



RESEARCH REPORT

Payment Methods and Benefit Designs: How They Work and How They Work Together to Improve Health Care

# Pay-for-Performance

*Robert A. Berenson*  
URBAN INSTITUTE

*Divvy K. Upadhyay*  
URBAN INSTITUTE

*Suzanne F. Delbanco*  
CATALYST FOR  
PAYMENT REFORM

*Roslyn Murray*  
CATALYST FOR  
PAYMENT REFORM

April 2016



CATALYST  
FOR  
PAYMENT  
REFORM



## ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is dedicated to elevating the debate on social and economic policy. For nearly five decades, Urban scholars have conducted research and offered evidence-based solutions that improve lives and strengthen communities across a rapidly urbanizing world. Their objective research helps expand opportunities for all, reduce hardship among the most vulnerable, and strengthen the effectiveness of the public sector.

# Contents

<b>Pay-for-Performance</b>	<b>1</b>
Background	1
Key Objectives	2
Strengths	3
Weaknesses	3
Design Choices to Mitigate Weaknesses	5
Compatibility with Other Payment Methods and Benefit Designs	6
Focus of Performance Measurement	7
Potential Impact on Provider Prices and Price Increases	7
<b>Acknowledgments</b>	<b>8</b>
<b>Statement of Independence</b>	<b>9</b>



---

Payment reform promises to substitute value for volume. Yet, value- and volume-based approaches typically are implemented together. All payment methods have strengths and weaknesses, and how they affect the behavior of health care providers depends on their operational design features and, crucially, on how they interact with benefit design. Those seeking greater value for their health care dollar are also turning to innovation in benefit design, which also typically involves the implementation of more than one approach at a time—each with its own strengths, weaknesses, and effect on consumer health care behavior. Although payment and benefit design each has received significant attention independently, the intersection between the two has received little if any. The Urban Institute partnered with Catalyst for Payment Reform to explore how established and proposed payment methods and benefit design options work on their own and together. We also examined how payment and benefit design can be blended to improve health care delivery. This chapter is one of the nine payment methods discussed in the report *Payment Methods: How They Work*. All reports and chapters can be found on our project page: [Payment Methods and Benefit Designs: How They Work and How They Work Together to Improve Health Care](#).

---

# Pay-for-Performance

## Background

A pay-for-performance (P4P) model consists of financial incentives or penalties based on a provider's ability or inability to meet certain performance expectations based on predetermined measures. Mostly, P4P has been associated with providers' performance in meeting quality of care standards or improving the quality provided to the patients for whom they are responsible. P4P models measure performance using clinical process and outcome measures and surveys on patients' experiences with care. MIPS, a prominent example of P4P, also includes measures of "meaningful use" of electronic health records and resource use. Often implemented as a performance-based bonus on top of usual compensation methods, P4P as adopted in Medicare also includes penalties with considerable financial impact. The measures used in P4P programs can be targeted to an individual physician, a group of physicians, or an organization, such as a hospital or a large integrated delivery system.

Typically, performance bonuses or penalties have represented a few percentage points of the base payment providers or health professionals would have received. However, because the economics of different providers varies substantially, a 1 or 2 percentage point bonus or penalty has much different impact. For example, hospital margins are usually in the low single digits, whereas “profits” for a physician practice, representing the physicians’ take-home income, are typically greater than 40 percent. The same P4P percentage of incremental payment, thus, can impact hospital behavior much more than physician behavior. Recently, perhaps recognizing that the current P4P programs for physicians in Medicare have had limited impact, Congress substantially increased the P4P amounts to approach 10 percent downside penalties and even greater upside bonuses. A P4P approach in the United Kingdom, labeled the Quality and Outcomes Framework, provided bonuses of more than 25 percent; yet the approach has elicited mixed reviews about whether the relatively modest quality improvements have been worth the financial investment.

Various formulations of P4P programs differ based on whether providers *attain* a certain level of performance or *improve* from a baseline performance enough to qualify for bonuses. Accordingly, there are various pros and cons regarding whether attainment or improvement might apply in particular situations—and some P4P programs use a combination of attainment and improvement metrics.

Despite their widespread adoption, especially in Medicare, evidence is limited on the success of P4P programs for physicians, hospitals, and other providers in improving the specific elements of care, and so far, it is not persuasively positive. Further, there has been little assessment of P4P’s effect on non-measured quality, supporting concerns that providers’ attention has been diverted to what is being measured and rewarded, and away from the intrinsic commitment to quality that professionals have.

## Key Objectives

Pay-for-performance is predicated on the concept that providing financial rewards and penalties will motivate providers to pay attention to quality. Thus, providers will attempt to improve care they provided commensurately with their greater attention quality. In addition, when combined with public reporting of quality and perhaps other aspects of performance, P4P can provide some accountability for the substantial funding providers receive. This oversight also can serve to guide consumers and patients in their choices of providers.

## Strengths

- Most payment methods' primary impact is on volume of services produced. P4P introduces into payment policy emphasis on the quality of care produced, a core element of care that has been missing in base payment methods.
- P4P permits payers and purchasers to emphasize which aspects of performance deserve priority—for example, aspects of care that might be compromised under the incentives of particular base payment models.
- P4P is complementary to public reporting of performance; together, public reporting and pay-for-performance impart transparency to better hold providers accountable for the large payments they receive, to help consumers and patients make informed choices of provider, and to support quality improvement efforts.
- P4P can be complement base payment methods without changing their basic structures. It offers payers the most practical approach to improving value with providers who are unwilling or unable to accept new forms of base payment.
- As an incremental payment method, P4P can be implemented with varying degrees of intensity, consistent with the context of application, the strength of the measures available for the clinical conditions to which it is being applied, or other relevant factors.
- Although there are clear gaps in what is accurately measureable, a commitment to P4P could create momentum to expand measure sets and approaches to achieving greater measurement accuracy.

## Weaknesses

- P4P introduces significant administrative complexity associated with acquiring data and verifying it for accuracy.
- Behavioral economics suggests that, in professions that require high cognitive skill and high intrinsic motivation, associating better performance with financial incentives could be counterproductive because it might compromise commitment to quality. P4P incentives for organizations such as hospitals may or may not impart this “crowd out” of intrinsic motivation.

- For many health professionals and providers, there are major gaps in what aspects of care are measurable using current data sources. Therefore, erroneous judgments about a provider's overall quality and value may occur.
- Most P4P programs tend to concentrate on clinical process measures rather than outcomes, which are what consumers or payers are most interested in achieving. Moreover, research indicates that process measures are not strongly related to significant healthcare outcomes.
- Apparent improvement in performance may simply reflect more extensive documentation of what was already being done for patients, thus reflecting improvement in reporting rather than actual performance.
- It is administratively easy and practical for payers to base measurement on objective administrative data, usually from claims such as laboratory test results, but that limits the choice of measures. Clinical data from records, self-reporting by providers, or patient-reported outcomes is more costly to obtain and not necessarily reliable (although this may change with improvements in electronic medical records).
- The small incremental reward and penalty payments common in P4P programs may not be sufficient to counter the much stronger incentives in the base payment methods that produce a larger share of provider payment.
- In a multipayer health care system, different P4P regimes may cause providers dissonance in responding to different measures, different measurement requirements, and different approaches to rewards and penalties.
- Attainment approaches in public reporting and P4P that compare providers' performance may not be fair to providers with more challenging patient populations, perhaps because of socioeconomic factors or unmeasured case-mix differences. Yet, improvement approaches compromise the goals of providing consumers with information for choice and for accountability—as well as making it easier for initially poor performers to receive rewards for their improvement.
- Clinicians might respond to P4P incentives by altering their professional activities so they perform better on the P4P measures, as they sometimes do in public reporting programs—for example, by not caring for high-acuity patients with greater likelihood of experiencing a poor outcome.

## Design Choices to Mitigate Weaknesses

Pay-for-performance continues to evolve, with no consensus about how best to design it. Indeed, there remain different viewpoints on whether P4P is even a good idea, especially when applied to individual clinicians. The major reason is P4P's potential to crowd out intrinsic motivation, even as some evidence suggests that P4P programs work best when they use highly targeted measures and reward individual clinicians directly. There remain tensions between adopting particular measures for specific clinical application versus having uniform, "core" measures across payers and providers. Further, there are different views about the relative merits of attainment models versus improvement models of P4P.

Some, including Congress (as reflected in the Medicare Access and CHIP Reauthorization Act of 2015), think that measurement gaps can be readily filled, and that this would be accelerated by broad adoption of P4P programs that would create greater urgency for filling measurement gaps. Others think that some measurement gaps cannot be filled in the foreseeable future, raising concerns that the rapid, broad adoption of P4P programs might perversely have an adverse impact on quality.

Process measures are useful because they are more readily "actionable" to improve performance (that is, the performance deficiency is readily evident in the measure itself). Yet, growing consensus holds that the measures used in P4P should over time de-emphasize processes and move to outcomes, although accurately measuring providers' performance on outcomes is more difficult. Outcome assessments could not only include clinical outcomes—for example, 30-day mortality after a procedure—but also consider patients' experiences and patient-reported outcomes from surveys as essential components.

There is concern that a clinician or institutional provider may face different—even conflicting—P4P programs that use different measures, different scoring, and so on. To address this, payers and providers might collaborate to develop common approaches, as the California P4P program accomplishes under the auspices of the Integrated Healthcare Association.

P4P program implementers should be clear about the purpose of their P4P program, because that will determine, among other things, how large the upside (and any downside) P4P payments should be. For example, a central decision might be whether the program's main purpose is to influence patients' choice of provider and achieve broad provider accountability or, whether the program should promote specific quality improvements for high-impact problems. The decision on purpose should help determine which measures to use in P4P and how much money to put on the table for the program—as well as make clear to providers what results are desired.

Finally, payers need to appreciate that they might more efficiently achieve the objective of increasing quality at a reasonable cost through base payments rather than through incremental P4P incentives. For example, to reduce preventable hospital readmissions, extending the episode of DRG case rate beyond discharge might be a more powerful and effective approach than measuring readmission rates and applying incremental payment penalties.

## Compatibility with Other Payment Methods and Benefit Designs

As an incremental payment based on measured performance, P4P is compatible with all base payment methods. P4P is also compatible with shared savings incremental rewards and penalties, with the former typically focused more on quality and the latter on costs. However, measurement of costs and cost increases could be included as part of P4P without a separate shared savings program. A prime example is the Medicare MIPS program, under which one of the four components of a physician's value score will be their resource use performance as measured by spending for their patients. More typically, P4P would include utilization data—for example, readmission rates—whereas shared savings by definition would focus on expenditures.

Public reporting and P4P can be crafted to target the aspects of quality at risk of adverse outcomes under a particular payment method. For example, there is potential for stinting under global capitation and strong two-sided-risk shared savings programs. Thus, P4P could be added for services subject to stinting. This would, when combined with public reporting on the same measures, serve as a counterincentive. Generally, it is difficult to measure underservice (except for primary and secondary prevention services) using claims data as the basis for determination. Unfortunately, some areas of stinting (e.g., not referring when it is clinically indicated) are not currently amenable to measurement. In particular, it would be desirable to have measures of appropriate referrals in relation to narrow network plans.

Finally, P4P programs are a provider-side strategy and should not conflict with consumer-oriented strategies that use cost-sharing to alter demand for services. Because many quality measures used in P4P programs are for primary prevention services, high-deductible plans that provide first-dollar coverage for prevention services may be compatible. However, secondary prevention services, which also are emphasized in P4P, are not exempt from high deductibles. This may create financial barriers to care that impede providers' ability to perform well on the measures. And the impact of the barriers may

be a function of patients' socioeconomic situations, thereby creating differential impact on performance across providers.

## Focus of Performance Measurement

Controversy persists over the breadth and focus of the measures that might populate P4P programs. In general, the measures should be the same or comparable to those used for public reporting. A key issue is whether the measures should be general core measures most consistent with assessing the Triple Aim goals for population health or whether the measures should be granular and specific to particular specialties and institutional providers. The former approach generally requires a focus on population health measures and would be less useful for assessing individual provider performance. The latter approach would assess the individual provider but faces particular challenges because measure gaps persist for many specialties and other providers and because smaller sample sizes impede statistical significance. Currently, there does not seem to be consensus on how best to proceed with P4P.

## Potential Impact on Provider Prices and Price Increases

As an incremental payment approach, P4P does not directly address providers' prices or incentives for price increases. Relative market power as expressed through negotiations over terms and conditions, however, can affect the specific P4P design implemented. Powerful providers can simply refuse to participate or can participate in ways payers would not prefer, for example, by compelling payers to provide upside-only rewards rather than rewards and offsetting penalties.

# Acknowledgments

This report was funded by the Robert Wood Johnson Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at [www.urban.org/support](http://www.urban.org/support).

A [technical expert panel](#) advised the project team and reviewed the reports at different stages.

## STATEMENT OF INDEPENDENCE

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.



2100 M Street NW  
Washington, DC 20037

[www.urban.org](http://www.urban.org)