-year. As such, the quantification of program-related benefits is quite difficult.

These and other similar complex issues arising in comprehensive return on investment calculations suggest that more narrowly focused questions about the effectiveness of P4Q programs are more likely to be answered accurately. QIs can be categorized as increasing or decreasing costs of care and on whether the benefits accrue immediately or only far out in the future. Positive return on investment reports are unlikely from quality processes

that do not reduce volume of services or cost per unit, or have benefits that accrue only far in the future.

CONCLUSION

As the P4Q movement in the U.S. continues to gather momentum, program sponsors are searching for evidence about the impacts of such programs. Our evaluation was unable to address in a comprehensive manner the impact of P4Q on QI. However, based on our evaluation of the RR demonstration, we expect that providers will take financial incentives for quality seriously and this may lead to improvements in quality of care. It is also clear that program sponsors face a significant learning curve for P4Q and so they, as well as policymakers and providers, will need to be vigilant in searching for best practices for designing and implementing P4Q programs.

REFERENCES

Belden, G.: *Leapfrog's Incentive and Reward Compendium Guide.* 2006. Internet address: http://www.leapfroggroup.org/leapfrog_compendium (Accessed 2007.)

Beckman, H., Suchman, A., Curtin, K., et al.: Physician Reactions to Quantitative Individual Performance Reports. *American Journal of Medical Quality* 21(3):192-199, May-June 2006.

Bokhour, B., Burgess, J., Jr., Hook, J., et al.: Incentive Implementation in Physician Practices: A Qualitative Study of Practice Executive Perspectives on Pay for Performance. *Medical Care Research and Review* 63(Supplement 1):73S-95S, February 2006.

Casalino, L., Gillies, R., Shortell, S., et al.: External Incentives, Information Technology, and Organized Process to Improve Health Care Quality for Patients with Chronic Diseases. *JAMA* 289(4):434-441, January 22, 2003.

Curtis, J.: *PULSE: Blue Cross of California's PPO Physician Pay for Performance Program.* Report to Rewarding Results: Lessons Learned & Implications for Medicare Pay for Performance CMS. Baltimore, MD. December 15, 2006.

Curtin, K., Beckman, H., Pankow, G., et al.: Return on Investment in Pay for Performance: A Diabetes Case Study. *Journal of Healthcare Management* 51(6):365-376, November-December, 2006.

Damberg, C., Raube, K., Williams, T., et al.: Paying for Performance: Implementing a Statewide Project in California. *Quality Management in Health Care* 14(2):66-79, April-June 2005.

deBrantes, F.: *Lessons Learned.* Institute of Medicine. May 4, 2005. Internet address: http://www.iom.edu/CMS/3809/25241/25255/26860.aspx (Accessed 2007.)

Felt-Lisk, S., Gimm, G., Peterson, S.: Making Pay-for-Performance Work in Medicaid. Web Exclusive. *Health Affairs* 26(4):W516-W517, June 26, 2007. Internet address: http://content.healthaffairs.org/cgi/content/abstract/hlthaff.26.4.w516 (Accessed 2007.)

Francis, D., Beckman, H., Chamberlain, J., et al.: Introducing a Multifaceted Intervention to Improve the Management of Otitis Media: How do Pediatrician, Internist, and Family Physicians Respond? *American Journal of Medical Quality* 21(2):134-143, March-April 2006.

Highsmith, N. and Rothstein, J.: *Rewarding Performance in Medicaid Managed Care.* Center for Health Care Strategies. March 2006. Internet address: http://www.allhealth.org/BriefingMaterials/RewardingPerformanceinMedicaidManagedCare-CHCS-497.pdf (Accessed 2007.)

Lindenauer, P., Remus, D., Roman, S., et al.: Public Reporting and Pay for Performance in Hospital Quality Improvement. *New England Journal of Medicine* 356(5):486-496, February 2007.

Mehrotra, A., Pearson, S., Coltin, K., et al.: The Response of Physician Groups to P4P Incentives. *American Journal of Managed Care* 13(5):249-255, May 2007.

Meterko, M., Young, G., White, B., et al.: Provider Attitudes Toward Pay-for-Performance Programs: Development and Validation of a Measurement Instrument. *Health Services Research* 41(5):1959-1978, October 2006.

Nahra, T., Reiter, K., Hirth, R., et al.: Cost-Effectiveness of Hospital Pay-for-Performance Incentives. *Medical Care Research and Review* 63(Supplement 1):49S-72S, February 2006.

Reiter, K., Nahra, T., Alexander, J., et al.: Hospital Responses to Pay-for-Performance Incentives. *Health Services Management Research* 19(2):123-134, May 2006.

Rosenthal, M., Landon, B., Normand, S., et al.: Pay for Performance in Commercial HMOs. *New England Journal of Medicine* 355(18):1895-1902, November 2, 2006.

Sautter, K., Bokhour, B., White, B., et al.: The Early Experience of a Hospital Based Pay-for-Performance Program. *Journal of Healthcare Management* 52(2):95-107, March/April 2007.

Teleki, S., Damberg, C., Pham, C., et al.: Will Financial Incentives Stimulate Quality Improvement? Reactions from Frontline Physicians. *American Journal of Medical Quality* 21(6):367-374, November-December 2006.

Thomson Medstat: *BTE Program Evaluation.* Internet address: http://www.bridgestoexcellence.org/assets/Documents/Program_Evaluation_Documents/BTE-Program-Evaluation-7-26-06.pdf (Accessed 2007.)

Wheeler, J., White, B., Rauscher, S.: Pay-for-Performance as a Method to Establish the Business Case for Quality. *Journal of Health Care Finance* 33(4):17-30, Summer 2007.

Williams, T., Raube, K., Damberg, C., et al.: Pay for Performance: Its Influence on the Use of IT in Physician Organizations. *Journal of Medical Practice Management* 21(5):301-306, March-April 2006.

Young, G., Meterko, M., Beckman, H., et al.: Effects of Paying Physicians Based on their Relative Performance for Quality. *Journal of General Internal Medicine* 22(6):872-876, June 2007a.

Young , G., Meterko, M., White, B., et al.: Physician Attitudes toward Pay-for-Quality Programs: Perspectives from the Front Line. *Medical Care Research and Review* 64(3):331-343, June 2007b.

Reprint Requests: Gary J. Young, Department of Health Policy and Management, Boston University School of Public Health, 715 Albany Street, T3W, Boston, MA 02118-2526. E-mail: health@bu.edu