

Measurement Of And Reward For Efficiency In California's Pay-For-Performance Program

How the Integrated Healthcare Association discovered the problems of using “episodes of care” as the basis for physician performance rewards.

by James C. Robinson, Thomas Williams, and Dolores Yanagihara

ABSTRACT: Pay-for-performance (P4P) programs are expanding their purview from quality to include efficiency, and many consider the episode of care as the appropriate unit of measurement. Two years' experience by the California P4P program, however, reveals that the requisite claims data often are incomplete or poorly coded and that even large physician groups have too few patients experiencing most types of episodes to permit statistically valid measurement for public reporting and incentive payment. The California P4P program is shifting its efficiency focus to metrics not reliant on episode measurement while shifting episode measurement to supporting bundled payment for acute surgical and medical interventions. [Health Aff (Millwood). 2009;28(5):1438-47; 10.1377/hlthaff.28.5.1438]

PAY-FOR-PERFORMANCE (P4P) DERIVES FROM dissatisfaction with reimbursement methods that reward the quantity rather than the quality of health care services. It emerged as part of the backlash against managed care, which for a few years seemed to have tamed cost inflation but raised concerns about inappropriate limits on access, and from growing evidence of quality deficiencies throughout the health system.¹ California was a crucible of the revised payment method as it was for managed care, with much of the physician community organized in large group practices, insurance enrollment in health maintenance organizations (HMOs), and provider payment through capitation (monthly payments to providers for each patient, in contrast to fees for each service).

The timing of P4P—launched in 2003 for six California health plans, 6.7 million enrollees, and 230 physician groups—was perfect. The nation was happy to hear that quality rather than cost control was the new priority. Health plans nationally

Jamie Robinson (james.robinson@berkeley.edu) is the Kaiser Permanente Professor of Health Economics and director of the Berkeley Center for Health Technology at the University of California, Berkeley. He is also senior director for medical technology at the Integrated Healthcare Association (IHA) in Oakland. Thomas Williams is the executive director of the IHA. Dolores Yanagihara is IHA program director, Pay for Performance.

were eager to rebrand themselves as quality improvement entities and applied P4P to their physician and hospital networks, and a reported 148 programs were in place nationwide by the end of 2007.^{2,3} In the background, providers were consolidating and demanding higher reimbursement from health plans, which in turn were consolidating and raising premiums to employers and employees.

The turn toward quality improvement and away from cost moderation improved quality but helped fuel the resurgence of cost inflation. After historically low rates of growth in health insurance premiums in the 1990s, inflation rebounded, with premiums rising by double-digit rates in the first half of the new decade.⁴ In California, premium increases in HMO products exceeded those outside the state, eliminating the cost advantage that purchasers had enjoyed relative to their peers nationally.⁵ Enrollment in network HMO plans, the sponsors of P4P, peaked in 2003 and then began to decline.⁶ A unilateral focus on quality suddenly seemed a luxury. The “performance” in pay-for-performance needed to include economic efficiency along with clinical quality. The question became: how to measure and reward efficiency?

This paper describes the process through which the Integrated Healthcare Association (IHA), an association of health plans, hospital systems, and medical groups in California that manages the state’s P4P program, expanded its purview from quality to include efficiency in the definition of “performance” to be evaluated and rewarded. It highlights the IHA’s efforts to measure and reward efficiency using the episode of care as the unit of analysis. This paper is informed by the authors’ discussions with IHA stakeholders about their experiences with the P4P program, discussions with other individuals and organizations, and review of available documents.

Efficiency Measurement Using Episodes Of Care

An episode of care is a series of health care services related to a particular condition or event. Most episodes have a reasonably well-defined beginning and end, but for management of chronic conditions, episodes are defined as all of the condition-related services in a certain period of time—usually twelve months. Improvement in the process of care may best be motivated to the extent that the many components and contributors are measured, evaluated, and reimbursed as a single bundle or “case rate.” Doug Emery, Michael Porter, Elizabeth Teisberg, Harold Luft, and others have highlighted the virtues of using the episode of care rather than the individual intervention (as in fee-for-service) or an entire array of services (as in capitation) as the basis for measurement, improvement, and payment in health care.⁷⁻¹¹

The diagnosis-related group (DRG) system embodied in Medicare’s hospital payment mechanism is an early and relatively successful application of episode principles, even if it truncates the episode to the inpatient admission and excludes physicians from the hospital payment bundle. The Medicare Payment Advisory

Commission (MedPAC) has advocated extending DRGs from the inpatient episode to include postdischarge care, and the recent Medicare Acute Care Episode Demonstration projects bundle together physicians' inpatient services along with those of the hospital.^{12,13} An earlier Medicare demonstration for case-rate pricing of coronary artery bypass graft (CABG) surgery was considered a success but was not replicated.¹⁴

Among private insurers, Oxford Health Plan pioneered episode payment in the 1990s only to be engulfed by data and computer system difficulties that forced the abandonment of the initiative. In California, independent practice associations (IPAs) pioneered "contact capitation" that set prospective payments for the ambulatory episode of care after referral of a patient to a specialist.¹⁵ Some health plans rely on episode measurement to create tiers of preferred providers whose services are available to enrollees at lower levels of coinsurance than those of nonpreferred providers are, even within the plan's contracted network.¹⁶

Pay-For-Performance: The California Model

California always has contrasted with mainstream U.S. health care because of the dominant role of multispecialty medical groups and IPAs in the organization and delivery of physician services. These physician organizations provided a unit of measurement more reliable than the individual physician as interest among purchasers grew during the 1990s for quality measurement and reporting.

■ **The beginnings of California's P4P program.** Working through the IHA, California insurers agreed to pool data in 2003 and use a single set of independently validated quality performance metrics. Each physician organization would receive a single score on each performance metric for all of its commercial HMO patients.¹⁷ A single "report card" on performance would be developed and disseminated by the state government, and each health plan would base its financial incentives on these performance metrics using the aggregated results from all plans.

Although P4P principles were adopted by Medicare and by private insurers across the nation, the California program remains the largest in terms of dollars distributed. The measurable impact of the \$264 million spent on P4P in California between 2003 and 2007 was modest, however, because even this large sum amounted to less than 2 percent of the physician organization's annual revenues. Performance on the clinical quality metrics improved by an average of 3 percent annually, while performance as measured by patient satisfaction surveys stagnated; adoption of the P4P-specified types of information technology (IT) increased annually by 7 percent.¹⁸ A survey of physician and plan leaders by Cheryl Damberg and colleagues reported that the majority felt that the P4P program had motivated improvements in the data systems and measurement capabilities but that no "breakthrough" quality improvements had been achieved.¹⁹

■ **Inclusion of "efficiency" in P4P.** Whatever the achievements of the California P4P program in its first years, the quality focus was at risk of being sidelined by

the continuing rise in health care costs and concomitant decline in health insurance coverage. Premiums grew by an average rate of 10.9 percent for the commercial HMO during these years, and enrollment dropped by 7.7 percent, despite strong population growth in the state.^{20, 21} Facing intense pressure from purchasers, the health plans tired of quality-only P4P and demanded that efficiency be included as a dimension of performance alongside quality.

Efficiency Measurement In The IHA's P4P Scheme

The medical group structure of managed care in California theoretically resolved the three thorniest problems besetting episode initiatives in other contexts: small numbers, attribution, and inconsistent benefit designs. The physician organizations had assumed responsibilities for tens of thousands of HMO enrollees and provided the full range of services that would be included in an episode. Simply put, it would be easier to measure episode performance and clearer where to send the episode-based efficiency bonus than it would be in a fragmented environment of small physician practices. The highly regulated HMO product in California had very similar levels of consumer cost sharing across competing health plans; thus, differences in the number of episodes per patient and in the average cost per episode would not be driven by differences in benefit design.

■ **Measuring efficiency.** The IHA obtained claims data for 2006 from six health plans covering more than five million HMO enrollees.²² The data were run through the Medical Episode Grouper (MEG) software developed by Thomson Reuters, one of the firms marketing episode measurement systems. The Medical Episode Grouper assigns each claim to one of 570 episode types based on primary and secondary diagnosis (rather than procedure) codes. The IHA analytic team examined the distribution of episodes and episode spending across physician organizations by health plan, by service category (for example, professional, hospital, outpatient facility, drug), and for selected major episodes (for example, asthma, diabetes). Efficiency at the medical-group level was to be measured based on the ratio of observed-to-expected costs for each episode and across all episodes. Expected episode costs were calculated as the average cost of each episode type across all physician organizations, adjusting for each patient's severity within the episode (stage of disease) and health status outside the episode (comorbidities).

■ **Structure of efficiency rewards.** Efficiency rewards would differ from the quality-based IHA P4P program awards in one important dimension. Financial payment for efficiency improvement would not be structured as a bonus paid on top of the usual negotiated payment rates. The whole point was to pay less, not more. Rather, efficiency would be rewarded using principles of gainsharing, according to which providers retain some portion of the cost savings they generated through changes in the clinical and administrative aspects of the care they deliver. If a physician organization did not engender any reductions in spending across the measured episodes, it would not receive any efficiency-based financial reward.

Small Numbers Of Patients Per Episode Type

The primary advantage of a physician organization–based health care system, for purposes of measuring efficiency, is that organizations as the unit of observation overcome the small numbers that preclude valid episode measurement at the individual physician level. As it turned out, however, even most physician organizations lack enough patients to have a large number of episodes to measure efficiency. The IHA technical committee decided that a physician organization must have at least thirty patients experiencing a type of episode during a year for the episode results to be valid for statistical purposes.

The physician organizations included in the California efficiency measurement program cumulatively assumed responsibility for more than five million HMO enrollees, but the number of enrollees affiliated with any one organization varied widely. The average organization had 27,000 enrolled patients, with a median of 13,000 and a range of 100–350,000.²³ These large numbers of enrollees generated only modest numbers of episodes because most people are healthy most of the time, and there is a wide range of episode types.

Of the 570 types of Medical Episode Grouper episodes, only 75 were sufficiently prevalent that a majority of physician organizations had 30 or more patients in one year, and only 28 were sufficiently prevalent that two-thirds of physician organizations had enough patients. Many of the most prevalent episode types were for relatively minor conditions that did not account for a large fraction of total spending. However, selected diagnoses that were important both clinically and economically did reach the threshold of thirty patients per year for a large number of medical groups, including type 2 diabetes (5.6 percent of expenditures, 84.9 percent of medical groups), chronic maintenance of hypertension (4.5 percent of expenditures, 88.5 percent of medical groups), and chronic maintenance of angina pectoris (4.3 percent of expenditures, 66.7 percent of medical groups) (Exhibit 1). For individual physicians, even highly prevalent conditions such as diabetes do not generate enough patients per physician to support reliable public reporting and performance bonuses at the episode level.²⁴

Data Completeness

The problem of small numbers proved to be compounded by problems of claims quality and completeness. Here the physician organization as the foundation of health care proved to be a liability rather than an advantage. Physician organizations in California must submit data to HMOs that describe the “encounters” in which they provide clinical services to insured members but are not paid on the basis of these encounter data (they are paid prospectively on the basis of negotiated per-member-per-month capitation). Prior to P4P, there was little incentive for the medical group to fully code what is done to the patient (procedures) and why it is done (diagnoses) on the encounter forms. The IHA found evidence of ex-

EXHIBIT 1
The Percentage Of Medical Groups In California With More Than Thirty Patients Per Episode, For The Highest-Cost Episode Types

| Rank | Episode type | Percent of total annual cost | Percent of physician organizations with more than 30 patients per episode |
|------|--|------------------------------|---|
| 1 | Diabetes mellitus type 2 and hyperglycemic states maintenance | 5.6 | 84.9 |
| 2 | Renal failure | 5.5 | 37.0 |
| 3 | Essential hypertension, chronic maintenance | 4.5 | 88.5 |
| 4 | Angina pectoris, chronic maintenance | 4.3 | 66.7 |
| 5 | Neoplasm, malignant: breast, female | 3.2 | 39.1 |
| 6 | Delivery, vaginal | 2.5 | 63.5 |
| 7 | Osteoarthritis, except spine | 2.3 | 77.6 |
| 8 | Asthma, chronic maintenance | 2.2 | 77.6 |
| 9 | Other arthropathies, bone and joint disorders | 2.0 | 88.0 |
| 10 | Human immunodeficiency virus type 1 (HIV) infection | 1.7 | 15.1 |
| 11 | Rheumatoid arthritis | 1.5 | 39.6 |
| 12 | Neoplasm, malignant: colon and rectum | 1.4 | 18.8 |
| 13 | Delivery, cesarean section | 1.4 | 34.4 |
| 14 | Other Inflammations and Infections of skin and subcutaneous tissue | 1.2 | 90.1 |
| 15 | Other gastrointestinal or abdominal symptoms | 1.1 | 85.9 |
| 16 | Complications of surgical and medical care | 1.1 | 47.9 |
| 17 | Multiple sclerosis | 1.0 | 15.6 |
| 18 | Infections of skin and subcutaneous tissue | 1.0 | 81.3 |
| 19 | Other ear, nose, and throat disorders | 1.0 | 89.1 |
| 20 | Acute myocardial infarction | 0.97 | 16.5 |

SOURCE: Integrated Healthcare Association.

tensive underreporting in some physician organizations. The number of episodes of care per 1,000 enrolled patients per year, a proxy for data completeness, varied more than tenfold, from 183 to 1,886.

Problems of data quality and completeness were not limited to the services that fell under capitation payment. The percentage of facility claims—generally paid by the insurers on a fee-for-service basis separate from the capitation payment—that included usable procedure codes varied widely across the six participating health plans. More important was the variance in the number of secondary diagnoses recorded on the claims. For physician visits, two health plans recorded no secondary diagnoses at all, and the remaining four plans had secondary diagnoses on fewer than half the claims. Lack of secondary diagnoses diminishes the accuracy of assigning claims to episodes and limits the ability to adjust for disease severity differences within and across episodes. Major problems also arose with claims for ambulatory surgical procedures. These procedures can be performed either in hospital outpatient departments or in freestanding ambulatory surgery centers. Place of service greatly affects the pricing and hence the cost but often could not be determined from the claims.

Actual Versus ‘Standardized’ Prices

Episodes of care are composed of services; to be measured and compared, the services must be converted into expenditures. This can be done by either using the actual prices paid by the insurer to the provider or using average (“standardized”) prices paid for those types of services. Actual prices vary widely across providers for the same services, based on the negotiating position of the provider. Physicians and hospitals with few meaningful competitors extract much higher prices than do otherwise similar providers in more competitive contexts.

Physician organizations in California were reluctant to have their efficiency measured using the prices charged by the hospitals where they admitted their patients. The hospital industry in California has consolidated extensively during the past decade, and the erstwhile excess in capacity has been eliminated through a combination of facility closures, bed reductions, and population growth.⁵

The IHA thus decided to use standardized prices to measure episode efficiency; this focused attention on variation in utilization but prevented it from getting at differences in prices (that is, unit costs), both of which need to be addressed to affect cost. The IHA is still exploring the feasibility of using a measure of annual change in total cost per patient, which would be adjusted for risk and include actual dollars paid for all services provided to members of the organization. It uses actual prices and costs but does not provide the granularity of episode measures. Corporate purchasers of care, represented in the IHA efficiency process by the Pacific Business Group on Health, are pushing for this measure.

Mid-Course Corrections

During the era of managed care’s ascendancy in the 1990s, health plans and physician organizations in California had negotiated economic incentives for cost moderation, particularly with respect to inpatient services. Entities whose hospital use and spending fell below negotiated targets reaped a share of the resulting savings, with the remainder going to the health plan or the hospital, or both.¹⁵ Hospital “risk pools” of this sort fell into disfavor because they provided no quality incentive, and the P4P program was embraced by both plans and providers as a partial alternative. Most health plans continued to reward reductions in inpatient utilization, measured in terms of annual admissions or inpatient days per 1,000 enrollees, and supplemented these with utilization metrics that targeted other dimensions of performance. As the challenges inherent in episode-based metrics became evident, the IHA efficiency initiative turned to these alternatives.²⁵

The efficiency metrics that the IHA now will use alongside quality metrics in the P4P program include generic prescriptions as a percentage of all prescriptions (for ten classes of medications where generic alternatives are available), ambulatory surgery procedures that take place in freestanding centers (in contrast to hospital outpatient departments) as a percentage of total ambulatory procedures,

readmissions as a percentage of total hospital admissions, emergency department visits per 1,000 enrollees, and nonmaternity hospital admissions and days per 1,000 enrollees. Beginning with the 2009 calendar year, each health plan will determine whether to adopt the IHA program's recommendation to share with medical groups a percentage of the savings resulting from improved performance on these metrics. None of the metrics is based on episodes of care.

Distinguishing Payment From Performance Measurement

The IHA had conceptualized three related but nevertheless distinct applications for episode-based measurement within the context of P4P: clinical process improvement, public reporting, and financial reward for efficiency.

■ **Clinical process improvement.** For clinical process improvement, physician organizations were to be given their own numbers and encouraged to compare them with benchmarks of best and average performance by their peers. For this purpose, having small numbers is not a major problem because each medical group can focus on whichever episode type it chooses, taking into account sample size and the completeness of claims data as it sees fit. However, for public reporting and efficiency-based gainsharing, where the stakes of misinformation are higher, sample size and data completeness are critical to ensure reliability and validity of measurement.

■ **Episode of care.** The episode of care potentially offers more immediate application to the IHA's value purchasing program, which focuses on a limited number of high-volume and high-cost procedures in orthopedics and interventional cardiology. The value purchasing initiative is separate from P4P and seeks to improve physician-hospital relationships related to the purchase of "physician preference items" such as artificial knee and hip joints, cardiac stents, spine surgery components, and implantable defibrillators. The health plans in California are interested in providing financial incentives for cost-reducing provider initiatives in these domains, and the IHA sees episode payment as more feasible for these acute, hospital-centric areas of care than for the more heterogeneous domain of chronic conditions. The Centers for Medicare and Medicaid Services has come to similar conclusions and focused its episode payment demonstration around orthopedic and cardiac procedures.²⁶

■ **Bundled payment.** Bundled payment methods require adequately large numbers of patients within each type of episode to develop the average cost of that type relative to other types, and thus for the development of a payment schedule analogous to Medicare's resource-based relative value scale for fee-for-service payment. However, each provider organization that is paid on a bundled basis need not have a large number of patients in each episode type, because it will be paid an amount derived from the relative costs of patients in that episode type from all provider organizations. In the commercial health insurance sector, bundled payment levels might be negotiated between plans and providers rather than established by formula, but again the relative payment rate for each episode need not be derived from the cost experiences of each provider organization individually.

The completeness of the procedure and diagnosis codes on claims data is important for purposes of bundled payment, because they influence which type of episode each intervention is assigned to and how episodes are adjusted for differences in patient severity of illness. For these reasons, bundled payment is more easily applied to major surgical and acute procedures, such as hip replacement, than for diagnosis-based chronic conditions such as osteoarthritis (a cause of hip replacement procedures). In part for these reasons, both the Medicare demonstration project and the IHA bundled payment initiative are focusing on acute, hospital-based procedures.

Best But Not Perfect Method

Payment incentives create powerful but blunt tools for improving performance in health care. Retrospective methods such as fee-for-service undermine incentives for providers to do the hard work of improving efficiency and slow the otherwise rapid growth in use and spending. Prospective methods such as capitation create incentives for efficiency but provide no guidance for how to achieve it, and they can motivate providers to avoid patients most in need of resources.

Despite high hopes, performance reward based on episodes of care has proved to be limited by problems of sample size and data completeness. The use of the episode of care as the basis for a bundled provider payment method is not subject to the problems of small numbers of patients in each episode type at each provider organization, although it does face other problems, such as how to allocate the bundled payment across the participants in each episode of care. Until data systems and organizational structures evolve to overcome these limitations, episode measurement and payment validly can be used for judiciously chosen procedures in judiciously chosen settings, but not for more.

.....
The authors have been directly involved in the planning and implementation of IHA initiatives in California.

NOTES

1. Institute of Medicine. Crossing the quality chasm: a new health system for the twenty-first century. Washington (DC): National Academies Press; 2001.
2. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *New Engl J Med* 2006;355(18):1895–902.
3. Baker G, Delbanco S. Pay for performance: national perspective 2006; longitudinal survey results with 2007 market updates. San Francisco (CA): MedVantage; 2007.
4. Ginsburg PB, Strunk BC, Banker MI, Cookson JP. Tracking health care costs: continued stability but at high rates in 2005. *Health Aff (Millwood)*. 2006;25(6):w486–95.
5. Baumgarten A. California health care market report [Internet]. Oakland (CA): California HealthCare Foundation; 2007 Mar [cited 2009 Aug 4]. Available from: <http://www.chcf.org/topics/healthinsurance/index.cfm?itemID=114640>
6. Cattaneo and Stroud Inc. 2006 statewide HMO and special programs enrollment study, California statewide commercial enrollment by health plan for months of March 2000–March 2006 [Internet]. Sacramento (CA): Cattaneo and Stroud; 2006 [cited 2009 Jun 19]. Available from: <http://www.cattaneostroud.com/2006HMO.htm>

7. Emery D. Global fees for episodes of care—new approaches to healthcare financing. New York (NY): McGraw-Hill; 1999.
8. Luft H. Total cure: the antidote to the health care crisis. Cambridge (MA): Harvard University Press; 2008.
9. Porter ME, Teisberg EO. Redefining health care: creating value-based competition on results. Boston (MA): Harvard Business School Press; 2006.
10. Mechanic RE, Altman SH. Payment reform options: episode payment is a good place to start. *Health Aff (Millwood)*. 2009;28(2):w262-71.
11. de Brantes F, Camillus JA. Evidence-informed case rates: a new health care payment model. New York (NY): Commonwealth Fund; 2007.
12. Hackbarth G, Reischauer R, Mutti A. Collective accountability for medical care—toward bundled Medicare payments. *New Engl J Med*. 2008;359(1):3-5.
13. Medicare Payment Advisory Commission. Report to the Congress: reforming the delivery system. Washington (DC): MedPAC; 2008.
14. Goodroe JH, Murphy DA. Contracting for episodes of care: a successful model. In: Emery D, editor. Global fees for episodes of care—new approaches to healthcare financing. New York (NY): McGraw-Hill; 1999. Chap. 14.
15. Robinson JC. The corporate practice of medicine: competition and innovation in health care. Berkeley (CA): University of California Press; 1999.
16. Robinson JC. Hospital tiers in health insurance: balancing consumer choice with financial incentives. *Health Aff (Millwood)*. 2003;22:w135-46.
17. Health plans operating in California have various physician performance bonus programs for their products, but they do not pool performance data across plans, rely on a single set of independently validated metrics, or coordinate with the performance incentives for commercial HMO enrollees.
18. During the period 2003-07, some clinical quality measures had specification changes, and additional measures were added. An overall average annual percentage improvement for clinical quality measures was calculated by the IHA considering these changes.
19. Damberg CL, Raube K, Teleki S, dela Cruz E. Taking stock of pay-for-performance: a candid assessment from the front lines. *Health Aff (Millwood)*. 2009;28(2):517-25.
20. California HealthCare Foundation. California employer health benefits survey, 2008. Oakland (CA): CHCF; 2008.
21. Commercial HMO enrollment in California for 2007 provided by Craig Paxton, Cattaneo and Stroud Inc. (personal communication).
22. Although Kaiser Permanente participates in the IHA P4P program for purposes of performance measurement and public reporting, it does not participate in the performance reward component. The Kaiser health plan is vertically integrated with the Permanente medical groups that serve its enrollees and has its own set of financial metrics and relationships with those groups. Kaiser does not pay Permanente using claims data and hence does not contribute data or otherwise participate in the efficiency gains sharing bonus program. One of the network health plans that does participate in the financial as well as the public reporting components of P4P was not able to contribute claims data to the efficiency measurement initiative.
23. These numbers relate to the number of commercial HMO enrollees who select the medical group or IPA for their physician services. The physician organizations also treat various combinations of Medicaid, Medicare, and commercial PPO enrollees.
24. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician “report cards” for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281(22):2098-105.
25. Lewis C. P4P program expands focus to include cost measurements. *California health plan analysis (HealthLeaders-InterStudy)*. 2009 Winter; 8(1):1-5.
26. Centers for Medicare and Medicaid Services. CMS announces demonstration to encourage greater collaboration and improve quality using bundled hospital payments [Internet]. Baltimore (MD): CMS; 2008 May 16 [cited 2009 Aug 4]. Available from: http://www.cms.hhs.gov/apps/media/press_releases.asp