

## PAY FOR PERFORMANCE

# Unintended Consequences: Getting the Most While Avoiding the Worst

*Kaveh Safavi, M.D., chief medical officer, Solucient, Evanston, Illinois*

**T**his year we examined a variety of issues that complicate a seemingly straightforward and intuitive premise: pay should somehow be tied to performance. The pay-for-performance experience, both in the United States and abroad, has identified a number of unintended consequences that should be considered by those advocating specific pay-for-performance approaches.

One study reviewed (through PubMed) 3,256 articles published over the last 15 years that address the relationship between financial incentives and high-quality healthcare; the review yielded only 17 studies that provide empiric results allowing a reasonable comparison of financial incentives to no incentives (Peterson et al. 2006). Four of those studies showed unintended consequences.

One of the challenges of all incentive programs is creating rules that can accidentally create the wrong result. To help increase the likelihood that pay-for-performance approaches will accomplish the desired goals without the inadvertent problems, this column presents a number of lessons from incentive programs that currently exist.

### KNOW THE PITFALLS

A number of studies document the problems with well-intentioned pay-for-performance programs. Four categories are discussed here.

**Gaming the system.** In the United Kingdom, an ambitious National Health Service (NHS) initiated in 2004 a national pay-for-performance program for all primary care physicians. The program tracked 146 quality indicators related to clinical care, organization of care, and patient experience. Significant incentives were made available, giving primary care physicians a chance to increase their income by 25 percent. A review of the program's first-year results showed that excluding patients from the denominator used for calculating performance, while not extensive, was the strongest predictor of achievement (Doran et al. 2006). Peterson and colleagues (2006) found three studies that indicate that documentation, rather than actual use of preventive services, improved, in a statistically significant fashion, with financial incentives.

**Focusing on the wrong thing.** This category is sometimes also known as "teaching to the test." Providers who attempt to maximize the measured attributes of care may, in fact, waste resources at the expense of actions that could yield a better result. An example of this is retinal eye exams for diabetics. Hayward and colleagues

(2005) focused on patients who received photocoagulation for diabetic retinopathy, evaluating their records for suboptimal treatment. None of the suboptimal care the researchers found involved patients whose screenings were two or three years apart, even though current guidelines call for annual screening. However, one third of those who never had screening presented with advanced retinopathy. In other words, all the efforts used to screen diabetics annually instead of every three years diverted resources that could have been used in finding never-screened patients for whom outcomes could really be improved.

Hayward and colleagues also found that the majority of suboptimal treatments were not the result of screening failures but factors like poor follow-up of known diseases (32 percent), delay in scheduling tests or treatments (14 percent), and rapid disease progression (23 percent). This finding points out the complex relationship between any single process of care and desired outcomes. It is also a reminder that even "gold standard" consensus guides are not bulletproof.

The consensus-building process is inherently political, often conservative, and reflects the agendas of competing interests that draw broad policy inferences from statistical analysis. In an even more general sense, pay for performance has been criticized for distracting hospital and physician leaders from the real efforts of performance improvement by overfocusing attention on maximizing a few discrete measures.

**Overtreatment bias.** A third risk of performance-incentive measures is that they can drive overtreatment that may run contrary to the goals of improved quality or efficiency. A study of Medicare beneficiaries discharged with a diagnosis of pneumonia found that 22 percent of these patients were treated for pneumonia within four hours after admission, although the diagnosis of pneumonia had not been definitely made (Metersky et al. 2006). The researcher suggests that insisting all pneumonia patients receive treatment within four hours of admission would mean that many patients would be treated before a final diagnosis is made and therefore may be treated in error.

The challenge here is contained in the limits of measurement. The intention is to treat patients as soon as a diagnosis is confidently made. However, all we may have available to count is a discharge diagnosis. A look at the patient's record may show different doctors rendering different diagnosis of the same symptoms. Therefore, we are unable to determine when the diagnosis became certain in order to start the clock. This reality will have to be part of any pay-for-performance measure that evaluates the timing of care.

**Setting the wrong targets.** The fourth challenge is setting the proper level of performance. How high should this bar be? Should we count relative improvement or absolute performance? How big should the reward be? For example, setting the bar too low might make a program attractive but unsustainable. This is a situation where one wins the battle but loses the war. In the NHS program, for example, English family practitioners attained a median 96.7 percent of available points

that far exceeded the 75 percent budget target. This increased the average family practice physician's income by \$40,000 U.S., or almost a 30 percent increase in earned income in one year (Doran et al. 2006). This caused a significant shock to the taxpayers and has resulted in a renegotiation of both the targets and the base compensation of family practitioners to rein in these unintended increases (Foster 2006).

Another target-setting problem is balancing the goal of incentives for improvement over incentives for threshold performance. Some studies have shown that those with the lowest baseline performance improved the most. These providers would be penalized if the payment was tied to absolute performance. Other programs have been shown to reward those who were already high performers without driving any improvement. Similarly, a balance between process and outcome measures must be considered. Process measures are easy to measure and influence. Outcomes are influenced by many things, not just the performance of the party who is measured.

## AVOID THE PITFALLS

Several lessons about using financial incentives to improve performance can be learned from both inside and outside of healthcare settings.

***Use pay for performance to support a vision.*** Pay for performance is a tool, not a goal. Pick tools that support a vision. Providers are faced with a variety of pay-for-performance programs they can participate in. Even private payers may often insist on pay for performance but may leave the specific measures to a negotiation between the parties.

Jack Billi, M.D., has led the University of Michigan and the Michigan State Medical Society in a number of forward-thinking pay-for-performance initiatives. He notes that the university selectively participates in programs that advance its vision of quality, satisfaction, efficiency, and affordability. A related view comes from Kevin Murphy, Ph.D., professor and vice dean at the University of Southern California Marshall School of Business. Professor Murphy is widely published on executive compensation structures for public companies. He has observed that successful pay-for-performance plans embrace a single "governing objective." In fact, he suggests that having multiple objectives does not help leaders know where to make trade-offs between competing values (Murphy 2006).

All of this speaks to the importance of having a clear purpose before selecting tactics. With respect for pay for performance, this approach forces a dialog about the true intent of programs. Is the true goal to make payment more fair, or is it to improve quality or drive out waste and inefficiency? For example, at some point, we might find that improved quality will not make a payment system more fair or more efficient. How will that affect what goals we set and how to measure them?

***Focus on principles, not on rules.*** One way to avoid gaming the system is to make sure everyone understands that all measures have limitations and that the real test

of performance is when progress toward the vision is made. The public accounting field is facing a similar problem in the post-Enron era. Corporate bankruptcies and public accounting scandals resulted in extensive rule making and high reporting costs for all public companies. By most accounts, the effort's intent is good, but the execution may have overshot its mark. A movement back to center, using the notion of "principle-based" accounting rather than "rule-based" accounting, is now underway. For executing performance incentives, Professor Murphy advises companies to retain some subjectivity to protect against measures that are too broad or too narrow. Many of the inherent measurement pitfalls listed here can be avoided by having up-front clarity about principles before selecting the specific rules for keeping score.

***Make the plan easy to use.*** This means easy to understand, easy to explain, and easy to compute. The more complex the measure, the more subject it is to gaming or argument. An inevitable trade-off exists here. Healthcare is a complex activity, and trying to translate this complexity into a web of measures can lead to an intricacy that is unpersuasive and unintelligible. If the real goal is to change human behavior, the strategist must keep in mind the failings of human cognition. Selecting the right measure is different from finding the perfect measure.

***Understand the role of money.*** The role of money as an incentive should not be overestimated or underestimated. The evidence is clear: Some people respond positively to financial incentives, while others respond negatively. We just do not know in advance who are influenced and in what way. Unfortunately, pay-for-performance programs are designed to affect large numbers of providers and operate under the assumption that these providers are all receptive to financial incentives in the same way the program's designers think they are. We know for a fact, however, that practice-pattern variations exist among physicians who are compensated identically for certain services. Thus, not all variations arise for monetary reasons and not all disappear with monetary solutions.

For some providers, merely reporting measures publicly is sufficient to stimulate improved practice. Others want to see the money. If money matters, then the size and timing of payments may matter also. For example, end-of-year payments may not be as effective as interim or concurrent payments (Peterson et al. 2006). What this really means is that what we call pay for performance is less about pay and more about measuring performance and making it more transparent, with payment being targeted to motivate a subset of providers for whom payment matters and changed behavior is a goal.

## **FUTURE OF PAY FOR PERFORMANCE**

The healthcare industry's payment system will evolve into something more sophisticated than simple activity-based payment. It will move to reflect the relationship between money spent and value received. This is particularly true given the large percentage of healthcare expenditure that passes through third-party payers and the

imperfect market for healthcare services. Providers have two options: (1) wait for a program to be forced on the organization or (2) advocate for an approach that supports its vision. The best way any program can achieve the benefits of pay for performance without exaggerating the pitfall of unintended consequences is to examine how pay for performance should look rather than how pay for performance should not look.

## References

- Doran, T., C. Fullwood, H. Gravelle, D. Reeves, E. Kontopantelis, U. Hiroeh, and M. Roland. 2006. "Pay-for-Performance Programs in Family Practices in the United Kingdom." *The New England Journal of Medicine* 355 (4): 375–84.
- Foster, A. 2006. "Experience of Primary Care Physicians in England." National Pay for Performance Summit, February 6–9, Los Angeles, California. [Online information; retrieved 8/30/06.] [www.ehcca.com/presentations/p4psummit/foster.pdf](http://www.ehcca.com/presentations/p4psummit/foster.pdf).
- Hayward, R., C. Cowan, V. Giri, and F. Makki. 2005. "Causes of Preventable Visual Loss in Type 2 Diabetes Mellitus: An Evaluation of Suboptimally Timed Retinal Photocoagulation." *Journal of General Internal Medicine* 20: 467–69.
- Metersky, M., T. Sweeney, M. Getzow, F. Siddiqui, W. Nsa, and D. Bratzler. 2006. "Antibiotic Timing and Diagnostic Uncertainty in Medicare Patients With Pneumonia." *CHEST* 130: 16–21.
- Murphy, K. 2006. "Pay for Performance Lessons Learned from Other Industries." National Pay for Performance Summit, February 6–9, Los Angeles, California. [Online information; retrieved 8/30/06.] [www.ehcca.com/presentations/p4psummit/murphy.pdf](http://www.ehcca.com/presentations/p4psummit/murphy.pdf).
- Peterson, L., L. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does Pay-for-Performance Improve the Quality of Health Care." *Annals of Internal Medicine* 145 (4): 265–72.

For more information on the concepts in this column, please contact Kaveh Safavi at [ksafavi@Solucient.com](mailto:ksafavi@Solucient.com).