CATO INSTITUTE

WORKING PAPER

Please cite to 7 Yale J. Health P. Law & Ethics (forthcoming 2007)

July 31, 2006

Pay-for-Performance
Is Medicare a Good Candidate?

by Michael F. Cannon

Executive Summary

In response to growing concern over the quality of medical care, private and public third-party payers are experimenting with financial incentives, known as "pay-for-performance" (P4P), that reward health care providers for recommended care. Although P4P has the potential to improve quality in some instances, policymakers should take a cautious approach to this new tool.

Creating and administering provider-focused P4P financial incentives are immensely complex tasks that require making tradeoffs amid considerable uncertainty. Provider-focused P4P incentives often improve quality of care for some patients at the same time they reduce quality of or access to care for others. In particular, provider-focused P4P incentives can encourage inappropriate care or reduce access to care for patients with multiple illnesses or low incomes.

Medicare, the federal health care program for the elderly and disabled, has begun experimenting with provider-focused P4P incentives. Yet Medicare faces additional challenges beyond those confronting private third-party purchasers. Given Medicare's patient population, size, and sensitivity to interest group lobbying, any harm that could result from a P4P scheme would be more likely to occur within traditional Medicare than elsewhere in the health care system.

Congress can realize the potential of provider-focused P4P incentives, while reducing the likelihood of harm, by confining provider-focused P4P to private Medicare Advantage plans and by encouraging greater participation in those plans. Further, P4P financial incentives can be targeted at patients as well as providers. Patient-focused financial incentives would offer greater transparency and allow patients and their doctors to deviate from treatment guidelines when doing so is in the patient's interest.

Michael F. Cannon is director of health policy studies at the Cato Institute (www.cato.org) and coauthor of Healthy Competition: What's Holding Back Health Care and How to Free It *(Cato Institute, 2005).*

Medicare is an example of a quality-blind third-party purchaser.

Introduction

According to one prominent study, adults in the United States receive the generally accepted standard of preventive, acute, and chronic care only about 55 percent of the time. The likelihood that patients would receive recommended care "varied substantially according to the particular medical condition, ranging from 78.7 percent of recommended care . . . for senile cataract to 10.5 percent of recommended care . . . for alcohol dependence."[1] Evidence of low-quality care appears in Medicare, the federal health program for the elderly and disabled. Another study documents similar levels of recommended acute and preventive care and finds that patients are often less likely to receive such recommended care in regions where Medicare expenditures are highest. Patients in high-spending regions received 60 percent more care, but those higher Medicare expenditures did not translate into higher-quality care, decreased mortality, better functional status, or higher patient satisfaction.[2]

Third-party payment is a potential contributor to the underprovision of quality health care. Most health care payments in the United States are made by third parties to the delivery of care, such as employers, insurers, or government. Those purchasers typically reimburse health care providers on the basis of the volume and intensity of the services provided, rather than the quality or cost-effectiveness of those services.[3] The result is a financing system akin to paying academics on the basis of the volume and intensity of endnotes.[4]

Medicare is an example of a quality-blind third-party purchaser. Former Medicare administrator Tom Scully notes that, within a hospital referral region, Medicare pays "the exact same amount for hip replacement and the same amount for a heart bypass, if you're the best hospital or the worst hospital."[5] The Medicare Payment Advisory Commission has written:

> In the Medicare program, the payment system is largely neutral or negative towards quality. All providers meeting basic requirements are paid the same regardless of the quality of service provided. At times providers are paid even more when quality is worse, such as when complications occur as the result of error.[6]

Medicare's quality-blind payment system results not just in the underprovision of high-quality care but also in the overprovision of low-quality care.[7] For those and other reasons, Medicare has been the major focus of efforts to solve the third-party payer quality problem.

Those efforts have led third-party payers to experiment with financial incentives that encourage physicians and hospitals to provide recommended care. Such initiatives are termed "quality-based purchasing" or "pay-for-performance" (P4P). P4P is an outgrowth of the "evidence-based medicine" (EBM) movement. EBM advocates argue that providers too often rely on their own judgment. That is because scientific evidence on the effectiveness of medical interventions is too often unavailable and too often ignored when it is available. P4P attempts to use financial incentives to encourage providers to adhere more closely to evidence-based standards of care. As described by one academic proponent:

> The key to the quality-based payment system is that it differentiates between the intensity of medical care and the value of it. . . . Health-based payments . . . reward high-value services regardless of their intensity. Thus, there are no incentives to overprovide or underprovide services.[8]

By tying financial incentives to superior modes of care, advocates of third-party P4P hope to harness providers' self-interest in the service of higher-quality care.

A number of P4P initiatives are already under way in both the public and private sectors. Commercial insurers have been leaders in the field; some experimented with P4P as early

as 1994. Those private-sector programs reward physicians and facilities for meeting performance goals including patient satisfaction, preventive care, chronic care, acute care, and smoking cessation.[9] Medicare currently has 10 demonstration programs under way, which tie higher reimbursements to data reporting and a variety of quality indices (including structural, process, and outcome measures) across various types of care (though typically for chronic illnesses) and care settings. Such performance measures will be discussed in the next section.

Medicare recently released the first quality-based bonus payments in the program's history, following the promising results it announced from one such P4P demonstration. The Premier Hospital Quality Incentive demonstration was launched in 2003. It collects data on 33 quality indicators for joint replacements, coronary artery bypass grafts, heart attacks, heart failure, and pneumonia. For each clinical area, hospitals that score in the first and second deciles receive bonus payments from Medicare of 2 percent and 1 percent of Medicare payments for those services, respectively. After the first year, the demonstration used the bottom two deciles in each area of care to set baselines for poor performers. In the third and subsequent years, Medicare will reduce payments to hospitals that score below those baselines by 1–2 percentage points. Medicare predicts that most hospitals will improve and that "few, if any, hospitals would get a payment reduction."[10] Medicare officials estimate that the demonstration program, by encouraging the use of more effective care, has thus far saved the lives of 235 heart attack patients.[11]

Congress is considering proposals to expand on those initiatives within Medicare. Rep. Nancy Johnson (R-CT) has introduced legislation that would give larger payment increases to physicians who meet administratively specified performance targets or who make significant progress toward meeting them. Sen. Chuck Grassley (R-IA) has introduced even more expansive legislation, which would create P4P incentives for hospitals, physicians, Medicare Advantage plans, home health agencies, and other providers. Language that would have broadened the use of P4P in Medicare was removed from the fiscal year 2006 budget reconciliation package just before final passage.

Provider-focused financial incentives for high-quality care have the potential to improve quality in many instances. However, caution is in order. As discussed in the next section, creating a P4P program is an immensely complex task. That complexity derives from the difficulty that bureaucracies face in defining "quality" for large and diverse populations. As a result, creating provider-focused P4P incentives can improve the quality of care for some patients at the same time it reduces quality of or access to care for others.

Those difficulties suggest two approaches that would maximize the potential of P4P while minimizing any resulting harm. First, private experiments with provider-focused P4P incentives are preferable to public experiments. The current system of private P4P programs allows insurers and employers to conduct experiments and learn from each other's successes. As important, it confines any harmful failures to smaller populations. As discussed below, the politics of Medicare all but guarantees that any harm that might result from a P4P scheme is more likely to occur in Medicare, would harm more patients, and would take longer to correct. Congress should confine provider-focused P4P incentives to the Medicare Advantage program, under which beneficiaries can choose a private plan that provides Medicare-covered services. That would preserve the experimental process that has unfolded to date. Congress should resist the temptation to expand P4P in traditional Medicare.

Second, employers and insurers should experiment not only with provider-focused financial incentives but with patient-focused financial incentives as well. A weakness of provider-focused financial incentives is that they can affect the quality of care, or even a patient's access to care, without the patient's knowledge. In contrast, patient-focused financial incentives would engage patients in the pursuit of quality, while allowing them to

Private experiments with provider-focused P4P incentives are preferable to public experiments, and employers and insurers should experiment with patient-focused financial incentives as well.

deviate from "best practices" if doing so fits their needs.

Pitfalls of Third-Party P4P

Identifying and rewarding quality are difficult tasks for any third-party purchaser. As one study of P4P measures notes, "Experience in other industries has shown that developing performance measures for complex phenomena is difficult and that inappropriate measures can have significant negative consequences."[12] Defining quality health care is not as straightforward as it might appear. Quality is a complex and often subjective concept. Even relatively objective measures of quality can be difficult to translate into financial incentives that succeed in improving the quality of care. This section outlines the challenges faced by third-party purchasers, whether public or private, when attempting to improve quality through provider-focused financial incentives.

Defining Quality

The first challenge is to identify the dimensions of quality to be promoted. P4P programs typically rely on some mix of four types of quality measures: clinical outcomes, processes, structural factors, and patient satisfaction. Each dimension presents strengths and weaknesses as a measure of health care quality. Combining multiple dimensions can capture the strengths of each, but at the cost of added complexity.

Clinical outcomes, hereafter called patient outcomes, are the most obvious measure of health care quality. For example, outcome measures for heart attack patients could include patients' postintervention cholesterol levels, readmission rates, or mortality rates.

Defining quality health care is not as straightforward as it might appear.

However, outcome measures have limitations. First, patients may differ in their desired outcomes.[13] Second, patient outcomes can be influenced by factors other than the medical intervention. Readmission and mortality rates for heart attack patients may be influenced by the severity of illness. Patients' cholesterol levels may be influenced by their adherence to a prescribed drug regimen (e.g., statins). Such factors contribute to patient outcomes but say little about the quality of care provided. As a result, providers are understandably reluctant to be judged on the basis of factors they cannot control. A third and related limitation of outcome measures is that "although outcomes might indicate good or bad care in the aggregate, they do not give an insight into the nature and location of the deficiencies or strengths to which the outcome might be attributed."[14] Finally, measuring outcomes such as mortality can involve a considerable lag. Along with other factors, the desire to have a more immediate influence on quality has led many purchasers to focus on "aspects of care with proven relationships to desirable patient outcomes,"[15] which are more readily measured than patient outcomes.

One attempt to capture those aspects is process measures, which track a provider's adherence to accepted treatment guidelines that are based on scientific evidence. Rather than reward a provider on the basis of cholesterol levels or mortality rates for heart attack patients, a process measure would reflect how often a provider checks cholesterol levels or prescribes beta-blockers for heart attack patients. Providers who adhere to the recommended standard of care are rewarded with higher payments. Process measures are the most often discussed type of P4P quality measure; thus their potential shortcomings will be discussed in more detail below.

Structural quality measures attempt to evaluate the setting in which a provider delivers medical care. Such measures can include "the adequacy of facilities and equipment; the qualifications of medical staff and their organization; the administrative structure and operations of programs and institutions providing care; fiscal organization and the like."[16] Examples include whether a hospital uses health information technologies such as electronic prescribing, electronic medical records, or patient registries.

Structural quality measures have obvious appeal, but they also present limitations. The mere availability of sophisticated human and

physical capital offers no direct evidence of whether those resources are being used optimally. Meeting structural quality measures can also require large investments, which raise cost and cost-effectiveness implications.

Finally, patient satisfaction measures typically depend on surveys that ask patients about their experiences with a provider. Those measures presumably can capture aspects of quality that structural, process, and outcome measures cannot (e.g., convenience, waiting time, comfort, bedside manner, and level of trust between patient and physician). However, patient satisfaction measures also present shortcomings. Patient satisfaction is influenced by patients' expectations and their understanding of available alternatives. Both of those factors are influenced by the providers whose performance is to be judged.

Evidence-Based Quality Data

A third-party payer's ability to create financial incentives that guide providers toward recommended care depends on the availability of data that demonstrate a relationship between inputs and outcomes. Purchasers face significant challenges in accumulating accurate data that can be applied broadly.

Availability of Data

Pay-for-performance, also referred to as "quality-based purchasing" (QBP), depends on third-party purchasers having access to data that relate inputs to clinical outcomes. Such data exist for many but not all areas of care. According to one survey:

> A prominent barrier to QBP is that the science of performance measurement is still underdeveloped. Purchasers interested in QBP have limited choices for performance measures and these disproportionately target preventive care and structure or processes rather than outcomes. That is, the available set of metrics is not broadly representative of all care, while purchasers must pay for care across the entire clinical spectrum.[17]

At present, it is difficult or impossible to know for what share of health care expenditures such data exist.[18] Where widely recognized evidence-based data are not available, third-party purchasers have little ability to use financial incentives to drive quality improvements.[19]

Quality of Data

Where data are available, purchasers must consider the data's quality—that is, whether the available data lend themselves to performance measures that justify financial rewards. In order to serve as a basis for encouraging providers to change their behavior, the data employed must show a true relationship between a metric and a desired outcome. Moreover, accurate data can be misinterpreted or rendered out of date by subsequent research.

Ensuring that clinical data show a true relationship between a metric and a desired outcome is no small challenge. According to R. Brian Haynes, a prominent advocate of using more scientific data in clinical practice, "The advance of knowledge is incremental, with many false steps, and with breakthroughs few and far between, so that only a very tiny fraction of the reports in the medical literature signal new knowledge that is both adequately tested and important enough for practitioners to depend upon and apply."[20] Inaccurate findings are apparently not difficult to come by in the medical literature. Recent analyses suggest that one-third of frequently cited clinical studies are either incorrect or overstate the effect of clinical interventions[21] and that "most published research findings are false."[22]

Those analyses are themselves evidence that most research is vetted before it becomes the basis for clinical guidelines or performance measures. Nonetheless, concerns persist about the quality of data used in the clinical practice guidelines (CPGs) that often serve as the basis for performance measures and financial incentives. For example, "Professional organizations continue to issue recommendations on the basis of trials stopped early for benefit, including those . . . that seem most likely to overestimate effects."[23]

Concerns persist about the quality of data used in the clinical practice guidelines that often serve as the basis for performance measures and financial incentives.

Schemes that encourage providers to treat outliers like the average patient can create perverse incentives that encourage low-quality care.

And even accurate data grow old. Part of the challenge of P4P is to update performance targets and provider financial incentives on the basis of the most recent reliable data. That challenge is also not straightforward; experts often disagree about the significance or reliability of new clinical findings. According to one review of experts' use of new clinical information:

> Discrepancies were detected between the meta-analytic patterns of effectiveness in the randomized trials and the recommendations of reviewers. Review articles often failed to mention important advances or exhibited delays in recommending effective preventive measures. In some cases, treatments that have no effect on mortality or are potentially harmful continued to be recommended by several clinical experts.[24]

Whether a particular CPG's recommendations are overly cautious, hasty, or lack rigor is often a matter of opinion, and third-party purchasers have no clear guide for when they should incorporate new data. However, failure to assimilate accurate new data puts purchasers back where they don't want to be: paying for inferior quality.

Although assimilating new clinical data is essential, it also presents a tradeoff for purchasers. Collecting new data is costly. New data are often persuasive but not definitive. How often a purchaser chooses to integrate new data into the performance incentives it offers providers, and its threshold for the reliability of that data, will influence whether providers respond to the financial incentives and thus the effectiveness of those incentives.

Outliers

Another data-quality factor is the issue of outliers—patients who deviate from the mean either in their preferences or their response to treatment. Quality will have a different meaning for outliers than for most patients. P4P schemes that encourage providers to treat outliers like the average patient can create perverse incentives that encourage low-quality care. Because patients are often unaware of the financial arrangements between their insurer and provider, those perverse incentives can affect the quality of care without the patients' knowledge.

Some patients are clinical outliers. Even when randomized clinical trials accurately demonstrate the health benefits of an intervention, those benefits are not uniform among all patients within the trial, much less across the general population. A treatment's overall beneficial effects may hide different effects on subgroups, including no effect or even harmful effects. In addition, patients respond differently to a given intervention as a result of multiple illnesses or interactions with treatment regimens for such co-morbidities. Financial incentives that encourage providers to treat such outliers according to what benefits the majority of patients may inadvertently encourage low-quality or even harmful care.

The administration of beta-blockers to patients with cardiovascular disease is a common P4P quality measure. However, a recent study suggests that beta-blockers may not benefit all groups equally. Among acute coronary syndrome patients prescribed beta-blockers, certain genotypes had a lower rate of survival. The study's sample size was small (n = 735). Yet it plausibly suggests that compliance with a widely used P4P quality measure[25] could actually increase mortality among certain subgroups. The authors caution, "Further studies of the efficacy of b-blocker treatment . . . is [*sic*] warranted to be sure that we are not institutionalizing therapy through the adoption of health care quality performance measures that may offer little benefit, or even potential harm, to these patient subgroups."[26]

Another outlier challenge involves patients with co-morbidities. Many patients, particularly the elderly, suffer from multiple chronic diseases. Having multiple health conditions exposes patients to multiple treatment regimens and a correspondingly heightened risk of adverse drug events.[27] Most CPGs are

derived from clinical studies that exclude the elderly and patients with co-morbidities. Such guidelines in turn inform P4P measures.

Pay-for-performance measures that lack data specific to patients with co-morbidities can create significant perverse incentives for providers and quality problems for patients.[28] Following CPGs for each disease often results in multiple drug regimens. Yet little is known about how multiple medications, prescribed according to disease-specific guidelines, affect patients with numerous chronic conditions. It is thus possible that complying with P4P guidelines for treating each of a patient's chronic illnesses could lead to greater harm than if a provider made more individualized prescribing decisions.[29] The prospect of suffering financial penalties for providing individualized care to such patients could discourage providers from caring for patients with co-morbidities altogether.[30]

Other patient outliers deviate from the mean in their preferences for particular health outcomes.[31] Older patients and those with numerous health problems often have treatment goals that conflict with P4P measures:

> Is a statin or a beta-blocker, for example, as part of an 11-drug regimen, likely to provide greater benefit or greater harm to a 73-year-old whose priority is maximal energy, strength, and alertness today and who is willing to take on an increased risk of myocardial infarction or stroke over the next 5 or 10 years?[32]

For reasons of practicality, a P4P scheme might measure outcomes such as readmission or mortality rates, or processes such as statin or beta-blocker prescriptions, but not outcomes such as energy, strength, or alertness. Under such a payment system, a provider who treats such patients according to their wishes would be penalized for providing quality care.

In a P4P framework, outliers raise issues of equity between physicians. The inequities created by Medicare's quality-blind payment system are a primary reason why many observers consider that system unfair to high-performing providers.[33] A P4P scheme that encourages inappropriate care for outliers would create similar inequities.[34] Physicians who treat outliers according to clinical guidelines would be rewarded. But a physician who correctly treats a patient as an outlier—i.e., who deviates from the recommended protocol—would be penalized for doing the right thing.

The existence of outliers points to the limited usefulness of aggregate data in promoting quality. Much medical practice relies on the use of "unorganized knowledge"[35] about the circumstances of each patient and her preferences. As one advocate of EBM acknowledges: "Evidence from research can be no more than one component of any clinical decision. Other key components are the circumstances of the patient (as assessed through the expertise of the clinician), and the preferences of the patient."[36] Aggregate data will be applicable to large numbers of patients. However, it is difficult for a distant decisionmaker to identify those instances in which the data do not apply.

Tradeoffs

Beyond the challenges involved in collecting useful data, third-party purchasers face a second set of challenges: those associated with translating the data into performance measures and financial incentives. Here, too, the exercise is far from straightforward. Creating and administering P4P measures and financial incentives require making tradeoffs amid uncertainty. Different choices can potentially result in no effect, higher expenditures, inequities, reduced access to care, or even low-quality and inappropriate care. Those choices also affect whether providers respond to a P4P scheme the way that purchasers desire or in ways that defeat the effort.

Identifying the Optimal Target of Incentives

The first challenge is to determine which provider should be the target of the incentive. Poorly targeted financial incentives may create perverse incentives for providers to overprescribe, underprescribe, or unnecessarily com-

Pay-for-performance measures that lack data specific to patients with co-morbidities can create significant perverse incentives for providers and quality problems for patients.

Creating and administering P4P measures can potentially result in no effect, higher expenditures, inequities, reduced access to care, or even low-quality and inappropriate care.

partmentalize care. For most performance measures, the question is resolved if the patient receives care from an integrated health care system, such as a Kaiser Permanente. In those cases, the incentive would be targeted at the institution.

However, patients typically receive care in nonintegrated settings. For example, "An adult with diabetes mellitus could receive care regularly from an internist, cardiologist, ophthalmologist, and podiatrist, each of whom could adjust medications and share in the monitoring of disease complications and the side effects of treatment."[37] Which provider or providers should be penalized if the patient is not prescribed a recommended drug therapy, such as an angiotensin-converting enzyme (ACE) inhibitor? If the internist prescribes an ACE inhibitor and the cardiologist does not, should the cardiologist be penalized? What if the situation is reversed? Holding both responsible could lead to overprescribing and even less coordination of care. Holding only one responsible (say, the cardiologist) could also lead to overprescribing but also could lead to unnecessary compartmentalization of certain aspects of care (i.e., only cardiologists prescribing ACE inhibitors).

Whether a provider is responsible for a given outcome—and should therefore be the target of outcome-based financial incentives—can be even less clear. Continuing with the example: "If the patient requires a toe amputation that should have been preventable, which of several physicians and nurses caring for the patient should be considered responsible? To what degree does the patient bear responsibility?"[38]

What Types of Performance Targets?

Purchasers must also select the performance targets against which providers will be judged. Options include holding providers to an absolute standard (achievement), judging them against their peers (relative performance), judging them against prior performance (improvement), or some combination thereof. Each option presents tradeoffs that will affect providers' response to the financial incentives and the quality of care provided. Combining different types of performance targets can capture the benefits of each, but at the cost of added complexity.

An example of a performance measure based on absolute achievement would be one that rewarded all providers who prescribe beta-blockers to 90 percent of patients who suffer acute myocardial infarction (AMI). Such a measure gives each provider a clear picture of what is required to obtain the reward. An absolute goal helps providers plan their responses and can reduce uncertainty about whether an investment in improvement will pay off.

At the same time, absolute performance targets mostly reward providers who are already performing at the desired level. In one P4P initiative in which rewards were based on a fixed performance target, 75 percent of bonuses went to providers who were already exceeding the performance target. Such results may correct inequities that third-party payment systems create among providers. However, some observers note that absolute targets "may produce little gain in quality for the money spent."[39] Furthermore, fixed performance targets provide no incentive for providers to improve beyond the uppermost target.

In contrast, a performance target set relative to a provider's peers might reward the 10 percent of providers who have the highest rates of prescribing beta-blockers to AMI patients in a given year. The Medicare P4P demonstration program mentioned earlier awards bonuses to providers in the top two deciles in each of a number of metrics.[40]

A relative performance target is a moving target that depends on the behavior of other providers. Because the target cannot be known in advance, providers possess less certainty that a given compliance strategy will lead to a reward. As a result, performance targets that judge providers relative to their peers may result in increased compliance efforts among top performers but little effort at improvement among those who begin the competition farthest from the target. What research has been done on relative performance measures suggests that they may discourage compliance.[41]

Fixed and relative performance targets, which tend to set high standards for all providers, may fail to elicit responses from providers who begin the game far from the target. In contrast, financial incentives based on improvement over past performance create incentives for even those providers to improve. Such incentives, for example, could reward providers for every 10 percentage point improvement on a given metric.

Although such a performance target would increase the likelihood that poor performers would try to improve, it would do little to encourage improvement among providers in the top decile. Whether increments of improvement are judged along an absolute scale (percent of AMI patients who receive beta-blockers) or a relative scale (ranking providers on the basis of the rate at which they prescribe beta-blockers), providers already above the 90th percentile would have little incentive to improve. Moreover, using improvement as the sole criterion for financial rewards would create equity problems: poor performers could receive higher bonuses than providers with consistently high performance.

For a P4P arrangement to encourage improvement among all providers, it must employ some combination of financial incentives tied to absolute or relative performance targets, plus separate rewards for improvement. However, including both types of rewards adds complexity and essentially would give all providers an opportunity to increase their incomes, which creates affordability problems. One solution to that problem would be to offset the cost of rewards through the use of penalties—that is, by reducing payments to poor performers. However, the prospect of reduced incomes makes it more likely that providers would resist a P4P scheme.

Size of Financial Incentives

The size of financial incentives offered to providers is a key consideration. Incentives that are too small will fail to induce behavioral change.[42] On the other hand, incentives that are too large can encourage cost-ineffective or even inappropriate care, as well as make a P4P program unaffordable. The task is further complicated by the fact that provider characteristics will cause providers to respond to the same incentive in different ways.

Some observers suggest that financial incentives must account for at least 10 percent of a physician's income.[43] However, a provider's response to a financial incentive depends primarily, not on the absolute size of the incentive, but on the *net* size of the incentive. Suppose a provider could obtain a $90,000 bonus by implementing an electronic patient registry. If the cost of implementing the registry is $100,000, the incentive would have no effect. To cause this provider to change his behavior, the bonus must exceed $100,000. That is, to change a provider's behavior, net revenue (R_N) must be positive, meaning the actual financial incentive (R_A) must exceed the cost to providers of compliance (C_C):

$$R_N = R_A - C_C$$

This insight is important because providers will have different compliance costs.

For example, the largest award Medicare granted to a hospital in its Premier demonstration in 2005 was $326,000 given to Hackensack University Medical Center for compliance with performance targets for coronary artery bypass grafts. If Hackensack University Medical Center were already a top performer in that area, its compliance costs would be close to zero. But lower performers would have higher compliance costs. Does $326,000 per year provide enough of a net incentive to encourage average- or poor-performing hospitals to make the investments necessary to reach the desired level of quality? If not, few will make those investments.

Another factor complicating the calibration of financial incentives is providers' income goals. For example, "A provider whose income is at or near a preferred income target may be less likely to respond to an incentive of a given amount than a provider who is not yet achieving his or her target income."[44]

One way to encourage greater compliance is to ensure that the corresponding disincentives

For a P4P arrangement to encourage improvement among all providers, it must employ some combination of financial incentives. However, that adds complexity and creates affordability problems.

for noncompliance are large enough to encourage providers to invest in meeting the performance targets. Such disincentives could be merely relative—that is, noncompliant providers would be held harmless in real terms but paid less than compliant providers. A more controversial option is to impose real payment reductions on noncompliant providers. Financial penalties can improve overall affordability, but at the cost of provider resistance. Nonetheless, without financial penalties, P4P can only increase health expenditures. Dartmouth's Jack Wennberg notes, "Unless pay-for-performance focuses on rewarding providers who are *efficient* in the delivery of these services it will have little impact on overall costs and poor quality associated with too much care."[45]

Other important choices pertain to the timeliness of rewards and the frequency with which they are altered or updated. Collecting compliance data takes time, but long delays between desired behavior and rewards reduce the value of those rewards. Some P4P programs can involve reward lags of six months or more.[46] Likewise, third-party payers may want to update the size or other aspects of financial rewards on the basis of new information. However, frequent changes to performance targets or financial incentives reduce the certainty and thus the value of those rewards.

Cost-Effectiveness

Designing a P4P program that is cost-effective is also a significant challenge. Collecting evidence-based quality data, translating those data into performance measures, collecting data on provider compliance, distributing rewards, defending penalties, and continually updating a P4P scheme all involve significant financial commitments. An important cost dimension is the hidden costs that P4P might impose by encouraging inappropriate care or reducing access. The costs of P4P have yet to be quantified, much less compared to the potential benefits.

The costs of P4P have yet to be quantified, much less compared to the potential benefits.

Even if P4P delivered demonstrable improvements in health care quality, that would not demonstrate that P4P is worthwhile. Despite significant health gains, many or all P4P designs could impose even larger costs. For example, the cost of implementing a P4P scheme could lead private insurers to increase premiums. That in turn could reduce access to health coverage and lead to offsetting health losses. The problem exists in public programs as well. The cost of implementing P4P in Medicare could require spending reductions that reduce the quality of care elsewhere in the program, or higher taxes that make it more difficult for the nonelderly to afford coverage. In sum, the number of quality-adjusted life-years a P4P scheme "purchases" could be outweighed by the number of such years lost to a combination of inappropriate outlier care and reduced access.

How Will Providers Respond?

Quality-based purchasing is designed to affect the behavior of providers for the benefit of patients. Yet providers are highly suspicious of P4P efforts,[47] which have the potential to reduce provider incomes. The impact of a P4P scheme will be shaped in part by whether providers respond to financial incentives in the desired manner. Providers may respond in ways that defeat the exercise and even leave some patients worse off.

Will Providers Buy In?

For financial incentives to encourage providers to change their behavior, providers first must believe the performance targets are attainable. Yet many factors that influence a provider's ability to meet performance targets are beyond a provider's control. For example, outcome measures are affected by a patient's underlying health status. Providers with sicker-than-average patients could be penalized for below-average outcomes, even if the care provided is of the highest quality. Most efforts to judge providers on patient outcomes are risk adjusted, that is, they attempt to hold constant the severity of illness so that providers will not be penalized for treating sicker patients. Risk adjustment is meant to address the concern that "a

provider could be rewarded and penalized based on the patients it attracts, rather than the quality of care it delivers."[48] Nonetheless, "Some hospital leaders [have] expressed the view that 'not in the near future, nor possibly ever, will we develop a reliable severity adjustment system.'"[49] Outcome measures and risk adjustment are likely to be perennial battlegrounds on which providers are pitted against those seeking to measure quality.

Other patient demographics may also influence a provider's ability to meet performance measures. A low-income patient is less likely to be able to afford all the prescriptions recommended for her multiple conditions. In such cases, providers may rationally choose to focus on a smaller number of affordable medications that offer the greatest benefit. If a P4P program financially penalizes such providers, it would punish them for the type of patients they treat, rather than their performance.

Some performance measures depend on patient cooperation, another factor often beyond the provider's control. One such measure is patient participation in smoking cessation programs. Providers may have limited ability to persuade smokers to enroll in such programs.

Other factors will affect providers' receptivity to P4P schemes. Are there too many schemes?[50] If so, "a physician paid for diabetes control one way by the government and another way by the private sector might simply throw up his hands and ignore them both."[51] Are they too complex? Are they changed too frequently? Providers' willingness to comply with P4P standards will decline as the number and complexity of schemes increase. Do providers perceive the standards to be based on reliable data? Are payers open about the quality criteria?[52] Providers' reactions to any of those factors will be magnified by the size of the financial incentives involved, especially if the financial disincentives include potential losses in income.

Physicians have expressed opposition to many potential P4P designs. The American Medical Association has issued an official policy on the development of P4P programs. It states, "The primary goal of any P4P program must be to promote quality patient care that is safe and effective across the health care delivery system, rather than to achieve monetary savings." According to the AMA, all P4P programs must (1) be completely voluntary; (2) reimburse physicians the cost of their participation; (3) finance rewards with supplemental funds; (4) use "the best-available risk-adjustment"; (5) keep program features stable for at least two years[53]; and (6) allow for deviation from guidelines when clinically appropriate with "minimal, but appropriate, documentation." In addition, P4P programs must not (1) employ financial penalties, (2) judge individual physicians relative to one another, (3) "threaten the economic viability of physician practices" that do not participate, (4) judge physicians on the basis of factors beyond their control, or (5) limit patient access to care.[54] Opposition from the AMA is one reason a P4P proposal was dropped from the fiscal year 2006 budget resolution.

Will Providers Revolt?

If providers believe performance standards are too complicated or lack merit, or that they are being penalized for factors unrelated to their performance, they may act to undermine P4P efforts—not necessarily without reason, but often in ways that could harm patients. For example, providers could exert no effort to reach a P4P scheme's performance targets. In that case, patients often would receive care no better than they would have received in the absence of P4P. Alternatively, providers could refuse to do business with third-party payers who tie payment to "unreasonable" performance measures. That response would disrupt many patients' access to care. Finally, providers could respond in ways that undermine the effectiveness of the financial incentives. Principally, those responses involve various ways that providers can "game the system"—preserving or increasing their incomes through technical compliance with performance measures but without improving (and often reducing) the quality of care.

Providers may act to undermine P4P efforts—not necessarily without reason, but often in ways that could harm patients.

It is generally accepted that the use of clinical evidence in treatment decisions has been suboptimal. However, each provider can expect to treat some patients for whom "quality" will not be defined by the results of clinical trials.

Providers who believe they are being penalized for variables beyond their control can be expected to influence the variables they can control in order to protect their incomes. One method—patient selection—could jeopardize many patients' access to care. "If hospitals are paid for good surgical outcomes, they will want to operate only on the healthiest people."[55] If third-party payers reward providers on the basis of their patients' cholesterol levels, providers will select patients who are most likely to stick to a cholesterol-lowering treatment regimen. And they will avoid those, such as low-income patients or those with multiple chronic illnesses, who will have the most difficulty complying with doctor's orders. Those patients could become "medical hot potatoes"[56] who would find it increasingly difficult to obtain care and could be relegated to low-quality providers.

Another method is data manipulation.[57] As many as 50 percent of physicians admit they have manipulated third-party reimbursement rules to secure coverage of a particular treatment for a patient (and payment for themselves). As many as 70 percent of physicians state they would be willing to do so under certain circumstances.[58] Physicians report manipulating data in ways that could easily be used to game and defeat P4P measures.[59] Returning to the example of the above-mentioned "73-year-old whose priority is maximal energy, strength, and alertness today and who is willing to take on an increased risk of myocardial infarction or stroke over the next 5 or 10 years,"[60] suppose the physician would be rewarded for prescribing the patient a statin or beta-blocker. The physician might prescribe such medications but counsel the patient not to fill the prescription. The third-party payer would see only that the physician had complied with the performance measure. That strategy would reward the physician yet defeat the purpose of the financial incentive because the course of treatment would not change. (Perversely, it would be precisely because the strategy avoided the desired behavior that it improved the patient's quality of care.)

Finally, providers may be able to defeat P4P incentives by manipulating the intensity of care. Research has documented wide regional variations in health care spending on similar patients.[61] Much of this spending is the result of greater intensity of care, such as more frequent hospital admissions and specialist consultations. Where third-party purchasers are blind to overuse, providers who are unwilling or unable to meet P4P performance targets may be able to preserve their incomes by increasing the intensity of the care they provide. Such strategies would increase expenditures while potentially reducing quality.

How Will P4P Affect Experimentation and Learning?

Another important consideration is the effect that financial incentives will have on experimentation and learning. It is generally accepted that the use of clinical evidence in treatment decisions has been suboptimal and that providers have traditionally relied too heavily on the "art of medicine" or "clinical judgment."[62] However, as discussed earlier, clinical trials report average effects of interventions on patients who are selected for their lack of co-morbidities. Thus, while clinical evidence is essential, each provider can expect to treat some patients for whom "quality" will not be defined by the results of clinical trials.

In those cases, incentives to treat outliers like average patients could discourage providers from using their clinical judgment where it is appropriate. In any P4P scheme, providers arguably should be free to deviate from an "average patient" standard when dealing with patient subgroups for whom no evidence-based CPGs exist. Allowing that flexibility would enable providers to discover and disseminate new modes of treatment that later may be scrutinized in clinical trials.

Whether and how a P4P scheme creates such flexibility will affect both provider participation and the quality of care for outliers. Many performance targets include some flexibility. A target that rewards physicians when 90 percent of AMI patients are prescribed beta-blockers allows physicians to deviate

from the standard in 10 percent of cases. But payers and providers will differ over whether 10 percent is sufficient flexibility or too little.

What Do the Data Say about the Effectiveness of P4P?

Preserving providers' ability to exercise clinical judgment is particularly important when one considers the lack of evidence showing that evidence-based guidelines actually lead to better patient outcomes. According to one pioneer of EBM:

> A fundamental assumption of EBM is that practitioners whose practice is based on an understanding of evidence from applied health care research will provide superior patient care compared with practitioners who rely on understanding of basic mechanisms and their own clinical experience. So far, no convincing direct evidence exists that shows that this assumption is correct.[63]

In fact, much the same can be said of the performance of P4P. Although the aim of P4P is to use evidence to drive higher-quality care, very little evidence has been collected that shows that P4P actually delivers on its promise.

A 2004 survey of the literature found that data on the effectiveness of P4P are "sparse."[64] Researchers could locate only eight randomized, controlled studies that measured the ability of performance-based financial incentives to change provider behavior or to improve patient outcomes. The results were mixed. The studies obtained both positive and negative results when financial incentives were targeted to individual physicians, individual providers, and provider groups. According to the authors, "There was no consistent relationship between the magnitude of the incentive and response, and in fact the largest single incentive (the bonus of up to $10,000) was ineffective."[65] As noted earlier, the studies examining relative performance targets, where providers lacked certainty about what would be required to obtain the reward, obtained negative results. Two types of financial incentives (fee-for-service payment enhancements and bonuses) both showed mixed results. Results were also mixed for studies that measured whether financial incentives encouraged the use of preventive care. Insofar as the studies provided any consistent evidence, it was that, "in a general sense . . . incentives to achieve performance were more effective when the indicator to be followed required less patient cooperation (e.g., receiving vaccinations or answering questions about smoking) than when significant patient cooperation was needed (e.g., to quit smoking)."[66] While P4P may be a useful tool for improving health care quality, its effectiveness at changing provider behavior or improving outcomes has not been established.

Nor has the cost-effectiveness of P4P been established. The expense of a P4P scheme is incurred with the expectation that it will be outweighed by improved health and cost savings. Yet little attention has been paid to whether those hoped-for results are worth the cost, largely because effectiveness must be established before cost-effectiveness can be measured.

The P4P movement proceeds from two premises: first, that clinicians tend to underuse evidence from randomized clinical trials and, second, that financial incentives can increase such use and improve the quality of care. Yet whatever enthusiasm exists for P4P is not derived from the type of evidence of effectiveness that P4P enthusiasts believe should guide clinical practice. Third-party financial incentives remain an unproven tool for improving health care quality at all, let alone in a cost-effective manner.

Special Challenges Posed by Medicare

The preceding discussion describes difficulties that confront any third-party payer seeking to create payment incentives that reward providers on the basis of quality. Incorporating quality-based financial incen-

Whatever enthusiasm exists for P4P is not derived from the type of evidence of effectiveness that P4P enthusiasts believe should guide clinical practice.

Incorporating quality-based financial incentives into traditional Medicare poses additional challenges not faced by private third-party payers.

tives into traditional Medicare poses additional challenges not faced by private third-party payers. Medicare is subject to the political process, which increases the potential for error at each stage of designing, implementing, and maintaining a P4P scheme. Medicare's size and market dominance guarantee that the impact of its errors (and successes) would be magnified. Finally, since Medicare influences the payment rates and coverage decisions of private insurers, the impact of those choices would also extend well beyond Medicare.

Greater Potential for Error

Medicare's ability to use provider-focused financial incentives to improve quality is encumbered by factors not faced by private purchasers. Medicare's patient population is more susceptible to P4P pitfalls. The politicization of Medicare's every decision gives that program less flexibility to design a P4P program for the benefit of patients. Those same forces also make it more difficult for Medicare to correct the errors it commits.

Any harm that a P4P system could conceivably cause is more likely to appear in Medicare. Medicare's patient base is more susceptible than those of private insurers to the unintended harms that can result from P4P programs. Medicare enrollees are older and less healthy than non-Medicare enrollees. Close to one-third of Medicare beneficiaries have four or more chronic conditions, and those patients account for nearly 80 percent of Medicare spending.[67] The large number of beneficiaries with chronic conditions increases the likelihood that a P4P scheme would create incentives to mistreat such patients and turn them into "medical hot potatoes"[68] that providers make an effort to avoid. Those patients are at the highest risk for the type of adverse drug interactions that can come from strict adherence to multiple CPGs. Moreover, Medicare patients are the most likely to have treatment goals that differ from those assumed by CPGs and P4P measures. As a result, a Medicare P4P effort is more likely to create harmful financial incentives than are private efforts.

Compounding those challenges is the fact that Medicare is a creature of the political process. The political forces that govern Medicare will shape each phase of a P4P initiative. Insulating that process from politics would be impossible. The choices involved would directly affect the incomes of up to half a million Medicare-participating physicians, plus thousands of health care facilities and manufacturers of medical products, all of whom depend on Medicare for their livelihood. The tradeoffs made in structuring a Medicare P4P program would also affect the quality and accessibility of care for some 40 million seniors, the tax burden of hundreds of millions of Americans, and the availability of federal revenues for other priorities.

Parties with a stake in the tradeoffs involved would seek to influence Congress, the Centers for Medicare & Medicaid Services (CMS), and whatever other bodies make or influence those choices. For nearly a decade, health care has led other sectors of the economy in terms of dollars spent lobbying Congress.[69] Employing P4P financial incentives in Medicare would help solidify that lead. That unavoidable pressure would reduce Medicare's flexibility to make timely, focused, and evidence-based adjustments to its payment structure and would put upward pressure on Medicare outlays. The politicization of quality-based financial incentives can in turn be expected to politicize the search for data to guide those incentives. Thus, interested parties may also seek federally financed research on modes of care that they believe should be rewarded for being of higher quality.

Indeed, rent-seeking behavior will attend every aspect of a Medicare P4P system. As do Medicare's payment systems broadly, a P4P system would guarantee congressional and administrative lobbying by providers who seek to protect or increase their incomes, who fear being penalized for factors beyond their control, who don't want to change the way they practice, who want additional research devoted to their modes of care, who seek to gain advantages over their competitors, who wish to ensure the performance measures

can be gamed, who don't want the P4P system updated too frequently, and who want only one set of performance targets set by Medicare and adopted by private insurers.

Political pressures would also push in the opposite direction. A Medicare P4P system would be of perennial interest to those who see P4P as a way to reduce Medicare outlays. Given the future financial pressures facing Medicare[70] and the growing scarcity of federal resources for other congressional priorities (e.g., tax cuts, deficit reduction, new spending), politicians and interest groups can be expected to exploit opportunities to squeeze provider payments. In contrast to past reductions in Medicare provider payments, P4P would allow future cuts to be packaged as quality enhancing.

However, it is reasonable to predict that provider groups and Medicare beneficiaries would have the greatest influence over a Medicare P4P scheme and that such a scheme would increase Medicare outlays significantly. Providers and seniors have a more direct stake in how such a scheme is structured than do others. The benefits of their preferred policies are large and concentrated on relatively small groups that are relatively easy to organize for political action. By comparison, the per capita benefits of using P4P to constrain Medicare spending are smaller and more diffuse. That is, those smaller benefits are spread out among a larger group of individuals that is more difficult to organize. Trade groups often wield a disproportionate influence over public policies that affect their incomes and can be expected to prevail over other interest groups more often than not.[71] Providers are therefore likely to block P4P rules that would reduce their incomes. Likewise, Medicare beneficiaries (with the help of provider groups) are likely to block rules that reduce their access to care. The likely result is that a Medicare P4P effort would be able to create financial incentives only by increasing outlays.

The political forces governing Medicare would also make a Medicare P4P system more rigid and slower to adapt than private P4P schemes. As a result, Medicare would take longer to correct errors than do private third-party purchasers.

A Medicare P4P system essentially would build on Medicare's administrative pricing responsibilities. Medicare currently operates 16 different payment systems for various types of providers and health plans. The physician payment system alone must divine prices for more than 7,000 distinct services in each region of the country.[72] P4P would pile even more complexity on top of that system. In essence, where CMS need now divine only one quality-blind payment for a service, P4P would require the agency to devise two payment levels: one for high-quality providers and one for low-quality providers.

Medicare's administered pricing systems have been criticized for spurring overinvestment in some areas of care while reducing access in others[73] and for being slow to fix such errors.[74] Technological advances and productivity gains in ambulatory surgical centers (ASCs) have reduced the cost of care in those facilities. Yet Medicare payments to ASCs have not been adjusted to account for any such changes since 1988. This has led to a situation in which ASCs are often paid far more than hospital outpatient centers for the same procedure. Moreover, ASC payments will not be adjusted for these factors until 2008.[75]

The same rigidity would govern Medicare's administration of performance-based payments. Inaccurate data, mistargeted or miscalibrated financial incentives, or perverse incentives that result in low-quality care could live on within Medicare well after a private purchaser might correct the error. This is particularly true where interest groups would have an incentive to preserve Medicare's error (e.g., where providers are receiving excessive financial bonuses).

A Large Wake

Not only would errors be more likely in a Medicare P4P program, but the resulting harms would be far more widespread than those of similar errors by private payers. In addition to having a higher *proportion* of

Inaccurate data, mistargeted or miscalibrated financial incentives, or perverse incentives that result in low-quality care could live on within Medicare well after a private purchaser might correct the error.

patients who are likely to be harmed by the unintended consequences of a P4P program, Medicare covers a much larger *number* of individuals than any private insurer. Medicare is the single largest purchaser of medical care in the United States, with 37 million elderly and disabled enrollees in traditional Medicare and another five million enrolled in private plans. Medicare beneficiaries consume more medical care than the nonelderly. Finally, they are also less able than non-Medicare patients (or those patients' employers) to switch insurers if a P4P program causes unintended harm.

Moreover, any harms resulting from a Medicare P4P system likely would spill over into the private sector. Private insurers tend to follow Medicare's lead regarding certain business decisions. Many private insurers make it a policy to cover whatever services Medicare approves for coverage. Likewise, private insurers' payments are often heavily influenced by Medicare's payment rates.[76] If the federal government creates one P4P system for the entire Medicare program, insurers would face strong incentives to adopt that system as well. Many third-party purchasers (including private insurers and state governments) likely would rather have Medicare incur the costs of creating and maintaining a P4P scheme than incur those costs themselves. As a result, a Medicare P4P scheme likely would crowd out more flexible private efforts, and any harm it created could spread beyond the Medicare population.

Confining P4P to private Medicare plans would create a learning process that would allow for testing and refining P4P strategies.

How to Address the Unique Pitfalls of Medicare P4P

P4P gives third-party payers considerable power to influence—for good or ill—the quality of care that patients receive. The potential for harmful errors generally, and in Medicare in particular, suggests that at a minimum individual patients should have the ability to move between health plans that employ P4P. The potential for harm can also be mitigated by cost-sharing features that assign patients greater financial responsibility for care rendered by providers who deviate from CPGs.

Confine P4P to Private Medicare Plans

One option that would enable Medicare beneficiaries to reap the benefits of P4P while minimizing any harms would be to restrict the use of provider-focused P4P incentives to private Medicare Advantage plans. Seniors could then select a health plan on the basis of a number of features, including its P4P scheme. Plans that pay for performance would be able to market themselves on the basis of quality and cost-effectiveness. An enrollee could switch plans during the annual open enrollment period (or perhaps more frequently) if she and her doctor determined that her health plan's P4P incentives were interfering with the quality of her care.

Confining P4P to private Medicare plans would also create a learning process that would allow for testing and refining P4P strategies. Instead of creating a single set of P4P measures and incentives, private Medicare plans would experiment with multiple, competing P4P efforts. Best practices could be retained and emulated by other plans. As important, the harms resulting from ill-conceived financial incentives would be confined to smaller populations, and those incentives could be discarded sooner. Over time, all plans would gravitate toward whatever successful P4P strategies emerged, while keeping unintended harms to a minimum.

Preventing CMS from developing P4P incentives for traditional Medicare is necessary to create this learning process. A P4P scheme in traditional Medicare would apply to 37 million seniors and would effectively crowd out private efforts to develop competing P4P programs. Private plans would be much less likely to incur the costs involved with P4P when they could adopt the Medicare standards at close to zero cost. Moreover, even when private insurers sought to create alternative P4P programs, those efforts would be less likely to be accepted by providers. "In many markets Medicare and Medicaid comprise over 65 percent of the

payments to hospitals, and more than 80 percent in some physician specialties."[77] Necessity would force providers to give highest consideration to Medicare's performance measures. Providers likely would ignore financial incentives offered by other payers, particularly if those incentives applied to small patient populations or entailed high compliance costs (roughly measured by the degree to which a provider would have to deviate from what Medicare requires). Even where private P4P programs were accepted, private payers would incur the entire cost of implementing those programs but only reap benefits above and beyond what Medicare's standards would provide for free. For the sake of constantly improving the performance of P4P, it is important not to create a P4P system in traditional Medicare.

One criticism of confining P4P to private Medicare plans is that those plans face incentives to screen out seriously ill seniors, and P4P would give them another tool for doing so. For example, a plan could require strict adherence to CPGs without regard to co-morbidities, which would encourage seniors with multiple chronic conditions to avoid that plan. This valid concern arises from a problem similar to the difficulty involved in applying CPGs to atypical patients. In each case, a third-party payer is trying to treat an outlier as though she were the average patient.

Private Medicare plans face incentives to screen out chronically ill and other high-cost patients if and when Medicare pays plans less than the cost of treating those patients. The solution to screening is for Medicare to adjust its payments on the basis of the expected health costs of each patient. Medicare is currently refining its risk-adjustment capabilities. A further step would be for Medicare to subsidize patients directly, which would encourage plans to compete for outliers by tailoring offerings to those patients.

Allowing P4P only in private Medicare plans would confine P4P to a maximum of five million Medicare beneficiaries at present, or 12 percent of enrollees. However, the reach of P4P could be expanded—and the refinement of P4P tools accelerated—by encouraging more Medicare beneficiaries to enroll in private plans. The bipartisan premium support proposals[78] advanced in late 1990s would encourage greater private plan enrollment and would give individual seniors a stake in evaluating the cost-effectiveness of P4P strategies, as well as health plans overall.

The refinement of P4P tools could be accelerated by encouraging more Medicare beneficiaries to enroll in private plans.

Use Patient-Focused Financial Incentives

Pay-for-performance, where third-party purchasers offer financial incentives to providers, is a tool of unknown value. Much more research is necessary before payers can know whether and where P4P will change providers' behavior and improve patient outcomes. Thus, it is important that payers not focus solely on this approach. Other measures may also induce providers to improve the quality of care.

One possibility is financial incentives that increase patients' interest in high-quality care. Most P4P initiatives attempt to influence the behavior of providers. A weakness of this approach is that it does not involve the patient—in fact, patients could be completely unaware of the financial incentives that affect the care they receive. Engaging the patient in the pursuit of quality could educate patients about superior modes of care, have a greater influence on provider behavior, and still allow patients to avoid the harms that may result from hidden financial incentives targeted to providers.

The same sort of data that third-party payers use to create financial incentives for providers could be used to create financial incentives that encourage patients to demand higher-quality care. Payers could adjust out-of-pocket exposure such that patients who receive recommended care would face lower out-of-pocket costs, while those who do not would face higher out-of-pocket costs.

Patients would soon know whenever a provider was not adhering to the plan's quality guidelines because that deviation would affect the patient's pocketbook. In such cases, a dialogue between the patient and provider (and perhaps the health plan) would ensue. Both the price signals offered by the

Patient-focused financial incentives offer benefits that provider-focused incentives do not.

plan and the subsequent dialogue would lead to better-educated patients who would help drive quality improvements.

Finally, patient-focused financial incentives would offer protection to patients who might be inadvertently harmed by inappropriate provider-focused incentives. When the financial incentives are targeted to the patient, they are transparent. When a patient and her provider disagree with the health plan's recommendations, they would be free to disregard the recommendations and pay the higher coinsurance. Because such tiered coinsurance would keep the locus of decisionmaking at the level of the patient, it may be more appropriate for traditional Medicare than are provider-focused incentives.

Patient-focused financial incentives would face some difficulties similar to those facing provider-focused incentives. For example, health plans would have to take steps to ensure that patients with co-morbidities and other clinical outliers would not be penalized for choosing appropriate care that happens to deviate from the standard. In addition, patients without the means to pay higher coinsurance would be in the same position as a patient whose care is influenced by provider-focused financial incentives that she can neither control nor see. Nonetheless, patient-focused financial incentives are another arrow in the quiver of third-party purchasers (including Medicare) and offer benefits that provider-focused incentives do not.

Conclusion

America's health care sector is marked by substantial variations in health care quality. The purchasing power of large third-party payers—such as Medicare—presents an opportunity to encourage low-performing providers to improve the quality of care they deliver. Pay-for-performance offers a way to steer providers toward modes of care that have been demonstrated to improve patient health.

However, P4P is an unproven tool with significant potential pitfalls. Developing, implementing, and maintaining the financial incentives required to steer provider behavior are not straightforward tasks. Any number of errors—including false or misinterpreted data and mistargeted or miscalibrated financial incentives—could inadvertently encourage low-quality care or reduce access to care. Even when financial incentives are based on accurate data, not all patients hew to the mean. Encouraging providers to treat each patient as though she were the average patient can harm outliers.

The potential for error that exists in any P4P effort would be magnified in traditional Medicare. The political forces that govern Medicare increase the potential for error and would increase the duration of such errors. Many seniors would have difficulty avoiding the resulting harms, given that traditional fee-for-service Medicare is often the only game in town. Moreover, the introduction of P4P into traditional Medicare likely would crowd out private efforts to develop P4P financial incentives. As they do with regard to coverage determinations and provider reimbursements, private insurers would face strong incentives to follow Medicare's lead. That would unnecessarily constrict experimentation and competition among P4P schemes.

A better approach to introducing P4P into Medicare would be to restrict the use of provider-focused financial incentives to private Medicare plans. Doing so would allow patients to avoid P4P designs that create perverse incentives and would allow private plans to experiment and learn from each other's successes and failures. Moreover, it would offer the benefits of P4P to Medicare enrollees without having traditional Medicare create a de facto national P4P program. If traditional Medicare is to use financial incentives to drive quality, those incentives would be better targeted to individual patients. In either case, the ultimate locus of decisionmaking would be at the level of the individual.

The potential risks of broadly applicable P4P systems are serious enough that those adversely affected should have the right to opt out of those systems—and perhaps the respon-

sibility of bearing the cost of that choice. Moreover, P4P holds enough promise that special interests should not be able to stymie its development through political pressure.

Notes

1. Elizabeth A. McGlynn et al., "The Quality of Health Care Delivered to Adults in the United States," *New England Journal of Medicine* 348 (June 26, 2003): 2635, http://content.nejm.org/cgi/content/abstract/348/26/2635.

2. Elliott S. Fisher et al., "The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care," *Annals of Internal Medicine* 138, no. 38 (February 18, 2003): 273–87, http://www.annals.org/cgi/reprint/138/4/273.pdf; and Elliott S. Fisher et al., "The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care," *Annals of Internal Medicine* 138, no. 38 (February 18, 2003): 288–98, http://www.annals.org/cgi/reprint/138/4/288.pdf. See also Priscilla Hollander et al., "Quality of Care of Medicare Patients with Diabetes in a Metropolitan Fee-for-Service Primary Care Integrated Delivery System," *American Journal of Medical Quality* 20, no. 6 (November–December 2005): 344–52, http://ajm.sagepub.com/cgi/reprint/20/6/344.

3. See, e.g., U.S. Federal Trade Commission and U.S. Department of Justice, *Improving Health Care: A Dose of Competition*, July 23, 2004, Executive Summary, pp. 5, 16, http://www.ftc.gov/reports/healthcare/040723healthcarerpt.pdf; and David M. Cutler, *Your Money or Your Life* (Oxford: Oxford University Press, 2004), p. 101.

4. Such a payment system may or may not improve the overall quality of analysis, but assuredly it would result in the use of low-quality endnotes.

5. U.S. Federal Trade Commission and U.S. Department of Justice, chap. 2, p. 30.

6. Medicare Payment Advisory Commission, *Report to Congress: Variation and Innovation in Medicare*, June 12, 2003, p. 108, http://www.medpac.gov/publications/congressional_reports/June03_Entire_Report.pdf. "When Intermountain Health Care, a Salt Lake City hospital system, improved care for its pneumonia patients by making sure they received the right drugs, it lost money because Medicare continues to pay less when patients have fewer complications and require less extensive care." Reed Abelson, "Medicare Says Bonuses Can Improve Hospital Care," *New York Times*, November 15, 2005, http://www.nytimes. com/2005/11/15/business/15care.html?ei=5070&en=37f0cc8cf663d311&ex=1133067600&adxnnl=1&adxnnlx=1132936064-WtzUTO8qAm9un QpLnO+BDA.

7. John E. Wennberg, "Variation in Use of Medicare Services among Regions and Selected Academic Medical Centers: Is More Better?" Duncan W. Clark Lecture, New York Academy of Medicine, January 24, 2005, p. 2, http://www.dartmouthatlas.org/atlases/NYAM_Lecture.pdf.

8. Cutler, p. 101.

9. Karen Ignagni, president and chief executive officer, America's Health Insurance Plans, Testimony before the Subcommittee on Health of the House Committee on Ways and Means, September 29, 2005, http://waysandmeans.house.gov/hearings.asp?formmode=view&id=3820#_ftn8.

10. U.S. Centers for Medicare & Medicaid Services, "Medicare Demonstration Shows Hospital Quality of Care Improves with Payments Tied to Quality," press release, November 14, 2005, http://www.cms.hhs.gov/media/press/release.asp?Counter=1729.

11. Reed Abelson, "Medicare Says Bonuses Can Improve Hospital Care," *New York Times*, November 15, 2005, http://www.nytimes.com/2005/11/15/business/15care.html?ei=5070&en=37f0cc8cf663d311&ex=1133067600&adxnnl=1&adxnnlx=1132936064-WtzUTO8qAm9unQpLnO+BDA.

12. R. Adams Dudley et al., "Strategies to Support Quality-based Purchasing: A Review of the Evidence," U.S. Department of Health and Human Services, Agency for Healthcare Research and Quality, AHRQ pub. no. 04-0057, July 2004, p. 68, http://www.ahrq.gov/downloads/pub/evidence/pdf/qbpurch/qbpurch.pdf, citing C. D. Ittner and D. F. Larcker, "Coming Up Short on Nonfinancial Performance Measurement," *Harvard Business Review* 81, no. 11 (November 2003): 88–95, 139.

13. For instance, "although fixing a congenitally dislocated hip joint in a given position is considered good medicine for the white man, it can prove crippling for the Navajo Indian who spends much time seated on the floor or in the saddle." Avedis Donabedian, "Evaluating the Quality of Medical Care," *Milbank Quarterly* 83, no. 4 (2005): 694, reprinted from *Milbank Memorial Fund Quarterly* 44, no. 3, pt. 2 (1966): 166–203.

14. Ibid., p. 694.

15. U.S. Agency for Health Care Research and Quality, "Patient Safety Network Glossary: Structure-Process-Outcome Triad," http://psnet.ahrq.gov/glossary.aspx (accessed November 26, 2005).

16. Donabedian, p. 695.

17. Dudley et al., p. 7.

18. The Centers for Medicare & Medicaid services provide the following slippery description of the availability of such data (emphasis added): "A preliminary assessment indicates that *the specialties* for which *some* measures have been developed *account for about half* of Medicare *physician* spending. Specialties *accounting for* another 40 percent of *physician* spending have measures under development. . . . In addition, virtually all specialties have noted that evidence-based guidelines for best practices have been developed for *many* important aspects of the care they provide. Such guidelines *do not apply to all patients* receiving care from a particular specialty, but they do generally reflect the state of medical evidence about what works best in the specialty for *many* of the common problems they treat." U.S. Centers for Medicare & Medicaid Services, letter to the Honorable William M. Thomas, June 24, 2005, p. 4.

19. However, one avenue of inquiry that bears examination is whether there is a "quality spillover" effect from measured to unmeasured areas of care.

20. R. Brian Haynes, "What Kind of Evidence Is It That Evidence-Based Medicine Advocates Want Health Care Providers and Consumers to Pay Attention to?" *BMC Health Services Research* 2, no. 3 (March 6, 2002), http://www.biomedcentral.com/1472-6963/2/3.

21. John P. A. Ioannidis, "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research," *Journal of the American Medical Association* 294 (July 13, 2005): 218–28, http://jama.ama-assn.org/cgi/content/abstract/294/2/218.

22. John P. A. Ioannidis, "Why Most Published Research Findings Are False," *PLoS Medicine* 2, no. 8 (August 2005): e124, http://medicine.plosjournals.org/archive/1549-1676/2/8/pdf/10.1371_journal.pmed.0020124-L.pdf.

23. Victor M. Montori et al., "Randomized Trials Stopped Early for Benefit: A Systematic Review," *Journal of the American Medical Association* 294, no. 17 (November 2, 2005): 2208. "Such recommendations include the use of perioperative b-blockers in patients undergoing vascular surgery" made by the American College of Cardiologists and the American Heart Association. Montori et al., p. 2208. http://jama.ama-assn.org/cgi/content/abstract/294/17/2203.

24. E. M. Antman et al., "A Comparison of Results of Meta-Analyses of Randomized Control Trials and Recommendations of Clinical Experts; Treatments for Myocardial Infarction," *Journal of the American Medical Association* 268, no. 2 (July 8, 1992): 240, http://jama.ama-assn.org/cgi/content/abstract/268/2/240?ijkey=6f2406a3760805e095f6089be6d9d834830d8398&keytype2=tf_ipsecsha.

25. Two of the 10 quality measures in Medicare's Hospital Quality Initiative demonstration program are administration of beta-blockers at (1) arrival and (2) discharge for acute myocardial infarction (AMI). U.S. Centers for Medicare & Medicaid Services, "Hospital Quality Initiative Overview," August 2005, p. 2, http://www.cms.hhs.gov/quality/hospital/overview.pdf. Not all acute coronary syndromes are AMIs (http://www.infoplex.northwestern.edu/academic/clinical/medicine5.pdf). The example is offered, not as proof that beta-blockers harm certain patient subgroups, but to demonstrate the plausibility that aggregate benefits may conceal harm among subgroups.

26. David E. Lanfear et al., "b_2-Adrenergic Receptor Genotype and Survival among Patients Receiving b-Blocker Therapy after an Acute Coronary Syndrome," *Journal of the American Medical Association* 294 (September 28, 2005): 1532, http://jama.ama-assn.org/cgi/content/abstract/294/12/1526.

27. Cynthia M. Boyd et al., "Clinical Practice Guidelines and Quality of Care for Older Patients with Multiple Comorbid Diseases: Implications for Pay for Performance," *Journal of the American Medical Association* 294, no. 6 (August 10, 2005): 716–24, http://jama.ama-assn.org/cgi/content/abstract/294/6/716.

28. Ibid.

29. Mary E. Tinetti et al., "Potential Pitfalls of Disease-Specific Guidelines for Patients with Multiple Conditions," *New England Journal of Medicine* 351, no. 27 (December 30, 2004): 2870–74, http://content.nejm.org/cgi/content/extract/351/27/2870?hits=20&where=fulltext&andorexactfulltext=and&searchterm=bogardus&sortspec=Score%2Bdesc%2BPUBDATE_SORTDATE%2Bdesc&excludeflag=TWEEK_element&searchid=1&FIRSTINDEX=0&resourcetype=HWCIT.

30. Boyd et al.

31. Tinetti et al.

32. Ibid., p. 2827.

33. "Pay for performance will also address an inequity in the current payment system: paying the provider who gives his patients better care the same as the provider who does not." Mark E. Miller, executive director, Medicare Payment Advisory Commission,

"Pay for performance in Medicare," Testimony before the Senate Committee on Finance, July 27, 2005, p. 9, http://www.medpac.gov/publications/congressional_testimony/Testimony_P4P.pdf?CFID=2044521&CFTOKEN=4577 2147.

34. One possible answer to the challenge of outliers would be "insight bonuses." If a provider judges adherence to a P4P metric to be inappropriate for a particular patient, he would receive an "insight bonus" if his judgment proved accurate but no bonus (and possibly penalties) if his judgment proved inaccurate.

35. Friedrich A. Hayek, "The Use of Knowledge in Society," *American Economic Review* 25, no. 4 (September 1945): 519–30, http://www.econlib.org/library/Essays/hykKnw1.html.

36. Haynes.

37. Alan M. Garber, "Evidence-Based Guidelines as a Foundation for Performance Incentives," *Health Affairs* 24, no. 1 (January–February 2005): 176, http://content.healthaffairs.org/cgi/reprint/24/1/174.pdf.

38. Ibid.

39. Meredith B. Rosenthal et al., "Early Experience with Pay-for-Performance: From Concept to Practice," *Journal of the American Medical Association* 294, no. 14 (October 12, 2005): 1788–93, http://jama.ama-assn.org/cgi/content/short/294/14/1788.

40. U.S. Centers for Medicare & Medicaid Services, Centers for Medicare & Medicaid Services, "Medicare Demonstration Shows Hospital Quality of Care Improves with Payments Tied to Quality."

41. A survey of P4P experiments found only two randomized, controlled studies that examined relative performance measures. "The two studies in which the provider faced significant uncertainty about whether they [*sic*] could achieve success—in each case because the incentive was tied to performance relative to other groups, and this benchmark was unknown during the time when performance was measured—were negative." Dudley et al., p. 28.

42. Rosenthal et al.

43. Gary J. Young et al., "Conceptual Issues in the Design and Implementation of Pay-for-Quality Programs," *American Journal of Medical Quality* 20, no. 2 (April 2005): 1–7, http://ajm.sagepub.com/cgi/content/abstract/20/3/144.

44. Dudley et al., p. 11, citing A. Krasnik et al., "Changing Remuneration Systems: Effects on Activity in General Practice," *BMJ* 300, no. 6741 (1990): 1698–1701.

45. Wennberg, p. 2. Emphasis added.

46. Rosenthal et al.

47. Jim Molpus, "Pay for Performance: Is the Payoff Worth the Effort?" *HealthLeaders*, August 1, 2005, http://www.healthleaders.com/magazine/roundtable1.php?contentid=70748&issueId=91 (accessed December 4, 2005).

48. Garber, p. 179.

49. Ateev Mehrotra, Thomas Bodenheimer, and R. Adams Dudley, "Employers' Efforts to Measure and Improve Hospital Quality: Determinants of Success," *Health Affairs* 22, no. 2 (March–April 2003): 65, http://content.healthaffairs.org/cgi/reprint/22/2/60.pdf.

50. According to one private P4P administrator: "We are hearing complaints from clients about the demands of complying with different programs that are not using the same measures. Or even if they are using the same measures, the methodology and the application of the measurement sets differ." Molpus.

51. Cutler, p. 102.

52. The Medical Group Management Association, which represents physician group practices, has attacked a P4P scheme implemented by United Health: "'They allege that they are using established scientific measures of quality, but they've not been willing to say what they are or where they came from other than that they're in a piece of software that is proprietary,' says William F. Jessee, M.D., MGMA's president and CEO. 'That makes people suspicious.'" Philip Betbeze, "Pay for Performance Tipping Point," *HealthLeaders News*, September 15, 2005, http://www.healthleaders.com/news/feature1.php?contentid=72244 (accessed December 4, 2005).

53. With the exception of changes based on clinical evidence.

54. American Medical Association, "Health and Ethics Policies of the AMA House of Delegates: H-450.947 Pay-for-Performance Principles and Guidelines," pp. 695–99, http://www.ama-assn.org/ad-com/polfind/hlth-ethics.pdf. This list of conditions is not exhaustive.

55. Cutler, p. 110.

56. Boyd et al.

57. Cutler, pp. 108–9.

58. Sidney T. Bogardus Jr., David E. Geist, and Elizabeth H. Bradley, "Physicians' Interactions with Third-Party Payers: Is Deception Necessary?" *Archives*

of Internal Medicine 164 (September 27, 2004): 1841–44, http://archinte.ama-assn.org/cgi/content/abstract/164/17/1841?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&fulltext=Physicians%92+Interactions+with+Third-Party+Payers%3A+Is+Deception+Necessary&searchid=1133745842493_2000&stored_search=&FIRSTINDEX=0&journalcode=archinte.

59. "Tactics reported by physicians have included exaggerating the severity of the patient's condition, changing the patient's diagnosis for billing, or reporting signs or symptoms that the patient did not have. Deceptions may involve brief changes in wording, as when physicians rule out cancer as the indication for a test rather than screening. Also, physicians may be willing to alter billing codes or to change elements of patient history (e.g., increasing the severity of a symptom or even creating nonexistent symptoms, such as claiming suicidal ideation to obtain a psychiatric referral) or results of physical examination (e.g., inventing findings such as breast lumps to obtain a referral for screening mammography)." Bogardus et al., p. 1842.

60. Tinetti et al., p. 2872.

61. Wennberg, p. 2.

62. David M. Eddy, "Evidence-Based Medicine: A Unified Approach," *Health Affairs* 24, no. 1 (January–February 2005): 9–17, http://content. healthaffairs.org/cgi/content/abstract/24/1/9 (accessed December 4, 2005).

63. Haynes.

64. Dudley et al., p. 63.

65. Ibid., p. 28.

66. Ibid., p. 29.

67. Robert A. Berenson and Jane Horvath, "Confronting the Barriers to Chronic Care Management in Medicare," *Health Affairs* Web exclusive, January 22, 2003, W3–38, http://content.healthaffairs.org/cgi/reprint/hlthaff.w3.37v1.pdf.

68. "Current pay-for-performance initiatives can create financial incentives for physicians to focus on certain diseases and younger or healthier Medicare patients." Boyd et al., p. 722.

69. "Money in Politics Databases: Leading Sector Spending for Federal Lobbying (1/1/05-6/30/05)," Political Money Line, http://www.tray.com/cgi-win/lp_sector.exe?DoFn=my&Year=05, and previous reports dating back to "Money in Politics Databases: Leading Sector Spending for Federal Lobbying (1/1/99–6/30/99)," Political Money Line, http://www.tray.com/cgi-win/lp_sector.exe?DoFn=my&Year=99 (accessed December 21, 2005).

70. See generally *The 2005 Annual Report of the Board of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds* (Washington: Government Printing Office, March 23, 2005), http://www.cms.hhs.gov/publications/trusteesreport/tr2005.pdf.

71. See, e.g., Milton Friedman, *Capitalism and Freedom* (Chicago: University of Chicago Press, 2002), p. 143.

72. Medicare Payment Advisory Commission, "Payment Basics: Physician Services Payment System," December 9, 2005, p. 1, http://www.medpac.gov/publications/other_reports/Dec05_payment_basics_physician.pdf.

73. See, e.g., U.S. Federal Trade Commission and U.S. Department of Justice, p. 9.

74. "Because of strict statutory constraints and its own burdensome regulatory and administrative procedures, [CMS] is slow to address overpricing and overutilization problems." Janet L. Shikles, U.S. General Accounting Office, "Medicare: Private Payer Strategies Suggest Options to Reduce Rapid Spending Growth," Testimony before the Subcommittee on Health of the House Committee on Ways and Means, April 30, 1996, GAO/T-HEHS-96-138, http://www.gao.gov/archive/1996/he96138t.pdf.

75. U.S. Federal Trade Commission and U.S. Department of Justice, chap. 3, pp. 25–26.

76. Uwe Reinhardt, "The Medicare World from Both Sides: A Conversation with Tom Scully," *Health Affairs* 22, no. 6 (November–December, 2003): 168–74, http://content.healthaffairs.org/cgi/reprint/22/6/167.pdf.

77. Former CMS administrator Tom Scully, quoted in Reinhardt, pp. 169–70.

78. See National Bipartisan Commission on the Future of Medicare, "Building a Better Medicare for Today and Tomorrow," March 16, 1999, http://thomas.loc.gov/medicare/bbmtt31599.html.

Cato Institute

Founded in 1977, the Cato Institute is a public policy research foundation dedicated to broadening the parameters of policy debate to allow consideration of more options that are consistent with the traditional American principles of limited government, individual liberty, and peace. To that end, the Institute strives to achieve greater involvement of the intelligent, concerned lay public in questions of policy and the proper role of government.

The Institute is named for *Cato's Letters*, libertarian pamphlets that were widely read in the American Colonies in the early 18th century and played a major role in laying the philosophical foundation for the American Revolution.

Despite the achievement of the nation's Founders, today virtually no aspect of life is free from government encroachment. A pervasive intolerance for individual rights is shown by government's arbitrary intrusions into private economic transactions and its disregard for civil liberties.

To counter that trend, the Cato Institute undertakes an extensive publications program that addresses the complete spectrum of policy issues. Books, monographs, and shorter studies are commissioned to examine the federal budget, Social Security, regulation, military spending, international trade, and myriad other issues. Major policy conferences are held throughout the year, from which papers are published thrice yearly in the *Cato Journal*. The Institute also publishes the quarterly magazine *Regulation*.

In order to maintain its independence, the Cato Institute accepts no government funding. Contributions are received from foundations, corporations, and individuals, and other revenue is generated from the sale of publications. The Institute is a nonprofit, tax-exempt, educational foundation under Section 501(c)3 of the Internal Revenue Code.

CATO INSTITUTE
1000 Massachusetts Ave., N.W.
Washington, D.C. 20001
www.cato.org