

# Using the Lessons of Behavioral Economics to Design More Effective Pay-for-Performance Programs

Ateev Mehrotra, MD; Melony E. S. Sorbero, PhD, MS, MPH;  
and Cheryl L. Damberg, PhD

The use of pay-for-performance (P4P) incentives in healthcare is widespread. In the United States, P4P incentives are used by half of all commercial health maintenance organizations and are found in contracts with ambulatory physicians, hospitals, and nursing homes.<sup>1-5</sup> Numerous state Medicaid programs also use P4P incentives, and the proposed Medicare hospital P4P program could include incentives totaling more than \$3 billion annually.<sup>6</sup> In the United Kingdom, almost 25% of family practitioner income is tied to P4P incentives.<sup>7</sup> Despite the widespread application of P4P, much of the published literature<sup>2,3,7</sup> on the effect of P4P has concluded that these incentives have resulted in small or no improvements.

There have been various interpretations of these results. Some investigators have raised concern that the premise underlying P4P is flawed.<sup>8</sup> Others researchers believe that the magnitude of the incentives has been insufficient.<sup>9</sup> Another potential reason for failure is that the current design of P4P programs does not reflect the psychology of how people respond to incentives. This is not surprising, as there has been scant literature<sup>10,11</sup> on the effectiveness of specific design features of a P4P program to guide health plans or government sponsors. Program sponsors do what seems reasonable, and there is great variation in the design of programs.<sup>12</sup> The behavioral economics literature reviewed herein can serve as a useful guide on how to structure provider incentives.

In this article, we discuss several design alternatives drawn from the behavioral economics literature that we believe could lead to greater provider response for the same amount of money devoted to a P4P program. We start by describing the importance of design features and a prototypical P4P program being used today. We then discuss several design features that could improve P4P programs. Last, we discuss some potential unintended consequences of P4P and how design changes could minimize these.

## The Goal of P4P and Why Design Features Matter

The primary goal of most P4P programs is to improve healthcare quality, but incentives have been applied for other goals, including improving patient experience, implementing electronic prescribing, increasing patient safety, and decreasing utilization.<sup>12</sup> In this article, we focus on an illustrative physician P4P program that focuses on healthcare quality, but we believe that our recommendations

**Objectives:** To describe improvements in the design of pay-for-performance (P4P) programs that reflect the psychology of how people respond to incentives.

**Study Design:** Investigation of the behavioral economics literature.

**Methods:** We describe 7 ways to improve P4P program design in terms of frequency and types of incentive payments. After discussing why P4P incentives can have unintended adverse consequences, we outline potential ways to mitigate these.

**Results:** Although P4P incentives are increasingly popular, the healthcare literature shows that these have had minimal effect. Design improvements in P4P programs can enhance their effectiveness.

**Conclusion:** Lessons from behavioral economics may greatly enhance the design and effectiveness of P4P programs in healthcare, but future work is needed to demonstrate this empirically.

*(Am J Manag Care. 2010;16(7):497-503)*

**In this article**  
Take-Away Points / p498  
[www.ajmc.com](http://www.ajmc.com)  
Full text and PDF

**For author information and disclosures,  
see end of text.**

### Take-Away Points

Although pay-for-performance (P4P) incentives are widely used in healthcare, the published literature has shown that P4P incentives result at best in only modest improvements.

- The design of P4P programs generally does not reflect what is known about the psychology of how people respond to incentives, which may contribute to a lack of success.
- This article discusses design improvements that can enhance the effectiveness of P4P programs and potentially mitigate the risk of adverse unintended consequences.

menu of options to be considered and are not meant to be applied altogether.

### Seven Design Features That Could Improve P4P Programs

***A Series of Small Incentives Is Better Than 1 Large Incentive.*** Why do

people go across town to save \$10 on a

clock radio but not to save \$10 on a large-screen TV? In both cases, \$10 is saved, but \$10 is not always viewed the same.

It is believed that an individual perceives the difference between \$0 and \$10 as being greater than the difference between \$100 and \$110.<sup>14</sup> Similarly, 10 payments of \$10 may be more motivating than a single \$100 payment.

For P4P program design, it may be more psychologically motivating to provide a physician with smaller and more frequent incentive payments than a larger single lump-sum incentive payment. As an example, consider that a total of \$1000 is available to give in incentives to the top physician performers. Applying this principle, a physician's behavioral response is likely to be greater if the \$1000 is divided into several payments (eg, 100 payments of \$10 each) rather than paid as a single payment. Each \$10 is perceived as a new \$10 gain.<sup>14</sup>

We recognize that a reward program with frequent payments is administratively more difficult. However, the more frequent incentive can be symbolic and still be effective.<sup>15</sup> For example, every time a physician's patient receives a mammogram, an e-mail could be sent: "Your patient Edith Jones received a mammogram on this date. We will credit you with \$10 at the end of the quarter." The combination of this frequent symbolic reward and a larger separate check at the end of the quarter might be doubly satisfying because the incentive is reinforced.

***A Series of Tiered Absolute Thresholds Is Better Than 1 Absolute Threshold.*** An individual's motivation and effort when faced with a goal greatly depend on his or her baseline performance. Economists and psychologists have described this phenomenon as a "goal gradient."<sup>16</sup> If baseline performance is far away from goal performance, the individual exerts little effort because the goal is viewed as not immediately attainable. As baseline performance gets closer to goal performance, the individual exerts more effort to reach the goal (eg, 75% mammogram rate). However, the motivation to improve decreases significantly when the goal is achieved.<sup>17</sup> A simple illustration of this phenomenon is a study<sup>18</sup> of a coffee shop reward program in which the 10th coffee purchased was free. Participants in this experiment decreased the time between coffee purchases as they got closer to the free coffee.

can extend to other care settings, types of providers, and their goals.

Most evaluations of P4P programs have measured change in performance based on quality metrics.<sup>2,3</sup> However, in designing a P4P program, one has to consider the more proximal goal. To improve quality, a P4P program has to change the behavior of physicians and, more specifically, to increase the time and resources they allocate to quality improvement. The goal is to render a desired behavior (eg, taking the time to speak to patients about receiving a mammogram) front and center in the physician's mind during a busy day.

We recognize that quality improvement, cost reduction, or any other goal of a P4P program often requires more than physician behavioral change, including system changes such as implementing an electronic medical record. Keeping physicians engaged is still critical, however. When P4P programs target physician groups, it is notable that the groups often create internal P4P programs for their individual physicians.<sup>13</sup>

### Current P4P Programs

To make our recommendations more concrete, we begin by describing a prototypical physician P4P program that is designed to increase the number of women who receive a mammogram. After claims from a previous calendar year have been processed, an incentive is paid out in 2 steps. The health plan (1) determines the number of women who should have received a mammogram and how many did and (2) ascertains which physician is responsible for each patient's care and calculates a physician-specific mammogram rate.

There is significant heterogeneity in how health plans structure some aspects of their P4P programs.<sup>12</sup> Some health plans give incentives to physicians who meet a relative threshold (eg, the top 25% of physicians in terms of mammogram rate), and others use an absolute threshold (eg, physicians with a mammogram rate >75%). These top physicians can receive their incentive in various ways. They most commonly receive an increase in their reimbursement for each visit in the following year (eg, \$106 vs \$100) or a lump-sum incentive payment at the end of the year (eg, \$1000).<sup>12</sup>

We next discuss 7 potential design changes to these commonly used P4P programs (Table). We recognize that some of the design changes conflict with each other. We see them as a

■ **Table.** Seven Design Features That Could Improve Pay-for-Performance Programs

Commonly Used Design	Suggested Improvement
Incentive given as a lump sum	Divide the lump sum into a series of smaller incentive payments
Relative thresholds (eg, top 25% of physicians)	Use tiered absolute thresholds (eg, 25%, 50%, 75%, and 90%)
Long lag time between care and receipt of incentive	Shorten lag time to as short as possible
Use of withhold payments	Consider bonus payment or use of deposit contracts
Complex uncertain structure of program (eg, shared savings program)	Simplify program so that uncertainty is minimized
Incentive often given as an increase in fee schedule reimbursement	Decouple incentive payment so that it is given separately; consider a lottery
Monetary incentives	Use in-kind incentives

Goal gradient theory has several applications in a physician P4P program. In aggregate, a greater behavioral response is likely if there was a series of quality performance thresholds to meet (eg, increasing amounts of money for achieving 50%, 60%, 70%, 80%, and 90% performance thresholds) rather than 1 (eg, a 75% performance threshold). In a single-threshold system, physicians who at baseline have low performance (eg, 25%) or high performance (eg, 80%) have little reason to devote more resources to attempting to improve quality.

Some researchers have proposed eliminating thresholds entirely and using a continuous gradient (eg, a physician receives  $\$1000 \times 76\%$  performance = \$760).<sup>19</sup> Our opinion is that such a continuous gradient may be less effective than a series of thresholds because there is some benefit in having the clear bright-line goal of a threshold. However, this needs to be proven empirically.

**Reducing the Lag Times Between Care and Receipt of Incentives Increases the Behavioral Response.** Money received right away is perceived as different in value than money to be received in the future, even the near future.<sup>20</sup> This steep initial discounting is much greater than would be expected by “rational” economic discounting and has been termed *hyperbolic discounting*.<sup>20</sup> In a typical P4P program, the time required to collect and validate the data, create physician scores, and make the payout often means that the incentive payment comes many months or even 1 or 2 years after the actual delivery of care. We believe that this long delay undermines the behavioral response of physicians. Ideally, there should be little or no lag time between the behavior being rewarded and the receipt of the incentive; otherwise, the competing incentives of a busy day might trump the P4P incentive. For example, a physician might have to make the following choice: “If I spend 5 more minutes with Mrs Jones discussing the advantages of a mammogram, I might receive an incentive next March versus if I skip this discussion, I will catch up on my schedule for the day.” The immediate incen-

tive of not being behind in his or her schedule will likely trump the P4P incentive in the physician’s thinking. On the other hand, if the physician knows that the discussion might result in an immediate \$10, the cost-benefit equation might change.

**Although Withholds Have More of an Effect Than Bonuses, One Needs to Be Cognizant of the Negative Psychological Response.** Previous research has found that individuals are more sensitive to incentives when they perceive that they are losing something as opposed to gaining something.<sup>21</sup> Physicians in an experiment were asked to make a choice of treatment (surgery or radiation therapy) for a patient with cancer.<sup>22</sup> In some cases, the choice was framed as a loss (probability of dying after surgery) or as a gain (probability of surviving after surgery). Physicians were more likely to choose the surgical option when the surgical risk was framed in terms of the probability of living rather than the probability of dying. The difference in the behavioral response for a choice framed as a loss rather than as a gain can be significant, almost 2-fold in magnitude.<sup>21</sup>

This loss aversion has implications for structuring P4P incentives. Incentive payments can be structured as a withhold (a perceived loss in income) or as a bonus (a perceived gain in income). If the goal is to drive physicians to make changes that improve quality, withholding money (ie, framing the incentive as a possible loss) may lead to a greater behavioral response than framing the incentive as a “gain” in the form of a bonus, even if the same amount of money is at risk.

Although framing something as a loss rather than as a gain may result in a greater behavioral response, experiments have shown that doing so generally causes a significant negative psychological reaction and violates what the parties exposed to the incentive believe to be fair.<sup>23</sup> Therefore, while the behavioral response is stronger with a withhold, this benefit is likely outweighed by the risk of angering physicians.

A possible way to take advantage of loss aversion without the negative reaction is through the use of a deposit contract. The P4P program would have 2 options. The first option would be for the physician to receive \$1000 if his score increases by 10 percentage points from the previous year. However, the physician has a second option to enter a “deposit.” If the physician submits this \$500 deposit and his or her mammogram score increases by 10 percentage points, the physician receives \$2000 (instead of \$1000). If the mammogram score is not increased by 10 percentage points, the physician loses the \$500 deposit. Such a program has several advantages. First, it introduces loss aversion, as the physician will be motivated not to lose the \$500. Second, it takes advantage of the fact that individuals are overly optimistic in predicting their success and that physicians electively enter the program. Third, it will make it easy for the health plan to identify physicians who are engaged in the program. The downside is that the health plan’s fraction of the incentive has increased from \$1000 to \$1500.

***Reducing the Complexity of an Incentive Plan Increases the Behavioral Response.*** When given a choice of potential rewards, most people are risk averse; they will choose an option with absolute certainty over an option involving an uncertain but likely more valuable outcome. This principle of risk aversion is illustrated in a study<sup>21</sup> whereby subjects were given a choice between a 1-week vacation that was certain or a 3-week vacation that they had a 50% chance of winning. Most subjects chose the 1-week vacation. Although the 50% chance of a 3-week vacation might be considered a more rational choice in strict economic terms because the expected return of such a choice is 1.5 weeks of vacation, most people will choose the sure thing because they perceive it to be a better choice than the possibility of getting nothing at all.

A related phenomenon is that individuals often cannot process complex decisions that are tied to a financial incentive. Current P4P incentive programs are complex for a physician. It is cognitively difficult to keep track of complex trade-offs such as the following: “If I spend 5 more minutes with Mrs Jones discussing the advantages of a mammogram, I could increase my overall mammogram rate to  $x\%$ , which might put me in the 75th percentile for my peer physicians and possibly lead to an incentive at the end of the year, versus spending 5 minutes with Mrs Jones might put me behind for my morning.”<sup>24</sup> Because the P4P program decision is complex, while the concern of being late is clear and tangible, the physician is going to push off discussing the mammogram so that he or she is not late in the patient schedule.

How can P4P programs decrease uncertainty and complexity? As already noted, some health plans use relative thresholds

such as paying those physicians in the top quartile of performance as the basis for determining who “wins.” This type of payout scheme creates great uncertainty for the physician. The level of performance necessary to earn the incentive is unknown until after the fact, frequently 6 to 12 months later when physicians can be sorted by rank order of performance. A new form of incentive payment being used is a “shared savings” program. If the costs of care for a patient are less than what would be expected and quality measures are met, the health plan and the physician group share the savings.<sup>25</sup> In a shared saving program, there is uncertainty about whether there will be any cost savings and about the complexity of determining how much cost savings there will be to fund incentive payments. In contrast, absolute thresholds known in advance provide greater certainty to the physician trying to hit the target.

The least complex and most certain P4P program would likely be with the sure thing of a payment for each mammogram received. The primary care physician typically is not paid when a patient receives a mammogram, but he or she would receive an extra \$10 under such a system described herein if a patient receives a mammogram. In such an incentive system, physicians know that they will receive an incentive if they convince the patient in front of them that she should receive a mammogram.

***P4P Program and Incentive Payments Should Be “Decoupled” From Usual Reimbursement.*** As illustrated in 1 of our prototypical P4P programs, a common design feature is that the incentive payment is an incremental increase in usual reimbursement (eg, increasing the per-visit reimbursement from \$100 to \$106). We believe that this percentage increase of existing payment undermines the behavioral response of physicians. First, as already noted, an individual perceives the difference between \$0 and \$6 as being greater than the difference between \$100 and \$106. Second, in making financial decisions, individuals use mental accounting. Mental accounting describes how individuals organize, evaluate, and keep track of financial activities.<sup>26</sup> By linking the incentive payments to usual reimbursement for a visit, the incentives are mentally linked to usual reimbursement. In that context, the incentive payment is minimized because it seems miniscule compared with usual reimbursement.

If the incentive payment is decoupled from usual reimbursement, we believe that the incentive will garner more of a behavioral response. Instead of incorporating the incentive into the usual fee schedule, the health plan should make the incentive payment separate and special. Practical means of decoupling are to keep correspondence related to usual reimbursement and P4P separate and to issue incentive payments using a separate paycheck.

Another way to decouple is to use a lottery, which has been successful with patient incentive programs.<sup>27</sup> Every week, the health plan might hold a lottery for a \$10,000 payment. For each of his or her patients who received a mammogram in the previous week, a physician gets a virtual “lottery ticket,” and the odds of winning are a function of how many tickets he or she has. Every week, an e-mail or letter is sent to all physicians about who won the lottery and about how many chances to win they had “earned.” Beyond decoupling the payment, the perceived value of the incentives is higher (\$10,000 is likely a significant amount of money for any physician), although in aggregate the health plan is paying the same amount per week. This magnifies the incentive for all participants. Perhaps most important, we believe that under such a system the physician will perceive the incentive to be a pleasant surprise.

**“In Kind” Rewards May Be a Stronger Driver of Change Than a Cash Reward of the Same Amount.** Monetary incentives might be less effective in driving behavioral change than an object or service of equal value. This is illustrated in how the National Football League creates an incentive for its top players to play in the Pro Bowl. In the past, when offered a financial incentive to play in the game, many players declined. For players with 7-figure salaries, an incentive of several thousand US dollars was insufficient to play an extra game.<sup>26</sup> However, when the National Football League moved the game to Hawaii and provided 2 first-class tickets (for girlfriend or spouse) and accommodations for the players, this in-kind incentive became more effective.

In the same manner, an incentive of an all expense–paid dinner at a fancy restaurant (worth \$250) would be more valuable to a physician than \$250 in cash (presuming that the physician enjoys fine dining). Because the physician sees spending \$250 at the restaurant as a splurge, it makes the dinner that much more valuable. If fine dining appears unseemly, other options include a bagel breakfast for the practice or the latest and most expensive stethoscope. It could even be a choice of several options such as those used by credit card reward programs. Ideally, the object should be something that the physician would not normally buy for himself or herself.

### Potential Ways to Mitigate Unintended Consequences of P4P

We believe that the design changes described herein can be applied to a P4P program to maximize the response of physicians to an incentive. However, we acknowledge that a major drawback of using financial incentives is the potential for unintended and negative consequences. For example, in a recent evaluation of a large P4P program in the United King-

dom, researchers found that there was a decline in performance on measures excluded from the P4P program.<sup>17</sup> After a period of rapid improvement, there was also concern that the improvement had slowed, as physicians had achieved most of their potential incentives and saw little reason to focus their energies on further improvement.

**“Teaching to the Test.”** Multidimensional output, or multitasking, refers to situations in which the responsibilities of an individual include multiple activities or outputs that may require different types of skills to accomplish.<sup>28</sup> A physician’s “output” includes many different components such as managing a patient’s chronic illness, timely and efficiently diagnosing a patient’s new symptom, counseling and advising on how to prevent illness, and providing emotional support.

Multitasking is relevant to P4P programs because the performance measures in these programs typically address only a narrow portion of a physician’s outputs or the processes that contribute to outputs. For example, a P4P program may reward a physician for patient receipt of a mammogram but not other processes or outputs that are difficult to measure, such as diagnostic acumen for a patient presenting with unclear symptoms. If a large incentive is applied to a single type of output, other outputs may be neglected, and overall care might worsen.<sup>28</sup> Therefore, a large financial incentive based on a narrowly focused set of measures may lead to the unintended consequence of having a physician teach to the test, devoting resources to those items being measured and neglecting other important outputs that are not being measured. Teaching to the test is why few private-sector corporations place risk incentives on a large fraction of employee income.<sup>29</sup> There is mixed evidence about whether current P4P programs in healthcare have actually led to the adverse consequence of teaching to the test.<sup>17,30</sup>

A classic method of minimizing the likelihood of teaching to the test is to create an incentive program that addresses an extensive array of a physicians’ output by applying a broad dashboard of performance measures. This approach has been adopted by a primary care physician P4P incentive program in the United Kingdom that has more than 146 quality indicators.<sup>7</sup> The challenge with this approach is to avoid creating a program that may be overly complicated and expensive. Collecting and auditing quality data are inherently expensive and may outweigh benefits of the P4P program.

**Intrinsic Versus Extrinsic Motivation.** Meta-analyses<sup>31-33</sup> of studies that examined incentive programs in non-healthcare settings show that, while some programs have a positive effect, other programs have a negative effect. One theory to explain these mixed findings is that incentive might cause a conflict between intrinsic motivation, which is a person’s inherent desire to do a task, and extrinsic mo-



tivation, which is the external incentive such as might be provided in a P4P program. Researchers theorize that, instead of supporting intrinsic motivation, an extrinsic incentive “crowds out” intrinsic motivation.<sup>31,32,34</sup> Another explanation for this crowding-out effect is that, when a task is tied to an extrinsic incentive, people infer that the task is difficult or unpleasant.<sup>35</sup> Similar concerns have been raised about the effect of P4P in healthcare and how it may violate a physician’s sense of professionalism.<sup>8</sup> An alternative possibility is that a person usually concentrates on only the primary reason for a task rather than the sum of all possible reasons. This theory is used to explain why financial incentives for blood donation are ineffective: the financial incentive is less than the altruistic benefit of blood donation.<sup>36</sup>

The intrinsic motivation theory implies that a small P4P incentive could have no effect or could lead to lower performance if it is tied to something that physicians are intrinsically motivated to improve, such as quality of care. A potential way to address the crowding out of intrinsic motivation is simply to increase the size of the financial incentive. A large external incentive will crowd out any inherent intrinsic motivation; however, it may in turn create a greater behavioral response than would be obtained through intrinsic motivation alone. A study entitled “Pay Enough, or Don’t Pay at All”<sup>34</sup> illustrated this concept in an evaluation of IQ tests. Each of 4 groups was given a different incentive for each correct answer (no financial incentive or a small, medium, or large financial incentive). The group given no financial incentive outperformed the group given the small financial incentive (56% vs 46%), and the groups given the medium and large financial incentives (68% for both) outperformed both of the other groups.

## Conclusions

Taken together, the theories that we reviewed suggest that the way in which P4P incentives are structured or framed can influence whether they achieve the desired behavioral response. For a given amount of money, we suggest that the greatest behavioral response will occur with more frequent and smaller payments. We believe that establishing several stepped absolute thresholds and decoupling incentive payments from usual reimbursement may be more effective than current P4P designs. Lotteries and nonmonetary incentives are presented as other mechanisms to increase the behavioral response of physicians.

The potential unintended negative consequences discussed herein serve as a helpful counterpoint to our recommendations. They emphasize that P4P incentives could lead to the neglect of other important but unmeasured outputs of physicians and may have a negative effect on quality. Therefore, P4P programs should closely monitor for these unintended

consequences, and we have suggested some potential mechanisms to mitigate these risks. If unintended consequences arise, they will have to be considered in the cost-benefit analysis of a P4P program, just as a physician considers potential adverse effects before prescribing a medication.

There are several important limitations and caveats to our recommendations. The theories and studies we cite herein were used to describe the behavior of individuals and not institutions. Therefore, it is unclear to what extent the design changes we describe are applicable to hospitals and physician groups. Physician groups are more likely to have the resources to implement changes to improve their performance on a given quality measure that does not require increased effort by individual physicians (eg, having a nurse call women who have missed their mammogram). Although these efforts are important, it is important to engage individual physicians in quality improvement. Even within physician groups such as the Palo Alto, California, clinic with numerous quality improvement initiatives, physician-specific P4P programs have led to an incremental improvement in quality.<sup>13</sup>

Another caveat is that there are often practical reasons for not choosing the options suggested by these theories. For example, it was noted herein that more frequent payout might lead to a greater behavioral response. However, this result might be outweighed by the higher administrative costs to the health plan of more frequent processing of data and payouts. We have highlighted possible work-arounds to minimize these administrative costs.

Also relevant is the issue of financial risk. An absolute threshold with an associated incentive having a fixed US dollar amount might have advantages in terms of a behavioral response. However, such an approach leads to greater risk for the payer, who could face the prospect of paying out much more in incentives than was budgeted. At a P4P program in the United Kingdom, provider performance greatly exceeded what was expected, so the cost to taxpayers was considerably more than expected.<sup>7</sup> Although the design of incentive payments is important, we recognize that there are other aspects of P4P that affect how providers will respond. These include other P4P design elements (eg, measures used) and the provider’s practice environment (eg, availability of electronic medical records). The variation across health plans in P4P program design also makes it more difficult for providers as they face multiple and sometimes conflicting incentives.

Last and most important, we believe that the lessons from behavioral economics could greatly enhance the design and effectiveness of P4P programs in healthcare, but future work is needed to demonstrate this empirically. It will be critical to test the P4P program design enhancements that we propose to

determine if they are effective and whether they cause unintended consequences.

**Acknowledgments**

We acknowledge valuable input and suggestions from George Loewenstein.

**Author Affiliations:** From RAND Health (AM, MESS), Pittsburgh, PA; the School of Medicine (AM), University of Pittsburgh, Pittsburgh, PA; and RAND Health (CLD), Santa Monica, CA.

**Funding Source:** This study was supported by the Office of the Assistant Secretary of Planning and Evaluation, Department of Health and Human Services, and by the Centers for Medicare & Medicaid Services. The work was supported in part by a career development award from the National Institutes of Health. The funders had no role in the composition of the manuscript or in the decision to publish.

**Author Disclosures:** The authors (AM, MESS, CLD) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

**Authorship Information:** Concept and design (AM, MESS, CLD); acquisition of data (AM); analysis and interpretation of data (AM, MESS, CLD); drafting of the manuscript (AM, CLD); critical revision of the manuscript for important intellectual content (MESS); obtaining funding (MESS); and supervision (CLD).

**Address correspondence to:** Ateev Mehrotra, MD, RAND Health, 4570 Fifth Ave, Ste 600, Pittsburgh, PA 15213-2665. E-mail: mehrotra@rand.org.

REFERENCES

1. **Norton EC.** Incentive regulation of nursing homes. *J Health Econ.* 1992;11(2):105-128.

2. **Mehrotra A, Damberg CL, Sorbero ME, Teleki SS.** Pay for performance in the hospital setting: what is the state of the evidence? *Am J Med Qual.* 2009;24(1):19-28.

3. **Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S.** Does pay-for-performance improve the quality of health care? *Ann Intern Med.* 2006;145(4):265-272.

4. **Lindenauer PK, Remus D, Roman S, et al.** Public reporting and pay for performance in hospital quality improvement. *N Engl J Med.* 2007;356(5):486-496.

5. **Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM.** Pay for performance in commercial HMOs. *N Engl J Med.* 2006;355(18):1895-1902.

6. **US Department of Health and Human Services.** *HHS Reports to Congress on Value-Based Purchasing of Hospital Services by Medicare.* Baltimore, MD: Centers for Medicare & Medicaid Services. November 26, 2007.

7. **Doran T, Fullwood C, Gravelle H, et al.** Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med.* 2006;355(4):375-384.

8. **Berwick DM.** The toxicity of pay for performance. *Qual Manag Health Care.* 1995;4(1):27-33.

9. **Dudley RA, Frolich A, Robinowitz DL, Talavera JA, Broadhead P, Luft HS.** *Strategies to Support Quality-Based Purchasing: A Review of the Evidence.* Rockville, MD: Agency for Healthcare Research and Quality; 2004. Technical review.

10. **Frolich A, Talavera JA, Broadhead P, Dudley RA.** A behavioral model of clinician responses to incentives to improve quality. *Health Policy.* 2007;80(1):179-193.

11. **Rosenthal MB, Dudley RA.** Pay-for-performance: will the latest payment trend improve care? *JAMA.* 2007;297(7):740-744.

12. **Sorbero ME, Damberg CL, Teleki S, et al.** *Assessment of Pay-for-Performance Options for Medicare Physician Services: Final Report.* Washington, DC: US Dept of Health and Human Services; 2006. RAND

working paper; prepared for the Assistant Secretary for Planning and Evaluation.

13. **Chung S, Palaniappan LP, Trujillo LM, Rubin HR, Luft HS.** Effect of physician-specific pay-for-performance incentives in a large group practice. *Am J Manag Care.* 2010;16(2):e35-e42.

14. **Thaler RH.** Mental accounting and consumer choice. *Marketing Sci.* 1985;4(3):199-214.

15. **Diamond F.** Palo Alto clinic physicians design own P4P program. *Manag Care.* 2009;18(7):54-56.

16. **Heath C, Larrick RP, Wu G.** Goals as reference points. *Cognitive Psychol.* 1999;38:79-109.

17. **Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M.** Effects of pay for performance on the quality of primary care in England. *N Engl J Med.* 2009;361(4):368-378.

18. **Kivetz R, Urminsky O, Zheng Y.** The goal-gradient hypothesis resurrected: purchase acceleration, illusory goal progress, and customer retention. *J Marketing Res.* 2006;43(1):39-58.

19. **US Department of Health and Human Services.** *Report to Congress: Plan to Implement a Medicare Hospital Value-Based Purchasing Program.* 2007. [http://healthcaredisclosure.org/docs/files/CMS\\_Report1107.pdf](http://healthcaredisclosure.org/docs/files/CMS_Report1107.pdf).

20. **Loewenstein G, Prelec D.** Anomalies in intertemporal choice: evidence and an interpretation. *Q J Econ.* 1992;107:573-597.

21. **Kahneman D, Tversky A.** Prospect theory: an analysis of decision under risk. *Econometrica.* 1979;47(2):263-292.

22. **McNeil BJ, Pauker SG, Sox HC Jr, Tversky A.** On the elicitation of preferences for alternative therapies. *N Engl J Med.* 1982;306(21):1259-1262.

23. **Kahneman D, Knetsch JL, Thaler RH.** Fairness as a constraint in profit-seeking: entitlements in the market. *Am Econ Rev.* 1986;76(4):728-741.

24. **Prelec D, Loewenstein GF.** Decision making over time and under uncertainty: a common approach. *Manag Sci.* 1991;37(7):770-786.

25. **Trisolini M, Aggarwal J, Leung M, Pope G, Kautter J; RTI International.** *The Medicare Physician Group Practice Demonstration: Lessons Learned on Improving Quality and Efficiency in Health Care.* February 2008. [http://www.cms.hhs.gov/DemoProjectsEvalRpts/downloads/PGP\\_SiteMeeting\\_Report.pdf](http://www.cms.hhs.gov/DemoProjectsEvalRpts/downloads/PGP_SiteMeeting_Report.pdf). Accessed May 13, 2009.

26. **Thaler RH.** Mental accounting matters. *J Behav Decis Making.* 1999;12:183-206.

27. **Volpp KG, John LK, Troxel AB, Norton L, Fassbender J, Loewenstein G.** Financial incentive-based approaches for weight loss: a randomized trial. *JAMA.* 2008;300(22):2631-2637.

28. **Holmstrom B, Milgrom P.** Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *J Law Econ Organ.* 1991;7:24-52.

29. **Asch B, Warner J.** Incentive systems: theory and evidence. In: Lewin D, Mitchell D, Zaidi M, eds. *Handbook of Human Resource Management.* Stamford, CT: JAI Press; 1996:175-215.

30. **Glickman SW, Ou FS, DeLong ER, et al.** Pay for performance, quality of care, and outcomes in acute myocardial infarction. *JAMA.* 2007;297(21):2373-2380.

31. **Cameron J, Banko KM, Pierce WD.** Pervasive negative effects of rewards on intrinsic motivation: the myth continues. *Behav Analyst.* 2001;24(1):1-44.

32. **Deci EL, Koestner R, Ryan RM.** A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol Bull.* 1999;125(6):627-668, 692-700.

33. **Rothe HF.** Output rates among welders: productivity and consistency following removal of a financial incentive system. *J Appl Psychol.* 1970;54:549-551.

34. **Gneezy U, Rustichini A.** Pay enough, or don't pay at all. *Q J Econ.* 2000;115(3):791-810.

35. **Freedman JL, Cunningham JA, Krismer K.** Inferred values and the reverse-incentive effect in induced compliance. *J Pers Soc Psychol.* 1992;62(3):357-368.

36. **Mellström C, Johannesson M.** Crowding out in blood donation: was Titmuss right? *J Eur Econ Assoc.* 2008;6(4):845-863. ■