

**Practical Issues in the Design and Implementation of Pay-for-Quality
Programs***

Gary J. Young, PhD, Department of Veterans Affairs and Boston University

and

Douglas A. Conrad, PhD, Department of Health Services, University of

Washington**

Acknowledgements: This paper was supported by a grant from the Center for Health Management Research to Drs. Young and Conrad, and Dr. Young's work was also supported by grants from the Agency for Healthcare Research and Quality.

Corresponding Author: Douglas A. Conrad, PhD, Professor, Department of Health Services and Center for Health Management Research, University of Washington, Room H660C, Box 357660, 1959 NE 59th Street, Seattle, WA 98195-7660. Phone: (206) 616-2923. Fax: (206) 543-3964. E-mail: dconrad@u.washington.edu.

Practical Issues in the Design and Implementation of Pay-for-Quality Programs

Executive Summary:

Health plans, health care purchasers, and provider organizations throughout the United States are crafting pay-for-performance programs with the intent of improving the quality of care, while also recognizing the need to restrain rapidly rising costs. Health plans and large, self-insured employers have typically led the movement toward quality scorecards for hospital and physician performance, coupled with the use of financial incentives directed at hospitals, physician group practices, and individual physicians and practice teams.

In this article we provide a conceptual perspective for understanding the objectives and constraints of payers and providers as they wrestle with the next generation of paying-for-quality (P4Q) programs. We next identify a set of practical issues that must be addressed in developing and conducting P4Q programs in different market environments. Those issues include specific strategies for choosing quality metrics, the unit(s) of accountability, size of the incentive, data and measurement systems, payout formula, and collaboration among payers.

We illuminate these issues by considering different approaches in the light of real-world P4Q demonstrations underway in the Rewarding Results Program, Bridges to Excellence, and in specific provider organizations that we have interviewed. The discussion of practical issues highlights principles and examples directly relevant for the managers of hospitals and physician organizations considering participation in P4Q, as well as those re-examining their internal compensation mechanisms for physicians.

Introduction

Across the United States, health plans and employer groups are adopting pay-for-performance programs that link financial incentives to quality of care. The basic concept is to present providers with financial incentives for achieving predefined quality targets. There are currently more than 150 such programs in operation and many more are under development (Baker and Carter, 2005). Most programs apply to primary care providers and feature quality targets consisting of preventive-oriented, process-based measures for chronic care conditions. However, a small number of existing programs focus on hospitals and many payers are now expressing strong interest in offering quality-related incentives to physician specialists.

Pay-for-performance itself is not a new concept to the healthcare industry. Managed care organizations and other payers have been offering providers financial incentives to achieve productivity and efficiency targets for well over twenty years (Conrad and Christianson, 2004). Gain-sharing arrangements between hospitals and physicians, a type of pay-for-performance program that typically focuses on cost savings goals, have also been around for some time. What is new is that the recent wave of pay-for-performance programs focuses, in some cases exclusively, on quality of care. This shift in emphasis may well usher in a new era in the reimbursement of US health care providers.

Several factors underlie this interest in the “pay-for-quality” (P4Q) concept. One such factor is a growing concern about the quality of US health care. Numerous studies indicate major deficits in provider compliance with evidence-based clinical practices (McGlynn et al., 2003)¹. There is also evidence that medical errors are occurring far more frequently than many health care experts expected. Another factor is health care expenditures, which have been rising

steadily for several years. Payers are hoping that improvements in clinical quality will mean healthier patients and healthier patients, ultimately, will translate into long-term cost savings – the so-called business case for quality. Further, a number of leading policy and scientific organizations have identified the lack of quality-related financial incentives in traditional provider reimbursement systems as a major barrier to quality improvement in the US (Institute of Medicine, 2001). Indeed, several proposals have been put forth to integrate a quality-related incentive program within the Medicare reimbursement system (e.g., the Medicare Value-Based Purchasing for Physicians' Services Act of 2005, S. 1356 109th Cong., 1st Sess. 2005).

While the motivation for P4Q programs is fairly clear, the best way to design and implement these programs is not. The pay-for-quality programs that are already in existence are by no means uniform. In fact, they vary markedly along such dimensions as their quality measures, financial incentives, unit of accountability, data and measurement systems, and payout formulas (Baker and Carter, 2005; Young et al. 2005). In this paper, we identify and discuss several practical issues in the design and implementation of P4Q programs. Our discussion is informed by relevant theory and research. However, because there is little scientific evidence to draw from on this topic, we have drawn extensively from our own field work with a number of prominent P4Q programs in the US. Over the last several years, we have conducted numerous interviews with individuals engaged in the design and implementation of P4Q programs about the decisions and challenges they are facing (Young et al 2005; Conrad et al 2005; Conrad and Christianson 2004). We have also had the opportunity to speak with senior leaders of provider organizations that are participating in P4Q programs about their perceptions of these programs and their efforts to adapt to quality-related incentives.

Conceptual Perspective

Agency theory offers a useful perspective for discussing issues in the design and implementation of P4Q programs. From this viewpoint, P4Q addresses what is known as a principal-agent problem. In this context a principal is someone who delegates work to a second party, the agent, for which the agent is compensated in some form. For the principal, a key issue or problem is how to structure the relationship in such a manner that the agent performs the work as desired. Two general approaches are contemplated. One is for the principal to monitor the agent's work activities. This approach assumes the principal is able to specify in advance what the appropriate work activities should be and identify deviations from these specifications. The other is for the principal to tie some or all of the agent's compensation to the achievement of the desired performance targets, an approach that assumes the ability to measure the agent's performance reliably. For any given principal, which approach is best depends on the relative effectiveness and cost of monitoring the agent's activities versus measuring the agent's performance.

In this vein, P4Q represents an effort on the part of payers (principal) to tie the compensation of providers (agent) to their quality-related performance. This approach is an alternative to monitoring the work activities of providers, which is difficult because providers typically possess superior knowledge about the effectiveness of particular diagnostic and treatment regimens. Agency theory would predict that high-powered financial incentives that directly reward providers' quality performance (or other principal) will result in improved quality performance. The performance measures must be sufficiently broad and robust to discourage providers from "treating to the test;" that is, focusing on explicitly rewarded dimensions of care to the detriment of other important, but excluded dimensions of performance

(Eggleston, 2005). This agency principle is termed **incentive compatibility**. Agency theory also implies that financial incentives must provide sufficient incremental compensation to cover the opportunity costs of the extra effort and resources required to produce improved performance. Theorists term this condition the **participation constraint** (Macho-Stadler and Perez-Castrillo 1997).²

Practical Issues in “P4Q” Design and Implementation

In the following section, we discuss a number of design and implementation issues that we have organized into six categories: quality measures, unit of accountability, financial incentives, data and measurement systems, payout formulas, and payer collaborations.

Quality Measures. Clearly, a defining feature of a P4Q program is its quality measures. In our experience, we find that payers struggle with several key decisions in their efforts to select these measures. One such decision concerns the types of clinical conditions on which to focus, specifically chronic and acute care. Some types of provider, namely physicians, deliver both kinds of care and an optimal incentive program would reward performance in both areas. The fact that most programs currently focus on preventive care and primary care of chronic conditions reflects both the high health payoffs of preventive care and the high concentration of expenditures among persons with chronic conditions, who account for more than 75% of direct medical care costs (Centers for Disease Control and Prevention 2004).³ Indeed, Agency theory implies that an optimal quality incentive program would be based on a broad range of measures, which capture a representative set of the preventive, diagnostic, and treatment processes typically

managed by the physician. A narrow set of measures is more likely to trigger "multi-tasking", in which providers pay less attention to unrewarded aspects of care in favor of the measures explicitly included in the quality incentive (Eggleston 2005). A broader set of metrics also helps contain the effects of random measurement error in individual quality measures.

For any given quality measure, whether it is for an acute or chronic condition, another key decision relates to whether to focus on clinical processes or outcomes. While outcome measures have an obvious "bottom-line" appeal, our own field work suggests that process measures may often be far more feasible for most P4Q programs. Process measures are the most controllable aspects of quality from the standpoint of providers. This is because a patient's severity of illness and compliance with treatment protocols as well as random events significantly influence patient health outcomes. Thus, other things being equal, it may be prudent to weigh process most heavily (the behavior being directly "incented") and use outcome measures primarily as a "check" on the predictive validity of the process indicators. Otherwise, one risks weakening the incentive effects by relying on outcome indicators that produce both "false positive" and "false negative" signals of the actual quality of the physician's diagnostic and treatment patterns.

A third key decision is whether to select measures based on national standards or those derived from local priorities and customs. Selecting measures based on peer-reviewed national standards of care both avoids re-inventing the wheel and enhances the credibility of the measures. At present, most P4Q programs appear to draw predominantly on such standards -- for example, the HEDIS metrics of NCQA for medical groups, the JCAHO Core Performance measures for hospitals. Still, at times local adjustments to national standards might sometimes

be advantageous to secure the commitment of providers to a program. Such adjustments can take into account clinical capacity, market conditions, and community norms.

Unit(s) of Accountability. In designing P4Q programs, payers must decide on the type (s) of provider to whom they want to direct financial incentives. A key choice is whether to direct incentives to individual providers, namely physicians, or to provider organizations such as medical groups and hospitals.

By and large, the production of quality is an exercise in team production. This reality is acknowledged in the General Medical Services Contract in the United Kingdom, which is currently engaged in a national P4Q program. In the UK, quality payments are made directly to the practice, not the individual provider (Smith and York 2004). Many programs in the US also direct incentives to medical groups and other physician organizations. The value of incentives directly applied to the level of the group practice derives from their impact on the organization's norms and "ambient risk" (Landon et al. 1998). The organization and group-level incentives play an important role in reinforcing the individual physician-level reward structure. Medical groups are designing individual physician compensation methods that complement organization-level incentives. These individual-level incentives currently center on primary care physicians for the most part, but some organizations have begun to craft incentives for non-primary care specialists (Conrad and Saver 2005).

However, other considerations may also need to be weighed into the decision of who to select as the unit of accountability. A recent study provides evidence that when physician organizations (i.e., medical groups and independent practice associations) serve as the financial intermediary in P4Q programs they may not distribute incentive money to individual physicians

or may distribute the money in ways that potentially attenuate the power of the incentives (Bokhour et al. 2006). Certainly, more research is needed to identify the factors that influence whether and how provider organizations transmit incentives to individual providers and how this affects the power of the incentives to influence clinical behavior.

Financial Incentives. The size of the incentive payment is obviously a central consideration in the design of pay-for-quality programs. Provided the expected payment level exceeds the provider's administrative and opportunity costs of responding to the incentives (i.e., the "participation constraint" of agency theory), some quality improvement is likely. Indeed, most programs appear to offer incentives that are relatively modest in size (e.g., \$2.25 per Health Net member per month in the IHA program or the maximum 4% increase in DRG payments for superior quality in the BCBS of Michigan program). If the "low-hanging fruit" in quality improvement is relatively abundant, as suggested by recent studies concerning the appropriateness and quality of medical care (McGlynn et al., 2003), modest financial inducements may be sufficient to stimulate significant early gains in clinical quality.

Thus modest size incentives may be very appropriate during the initial phases of a P4Q program. Moreover, small financial incentives allow payers to maintain "dynamic budget neutrality", in the sense that premium increases over time are approximately equal to the rise in the payer's costs. Such small incentives also minimize the likelihood that extrinsic financial rewards will "crowd out" providers' intrinsic motivation (Frey 1997), and also reduce the gains from provider multi-tasking or "treating to the test" (Eggleston 2005; Prendergast 1999).

Data and Measurement Systems. Not surprisingly, a major concern among providers participating in P4Q programs is the validity of data used for quality measurement (Bokhour et al. 2006). As demonstrated by the pay-for-performance program of the Integrated Healthcare Association (IHA) of California (Damberg et al. 2005), there are practical differences in the information capacities of the payers and providers that potentially can raise data issues for P4Q programs. For example, in the first year of the IHA-sponsored P4P program there were several systematic differences between the reported measures for plans and medical groups⁴:

- For preventive care metrics, physician organization-reported rates of screening and immunization were consistently higher than comparable plan-reported rates.
- Similarly, for diabetes and cardiac care, the screening rates for HbA1c and LDL screening were higher in the physician organization-reported results.
- In contrast, the compliance rates for appropriate use of asthma medications were higher for all age segments in physician organization-reported measures.

Gaps or reporting lags in pharmacy and other utilization data emerge when patients purchase “out-of-plan” services (e.g., over the counter drugs), or when health plans contract with pharmacy benefit managers (PBMs) and the information is not readily accessible from the PBM’s data warehouse. As illustrated in a IHA pilot study of quality measures, physician organizations are more likely to have complete lab data, while the health plans generally will have better access to pharmacy data – either through their own claims payment systems or their PBMs.

On balance, these patterns suggest that an incentive measurement system should blend claims data with abstracted medical records data. Periodic audits by qualified third parties of the

⁴ The group-reported rates were unaudited, whereas the plan-reported rates had passed NCQA audits of these HEDIS measures for the previous year

validity and reliability of claims and medical record information will be necessary to ensure sufficient accuracy for the data's use in quality incentive programs. There is a distinct role for government and quasi-public entities such as NCQA in this data quality assurance function, and the award of quality infrastructure grants from public sources such as the Centers for Medicare and Medicaid Services can expedite the improvement of electronic health information systems within provider organizations.

Payout Formulas. Other key design choices pertain to designing payout formulas. One particular choice is whether to reward absolute performance or relative performance. Most P4Q programs reward absolute performance based on whether a provider achieves a threshold value for one or more quality measures (e.g., 80% of diabetic patient undergo annual Hb_{A1c} testing). Each approach has certain potential benefits and liabilities. Specifically, because the attainment of absolute performance targets is more controllable by the individual provider, rewarding absolute performance constitutes a more powerful direct economic incentive than incentives based on relative performance since the latter partly depends on other providers' behavior. However, providers whose performance historically has been above the target will likely have little motivation to improve further because the status quo is sufficient to obtain the bonus. Indeed, this pattern of behavior was observed in one recent study of a California-based program that offered bonuses to medical groups for attaining predetermined quality targets (Rosenthal et al. 2005). Additionally, the all-or-nothing proposition of a bonus-type incentive may actually discourage some providers from trying to attain the target, particularly those whose past performance has been well below the level required for the bonus.

Payout formulas that reward providers based on their relative performance are uncommon at this time, though there is evidence from small area variation studies (cf., Keller et al. 1990)⁵ that physicians respond to peer comparisons, even in the absence of direct financial incentives. The Rochester Individual Practice Association/Excellus (RIPA) P4Q program is an exception in this regard. The program's financial incentive is largely funded through withholds of approximately 10 percent of participating physician fees. The potential payoff is between 50 and 150% of a physician's withhold based on physician's relative performance with respect to clinical quality, efficiency, and patient satisfaction.

Payer Collaborations. In designing P4Q programs, payers may also seek opportunities to collaborate with each other. Such collaborations may include coordinating the selection of quality targets and associated incentives to achieve greater market penetration of a P4Q program. Certainly, as payers expand the market penetration of their programs among providers, (i.e., the share of the provider's revenues affected by the program), their bargaining power with participating providers increases on such parameters as the level and structure of financial rewards.

Based on our interviews with leaders of provider organizations (Conrad and Saver 2005), it appears that they are very cautious about forming P4Q-related collaborations because of concerns that such collaborations may run afoul of the antitrust laws (i.e., price fixing and market allocation violations). However, even in the presence of these legal constraints, payers have opportunities to collaborate for purposes of selecting quality measures. IHA is an example of this type of P4Q collaboration. These collaborations potentially increase the power of the incentives and reduce administrative re-work.

Concluding Comment

This paper has attempted to frame key practical issues in designing and implementing P4Q programs, reflecting and adding to earlier work. By addressing the practical issues of design and implementation, and by comparing and contrasting actual program choices with theoretical predictions, we hope to challenge the thinking of those who organize and manage P4Q efforts and to stimulate future actionable research in this arena.

References

- Bokhour, B., J. Hook, J. Burgess, B. White, D. Berlowitz, M. Guldin, M. Meterko, and G. Young. 2006. "Incentive Implementation in Physician Practices: A Qualitative Study of Practice: Executive Perspectives in Pay-for-Performance." *Medical Care Research and Review* 63: 73S-95S.
- Baker, B., and B. Carter. 2005. *Provider Pay-for-Performance Incentive Programs: 2004 National Study Results* San Francisco: Med-Vantage, Inc.
- California Cooperative Healthcare Reporting Initiative, *P4P Clinical Measures Pilot: Final Report*. 2003. [Online article or information; retrieved 10/4/05].
<http://www.iha.org/CCHRI%20P4P%20Clinical%20Measure%20Pilot.pdf>
- Centers for Disease Control and Prevention: National Center for Chronic Disease Prevention and Health Promotion. 2004. *Chronic Disease Overview*. [Online article or information; retrieved 10/4/05.] <http://www.cdc.gov/nccdphp/overview.htm>
- Conrad, D.A., and J.B. Christianson. 2005. "Penetrating the 'Black Box': Financial Incentives for Enhancing the Quality of Physician Services." *Medical Care Research and Review* 61 (3S): 1-32.
- Conrad, D.A., and B. Saver, B. Court, and S. Health. 2006 (in press). "Paying Physicians for Quality: Evidence and Themes from the Field." *Joint Commission Journal on Quality and Patient Safety*.
- Damberg, C.L., K. Raube, T. Williams, and S.M. Shortell. 2005 (April-June). Paying for Performance: Implementing a Statewide Project in California." *Quality Management in Health Care* 14(2): 66-79.
- Eggleston, K. 2005 (December). "Economic Modeling of Methods to Stimulate Quality Improvement." *International Journal of Quality in Health Care* 17(6): 521-531.
- Frey, B.S. 1997. "On the Relationship Between Intrinsic and Extrinsic Work Motivation." *International Journal of Industrial Organization* 15: 427-39.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington D.C.: National Academy Press.
- Keller, R.B., D.N. Soule, J.E. Wennberg, and D.F. Hanley. 1990. "Dealing with Geographic Variations in the Use of Hospitals. The Experience of the Maine Medical Assessment Foundation Orthopaedic Study Group." *Journal of Bone and Joint Surgery* 72 (9): 1286-93.
- Landon, B.E., J. Reschovsky, M. Reed, and D. Blumenthal. 1998. "Personal, Organizational, and Market Level Influences on Physician's Practice Patterns." *Medical Care* 39 (8): 889-905.

Macho-Stadler, I. and J. D. Perez-Castrillo. 1997. *An Introduction to the Economics of Information: Incentives and Contracts*. New York: Oxford University Press.

McGlynn, E.A. S.M. Asch, J. Adams, J. Keeseey, J. Hicks, A. DeCristofaro, and E.A. Kerr. 2003. "The Quality of Health cCare Delivered to Adults in the United States." *New England Journal of Medicine* 348 (26): 2635-45.

Prendergast, C.R. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37: 7-63.

Rosenthal, M.B., R.G. Frank, Z. Lim, and A.M. Epstein. 2005. "Early Experience with Pay-for-Performance: from Concept to Practice." *JAMA* 294: 1788-93.

Smith, P.C. and N. York. 2004 (May/June). "Quality Incentives: The Case of U.K. General Practitioners." *Health Affairs* 23:3, 112-118.

Young, G.J., B. White, J.F. Burgess, D. Berlowitz, M. Meterko, M.R. Guldin, and B.G. Bokhour. 2005. "Conceptual Issues in the Design and Implementation of Pay-for-Quality Programs." *American Journal of Medical Quality* 20:3; 144-150.