
Emerging Issues of Pay-for-Performance in Health Care

Frederick G. Thomas, Ph.D., C.P.A. and Todd Caldis, Ph.D., J.D.

In the 1990s, managed care with capitation was regarded as the policy tool of choice that would control rising health care costs. An explicit emphasis on health care quality was largely absent, as administrative focus was on the cost per member/per month and provider organizations (POs) that fought to capture market share. In the new century, capitation has to some degree receded as the payment tool of preference as new thinking has spread through the health care marketplace. Health care is seen as entering an age in which both quality and efficiency are paramount in the purchase and delivery of health care. Taken together, the dual goals of quality and efficiency lead to a key concept: pay-for-performance (P4P). This concept is also referred to as value-based purchasing (VBP).

P4P can be broadly defined “...to include any type of performance-based provider payment arrangements including those that target performance on cost measures” (Dudley and Rosenthal, 2006). Various private sector programs have been implemented to reward providers for delivering high quality efficient care. However, these programs have been fairly limited in scope and scale. Opportunity abounds for demonstrating how P4P can be implemented to provide the right incentives to create quality, efficiency, and value in the delivery of health care.

Defining quality can be challenging. In cost effectiveness studies economists have

used the concept of the quality adjusted life years to depict the change in quality over time. This concept of value is simple: the increase in health care costs for some specific treatment is put into perspective by measuring longer life for those patients receiving the treatment, a highly targeted measure of quality. By contrast, in health care generally consumers, providers, and payers are interested in finding more direct and more timely measures of quality and value.

Quality indicators (QIs) and process measures have been developed, and continue to be refined. Such measures have entered into consumer awareness through Web-based portals. End stage renal disease patients can compare dialysis facilities on several quality measures: anemia, hemodialysis, and patient survival (<http://www.cms.gov/DialysisFacilityCompare/>). Beneficiaries can compare the QIs for heart attack, heart failure or pneumonia, or patients having surgery on Medicare’s Hospital Compare (http://www.cms.hhs.gov/HospitalQualityInits/25_HospitalCompare.asp). Quality measures for nursing home services are available on the Nursing Home Compare (http://www.cms.hhs.gov/NursingHomeQualityInits/01_Overview.asp).

Congress has also mandated that value be part of the payment equation in the purchase of Medicare hospital services. The 2005 Deficit Reduction Act required CMS to designate by October 1, 2007, hospital-acquired conditions that starting on October 1, 2008, will preclude assignment of a hospital stay to a higher paying diagnosis-related group unless it can be documented that the condition was already

The authors are with the Centers for Medicare & Medicaid Services (CMS). The statements expressed in this article are those of the authors and do not necessarily reflect the views or policies of CMS.

present on initial admission. In accordance with the law, CMS has targeted specific hospital-acquired conditions: objects left in surgery, air embolisms, blood incompatibility, catheter induced urinary infections, bedsores, vascular catheter induced infections, mediastinitis after coronary bypass, and falls occurring in a hospital ([http://cms.hhs.gov/Acute InpatientPPS/downloads/CMS-1533-FC.pdf](http://cms.hhs.gov/Acute%20InpatientPPS/downloads/CMS-1533-FC.pdf)).

The 2006 Tax Relief and Health Care Act authorized the establishment of a physician quality reporting system for Medicare, which CMS is implementing as the physician quality reporting initiative. The law builds on a voluntary quality reporting program for physicians that CMS initiated in 2006. Under this act, physicians who successfully report a set of quality measures from July 1 to December 31, 2007, may earn a bonus payment, subject to a cap, equal to 1.5 percent of total their total allowed Medicare charges for the six-month period.

Over the last year, the policy community has focused increasing attention on using episodic groupers as part of a move to a VBP system in Medicare. Episodic groupers classify patient claims into episodes of illness, making it in principle feasible to construct physician efficiency reports by types of illness. MedPAC has issued two reports in the past year related to episode groupers (http://www.medpac.gov/documents/Jun06_EntireReport.pdf). These analyses suggest that episode grouping techniques might have potential as tools to be used by Medicare to profile physicians and identify physicians with unreasonably high costs per episode.

Using annual Medicare claims data, the U.S. Government Accountability Office (2007) found substantial cost variation across patients within disease types. Extending the results of their study to a VBP perspective, it might be possible to generate physician profiles from claims data and

identify those practitioners associated with higher care costs (which could indicate less efficient practice patterns). In addition, various umbrella organizations, such as Ambulatory Care Quality Alliance and the National Quality Forum, are working to define a set of conceptual standards that can be used in constructing physician profiles.

In this special edition issue of the *Health Care Financing Review*, the five articles add to our understanding of VBP by focusing on problems of implementation and design of P4P. In the article by Cromwell, Drozd, Smith, and Trisolini, models of P4P payment arrangements are developed and simulated results are shown. A "...decision-making under uncertainty..." approach is used to structure sensitivity analysis of results. Based on the simulations, the authors conclude that the best P4P payment strategies are those that link QIs to patient outcomes, set challenging target rates of improvement, and tie bonuses to true improvements over baseline levels.

In their article Kautter, Pope, Trisolini, and Grund present bonus calculations and other design aspects of the Physician Group Practice Demonstration. A legislative mandate for the demonstrations was included in the 2000 Medicare, Medicaid, and State Child Health Insurance Program Benefits Improvement and Protection Act. Ten physician groups participated in this 3-year demonstration, which started on April 1, 2005. "The PGP Demonstration seeks to align incentives for physician groups to manage the overall care of their patients, especially beneficiaries with chronic illnesses and high-risk patients who account for a significant portion of Medicare expenditures. The demonstration provides a financial incentive similar to those used by managed care organizations and other commercial payers to reward quality improvement and encourage efficiency" (Leavitt, 2006). The article

provides valuable insights into the calculation performance bonuses and the associated problems of quality measurement and reporting.

Pope and Kautter in their article present a methodology for profiling the cost efficiency and quality of care of 30 large physician organizations and 4 physician networks in the Boston metropolitan area. They developed a profiling system that operationalizes the attribution of costs to physicians while controlling for patient-level risk and quality of care. They then construct an efficiency index to compare the relative costs across the 30 practices.

Attributing services to a particular physician is extremely important in specifying a reliable and valid physician profiling system. Physician attribution on Medicare claims proves to be a somewhat elusive assignment, as identifying information on Medicare claims may be at the physician group or tax identification levels. To identify physician networks, Pope and Kautter developed a physician assignment algorithm using a plurality of outpatient evaluation and management visits. They found that 74 percent of visits are reliably assigned by their algorithm.

Pope and Kautter standardized costs across the practices by assigning claims adjusted for risk, using the hierarchical conditions categories model, teaching and disproportionate share hospital payments. Finally, they present a composite quality score with their claims-based measures. Geographic variation is not adjusted, since the study is focused on only one metropolitan area. Using this methodology, and excluding one organization that specialized in oncology, they found that only 1 of the 30 practices has an index value that is suggestive of an “inefficient network.” They conclude that the potential savings from redirecting patients to more efficient providers may not be particularly large, since

so few organizations would be considered inefficient. However, the authors reason that the threat of losing patients may be potent in modifying physician behavior.

Davidson, Moscovice, and Remus in their article investigate the impact of hospital size on composite quality scores. It is a truism in applied statistics that the number of observational units (size) affects the variability of a sample. They use Bayesian hierarchical models to assess the impact of size on the ability to infer “true” ranks in P4P programs. The source of their research data is the Premier Hospital Quality Incentive Demonstration that originally operated for 3 years from October 1, 2003 - September 30, 2006, and is presently continuing under an extension through 2009. In the demonstration, hospitals have an incentive to report and deliver higher quality care. They are ranked by their composite quality scores, which are reported for five conditions. Hospitals in the top decile receive a bonus, a percentage of their inpatient payments, while hospitals ranked in the bottom decile incur a penalty.

The authors focused their attention on the performance of 47 rural and critical access hospitals that are participating in the demonstration. They obtained additional data for ordinal ranking of hospitals from Medicare Hospital Compare data by critical access hospitals. Their resulting analysis includes hospital rankings on three conditions which are often encountered in hospitals located in rural settings: acute myocardial infarction, heart failure, and community acquired pneumonia. Their study finds an inverse relationship between size and the statistical confidence interval, or uncertainty, around a mean ranking for the three conditions analyzed, and they conclude that hospital size should be considered in measuring quality, so as not to penalize smaller providers unduly.

Lastly, Young, Burgess, and White describe the lessons learned as the national evaluator for the Rewarding Results Demonstration, which operated in seven sites. These pay-for-quality (P4Q) programs varied by site, having individualized quality measures, financial incentives, and physician organizations. A survey of 1,500 physicians and telephone interviews with plan executives were used to gather information for this study. Some of the relevant lessons from this study include: (1) POs can develop the necessary infrastructure in response to quality incentives; (2) making physicians aware of P4Q programs is a challenge; (3) data must be accurate and valid; (4) the lack of infrastructure is a major barrier to implementing a P4Q program; (5) incentives directed to POs tend to limit physician involvement; and (6) the cost of the infrastructure may exceed the benefits, at least in the short run.

The articles in this edition of the *Review* address emerging issues of this important new area of the health services research. With a focus on value, our system will be in a state of transition to approaches that can

demonstrate quality and efficiency in the delivery of health care. As the baby boomer generation becomes eligible for Medicare, and begins the strain on the health care system within a decade, value for the health care dollar becomes even more important. To implement P4P systems, substantially more research will be needed to operationalize the concept of value in a meaningful way for providers and patients.

REFERENCES

Dudley, R.A. and Rosenthal, B.: *Pay for Performance: A Decision Guide for Purchasers*. Agency for Healthcare Research and Quality, Publication Number 06-0047, April 2006. Internet address: <http://www.ahrq.gov/qual/p4pguide.pdf>. (Accessed 2007.)

Leavitt, M.O., Department of Health and Human Services: *Report to Congress: Physician Group Practice Demonstration First Evaluation Report*. 2006

U.S. Government Accountability Office: *Focus on Physician Practice Patterns Can Lead to Greater Program Efficiency*. GAO-07-307, April 2007.

Reprint Requests: Frederick G. Thomas, Ph.D., Centers for Medicare & Medicaid Services, 7500 Security Boulevard, Mail Stop C3-19-07, Baltimore, MD 21214-1850. E-mail: fred.thomas@cms.hhs.gov