

PAY-for-PERFORMANCE

Hit or Myth?

In this article...

Examine why some think pay-for-performance programs need significant restructuring before they can make an impact and improve health care.

By Alan P. Marco, MD, MMM, CPE, FACPE

The idea that placing financial incentives on the performance of physicians, hospitals, or other components of the health care system will lead to immediate and substantial improvement in performance—the care of patients—appeals to common sense.

However, it is controversial whether or not this will actually happen. As P4P is currently constructed, it is largely a medical myth, much like other things that physicians do or recommend that are just “common sense” such as shaving the operative site before surgery (which can actually increase infection rates).¹

Careful assessment of the evidence behind P4P in its current form will reveal that it is largely mythical and the link between pay and performance is not as clear as the public, physicians or payers would hope.

The inherent problem with P4P is that most of the “quality measures” are not measures of quality, but of process.² Even the Physician Quality Reporting Initiative (PQRI) measures, which the Centers for Medicare and Medicaid Services (CMS) call “quality measures,” are often process measures.³

Continued on 36

Why it Won't Work

In this article...

Consider how market forces and social conscience could be affecting the success of pay-for-performance programs.

By Kent Bottles, MD

“If one wishes to be a true scientist—an explorer not in search of what one desires to be true but rather in search of whatever truth there is—then one must be willing to accept, to engage, even to pursue further the most unwelcome and confounding data. One must be willing to make discoveries that shatter one’s most deeply held beliefs. Maybe it turns out that Earth is not the center of the universe.”

Rivka Galchen, *Atmospheric Disturbances: A Novel*¹

The conventional wisdom is that pay-for-performance programs make a lot of sense. If physicians and hospitals are motivated by profit and if there is room to improve on the quality of medical care, then pay for performance will motivate providers to provide more value.

There are four major reasons why pay-for-performance programs are gaining popularity:

1. The idea that people are more likely to do something if they get paid more to do it has a strong intuitive appeal.

Continued on 36



Hit or Myth? continued

Some cost-containment programs masquerade as P4P.

For example, measure number 4, Screening for Future Fall Risk, is a process measure. Physicians are asked to report only that they “screened” the patients (a process), not whether they actually reduced falls (an outcome). Similarly, measures 5 through 8 refer to percentages of patients for whom certain medications are prescribed. This could be an outcome, but wouldn't it be more effective if the physician actually got the patients to take their medications rather than just get credit for writing a prescription?

While such an outcome measure has elements that are seemingly beyond the physician's control, such as patient compliance, that just means that the standard should not be an arbitrarily high 95 percent or some such. Many others of these measures are for reporting that screening or communication has taken place, but are not actually measures of improved care.

If one screens for a disease, but does nothing, credit would be given under PQRI without impacting patient outcomes. One review showed that of 17 studies examining the role of explicit financial incentives in improving the quality of health care, 13 were of process measures, not actual outcome measures.⁴

However, there are some studies supporting the notion that improvement in process measures can improve outcomes also.⁵

What's up in Hawaii?

A classic version of P4P as a “myth-understood” concept is on the MedQIC Website (www.medqic.org). Following the links through the Hospital-SCIP-Stories buttons brings the viewer to the news story of how Hawaii Medical Center East improved its performance by achieving a 0 percent surgical infection rate.

While this is laudable, close reading of the story as reported leads one to a different conclusion.⁶ The title is correct: “Hawaii Medical Center East Improves Quality Measures Through Participation in Surgical Care Improvement Project,” but the conclusion is wrong.

According to the report, Hawaii Medical Center East reported a 0 percent infection rate for the third quarter of 2005. They also noted that they did not follow the SCIP

Why it Won't Work continued

Ariely writes that we live in two worlds with different sets of rules. In one world social norms prevail; in the other world market norms prevail.

2. There are large gaps in quality and delivery of evidence-based guideline recommended care. McGlynn, and others, famously found an overall adherence with recommended adult care of only 55 percent.²
3. There is a lack of a relationship between quality and costs at a regional level.
4. The devastating impact of increasing costs on American companies have hampered their ability to compete in a global marketplace.

Conventional wisdom often masks “unwelcome and confounding data” because the subject has not been explored completely. Insights gained from neuroscience and behavioral economics question whether pay for performance will work; some believe that “extrinsic incentives” (financial compensation) can conflict with “intrinsic incentives” (the moral command to do one's duty). These investigators believe extrinsic incentives can crowd out intrinsic incentives and result in failures to do one's moral duty.³

O. Brafman and R. Brafman in *Sway* describe how the Swiss government identified a small town that seemed to be the ideal location for a necessary nuclear waste depository.

When they presented the plan to the town hall meeting, about half the citizens said they would approve the plan and about half said not in my backyard.

Believing they needed more public support, the government presented a plan to give all residents 5000 francs a year if the waste dump was built in their town. With financial compensation, the percentage of Swiss citizens who would approve the plan went from 50.8 percent to 24.6 percent, hardly the result that the officials anticipated. Raising the compensation above 5000 francs did not solve the problem.⁴

Dan Ariely in *Predictably Irrational* writes about a day care center with a vexing problem of parents arriving late to

Hit or Myth? continued

standards and undertook a broad restructuring of their perioperative process to improve their compliance with SCIP measures.

But, with a “baseline in July 2005... of 0 percent,” what were they improving? Certainly not outcomes—they just improved their processes. Yet, this “improvement” is touted as a shining example of the success of SCIP.

Some P4P programs, such as the Hospital Quality Initiative (HQI), can be successful. When used to create incentives for health systems, improvements in reported measures that more directly relate to patient outcomes, such as aspirin administration for acute myocardial infarction or “door-to-balloon” time in acute coronary syndrome, can be seen.⁷

However, in its current iteration, the HQI pays for reporting rather than improvement, although future linking outcomes to financial rewards is expected.⁸ A study assessing the validity of the Hospital Quality Incentive Demonstration (HQID) project showed that a hospital’s “performance” is highly dependent on process measures rather than outcome measures, with only 4 percent of the variability due to changes in the outcome of “survival.”⁹

Still, programs that rely on patient compliance, as many outpatient measures do, place the physician in the awkward position of bullying the recalcitrant patient or cherry-picking the patients who are most agreeable to the physician’s advice.

In a recent survey, 82 percent of physicians responding were concerned about unintended consequences of P4P such as avoiding high-risk patients.¹⁰

Yet, P4P programs can be useful. While there are significant costs involved both in the collection of data and the increased financial incentives, the cost for each quality-adjusted life year is well within that considered to be a good investment.¹¹

Other groups studying integrated systems have demonstrated savings to the payer.¹² Still, at least in California, a leader in the P4P movement, payers have yet to save money.¹³

Incentive programs organized on the institutional (hospital or health system) level with sufficient resources and leadership commitment can improve quality of care.^{14,15} As an industry, health care must keep in mind that the patient is the ultimate payer both through the use of direct payments (co-pays and self-insurance), indirect payments (employer-sponsored health care plans in lieu of direct compensation), and governmental mediated financial transfers (e.g., taxes).

P4P programs need to support patients’ free choice in providers. Without such choice, P4P programs may be as doomed as the cost-containment strategies of gatekeepers and capitation used by HMOs.¹⁶

Why it Won't Work continued

pick up their children after school. The board of directors imposed a fine on parents who were late to pick up their children. The fine made the matter worse, and when the board removed the fine parents continued to be late. Social relationships are hard to reestablish once they have been redefined.⁵

Going in circles

Ariely reports on studies that seem to find that paying research subjects does not always improve performance. In one such investigation the performance task was to move circles from one side of the computer screen to the other in a set amount of time.

Subjects paid \$5.00 moved 159 circles; subjects paid \$0.50 moved 101 circles; but unpaid volunteers moved 168 circles.

Ariely also reports that AARP requests to lawyers to lower their hourly rate of billing to \$30.00 for poor seniors were unsuccessful, but many lawyers were happy to perform the work for free.⁵

In addition, Ariely describes research where one set of subjects read the sentence “it’s cold outside” and another group read “high paid salary” before performing a set of tasks. Those assigned the second phrase took more time to ask for help, were less willing to help others, avoided working in teams, spent more time alone, and sat farther away from other research subjects.⁵

In medicine this tension between financial compensation and moral obligations has been played out in the arenas of blood donation and organ transplantation.

R. M. Titmuss’ *The Gift Relationship* compared blood donation in the United Kingdom (where the sale of blood is illegal) and the United States (where the sale of blood is legal). Titmuss found that the percentage of people who donated blood and the amount donated increased in the United Kingdom, when compared to the United States.

He concluded that the “commercialization of blood represses the expression of altruism.”⁶ S. M. Rothman and D. J. Rothman make a similar case in arguing against establishing a market for organs for transplantation.³

What are these peer-reviewed studies telling us about the conventional wisdom that paying humans more will result in desired human behavior? Perhaps Richard H. Thaler and Cass R. Sunstein are on to something when they emphasize the importance of choice structures: how problems are presented to people.⁷

Ariely writes that we live in two worlds with different sets of rules. In one world social norms prevail; in the other world market norms prevail.⁵

Hit or Myth? continued

Masquerade

Some cost-containment programs masquerade as P4P. One example, according to the American Medical Association, is UnitedHealthcare's (UHC) Premium Designation Program (PDP).

This appears to the public as a rating system (two stars must be better than one), but is more accurately labeled a cost-containment program since there is no direct economic link to performance and achieving the higher two-star rating is based on an economic efficiency rating without appropriate risk adjustment.¹⁷

Medicare has proposed modifying its P4P plan into a system of withholds for less than stellar performance by including "an incentive payment that makes a portion of the base DRG payment contingent on performance."¹⁸

If there were relevant outcome rather than process measures involved, this could be structured as a form of P4P rather than the current pay-for-reporting. Under this proposal, middle-performing hospitals have the most to lose, since the low-performing hospitals gain the most from potential improvement and high-performing hospitals already achieve the maximum incentive.

The key issue will be whether or not such a plan is tied to true patient outcomes rather than process measures. If process measures continue to be used, will they be linked through real evidence to improved patient outcomes? That is unclear, and the proposed plan even states that "[t]here is even less evidence of the effect of P4P on patient outcomes," which certainly casts doubt on the claim that "value-based purchasing" (the new moniker for P4P) will improve patient care, especially if it is based on lower base rates for care rather than new money.

Without new money in the system, these P4P plans are actually pay-reduction plans as the "incentive" pool is structured so only a fraction of eligible hospitals receive the full incentive.

What drives improvement?

One final aspect of the P4P myth is that it is P4P programs rather than public reporting of data that makes a difference. Certainly, everyone wants to look good in publicly reported measures, regardless of their validity.

Lindenauer reports that hospitals engaged in P4P programs in addition to public reporting achieved only modest additional gains over those merely reporting their data.¹⁹ Gains were inversely related to the baseline performance of the hospital.

Similarly, a study of Massachusetts physician groups failed to show significant gains from P4P rather than the

Why it Won't Work continued

On the brain

Brain imaging studies take this line of thinking one step farther. The pleasure center in the nucleus accumbens lights up when we are engaged in sex, drug use, gambling, and thinking about money. Is this the world where market norms prevail?

The altruism center in the posterior superior temporal sulcus is responsible for how we perceive others, how we create community, and how we relate to fellow citizens. Does this center explain how we behave when social norms prevail?

The latest neuroscience studies conclude that the pleasure center and the altruism center cannot both function at the same time; one of them has to be in control.⁴

Ariely has the simplest and best explanation that summarizes my line of argument. We humans get in trouble when we mix social and market norms. Ariely writes that you will get into trouble if, after a delicious Thanksgiving dinner cooked by our mother-in-law, you try to show our appreciation by paying her \$400.00.⁵

Should we at least try to understand what these studies are trying to tell us about the likelihood of success of pay-for-performance programs or should we just ignore the inconvenient scientific findings because conventional wisdom tells us such programs make perfect sense?



References

1. Galchen R. *Atmospheric Disturbances: A Novel*. New York, Farrar, Strauss and Giroux, 2008.
2. McGlynn EA. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348; 2635, June 26, 2003.
3. Rothman SM and Rothman DJ. "The Hidden Cost of Organ Sale." *American Journal of Transplantation* 2006; 6: 1524-1528.
4. Brafman O and Brafman R. *Sway*. New York, Doubleday, 2008.
5. Ariely D. *Predictably Irrational*. New York, Harper, 2008.
6. Titmuss RM. *The Gift Relationship*. New York, Vintage Books, 1971.
7. Thaler RH and Sunstein CR. *Nudge*. New Haven, Yale University Press, 2008.

Kent Bottles MD

President of the Institute for Clinical Systems Improvement in Bloomington, Minnesota.



Hit or Myth? *continued*

general improvement in care in during that time period in Massachusetts.²⁰

In a study of the National Health Service in Britain, it was observed that competition actually worsened outcomes in emergency admissions for acute myocardial infarction.²¹ The authors postulate that since wait times were measured and reported but mortality rates were not, resources were redirected into areas that improve reported scores rather than the outcome of improved survival.

Thus, P4P programs need to ensure that the Hawthorne Effect is not the primary effect induced and that unintended consequences do not result from implementation of these programs.

The final chapter on P4P has yet to be written. Like many myths, there is some basis in reality. Under certain circumstances, P4P as it is currently constructed could improve outcomes. While P4P is here to stay, there are still significant structural and procedural issues to resolve before P4P can be described as a hit rather than a myth.



References

1. Celik SE, Kara A. Does shaving the incision site increase the infection rate after spinal surgery? *Spine* July 1, 2007; 3(45): 31(3): 1575-7.
2. Sage WM, Kalyan DN. Horses or Unicorns: Can Paying for Performance Make Quality Competition Routine? *Journal of Health Politics, Policy and Law*, June 2006;531-60.
3. <http://www.cms.hhs.gov/pqri/> accessed February 24, 2008
4. Peterson LA, LeChauncy DW, Urech T, and others. Does Pay-for-Performance Improve the Quality of Health Care? *Ann Intern Med*. Aug. 15, 2006;145(4):265-272.
5. Millett C, Gray J, Saxena S, et al. Impact of a pay-for-performance incentive on support for smoking cessation and on smoking prevalence among people with diabetes. *CMAJ* June 5, 2007;176(12):1705-10.
6. Hawaii Medical Center East Improve Quality Measures Through Participation in Surgical Care Improvement Project <http://medqic.org/dcs/ContentServer?cid=1205442145582&pagename=Medqic%2FMQNews%2FNewsFeatureTemplate&c=MQNews> accessed January 23, 2009.
7. <http://www.premierinc.com/quality-safety/tools-services/p4p/hqi/resources/hqi-whitepaper-year2.pdf> accessed January 23, 2009.
8. Smoldt RK and Cortese DA. Pay-for-Performance or Pay for Value? *Mayo Clin Proc.* February 2007;82(2):210-3.
9. O'Brien SM, and others. "Exploring the behavior of hospital composite performance measures: an example from coronary artery bypass surgery." *Circulation*, Dec. 18 2007, 116(25) 2969-75.
10. Casalino LP, Alexander CG, Jin L, Konetzka RT. General Internists' Views On Pay-For-Performance And Public Reporting Of Quality Scores: A National Survey. *Health Affairs* March-April 2002; 26(2):492-9.
11. Nahra TA, Reiter KL, Hirth RA, et al. Cost-Effectiveness of Hospital Pay-for-Performance Incentives. *Medical Care Research and Review*, Vol. (Supplement to February 2006;63(1): 49S-72S.
12. Curtin K, Beckman H, Pankow G, et al. Return on Investment in Pay for Performance: A Diabetes Case Study. *Journal of Healthcare Management* Nov/Dec 2006; 51(6) p 365-76.
13. Terry K. Is P4P Getting Tougher? *Medical Economics*. July 20, 2007; 84(14): 36-8.
14. Sautter KM, Bokhour BG, White B, et al. The Early Experience of a Hospital-Based Pay-for-Performance Program. *Journal of Healthcare Management* March/April 2007; 52(2): 95-107.
15. Grossbart SR. What's the Return? Assessing the Effect of "Pay-for-Performance" Initiatives on the Quality of Care Delivery. *Medical Care Research and Review*, Vol. 63 No. 1, (Supplement to February 2006) 29S-48S.
16. Safavi K. Patient-Centered Pay for Performance: Are We Missing the Target? *Journal of Healthcare Management* July/August 2006 ;51(4): 215-8.
17. UnitedHealthcare's "Premium Designation" program. Available at http://www.ama-assn.org/ama1/pub/upload/mm/368/uhc_pd_chart.pdf accessed Jan. 23, 2009.
18. Plan to Implement a Medicare Hospital Value-Based Purchasing Program. <http://www.cms.hhs.gov/AcuteInpatientPPS/downloads/HospitalVBPPlanRTCFINALSUBMITTED2007.pdf> accessed Jan. 23, 2009.
19. Lindenauer, PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. Public Reporting and Pay for Performance in Hospital Quality Improvement. *N Engl J Med* Feb. 1, 2007;356:(5)486-96.
20. Pearson, SD, Schneider EC, Kleinman KP, Coltin KL, Singer JA. The Impact Of Pay-For-Performance On Health Care Quality In Massachusetts, 2001-2003. *Health Affairs*, Jul/Aug2008, p1167-76.
21. Propper C, Burgess S, Gossage D. Competition and Quality: Evidence from the NHS Internal Market 1991-9. *Economic Journal*, Jan. 2008, 118(525): p138-70.

Alan P. Marco MD, MMM, CPE, FACPE

is professor and chair of the department of anesthesiology at the University of Toledo in Toledo, Ohio.

alan.marco@utoledo.edu

