




2007

Pay-for-Performance: Is Medicare a Good Candidate?

Michael F. Cannon

Follow this and additional works at: <https://digitalcommons.law.yale.edu/yjhple>

 Part of the [Health Law and Policy Commons](#), and the [Legal Ethics and Professional Responsibility Commons](#)

Recommended Citation

Michael F. Cannon, *Pay-for-Performance: Is Medicare a Good Candidate?*, 7 *YALE J. HEALTH POL'Y L. & ETHICS* (2007).

Available at: <https://digitalcommons.law.yale.edu/yjhple/vol7/iss1/1>

This Article is brought to you for free and open access by Yale Law School Legal Scholarship Repository. It has been accepted for inclusion in *Yale Journal of Health Policy, Law, and Ethics* by an authorized editor of Yale Law School Legal Scholarship Repository. For more information, please contact julian.aiken@yale.edu.

ARTICLES

Pay-for-Performance: Is Medicare a Good Candidate?

Michael F. Cannon*

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% of the time.¹ The likelihood that patients will receive recommended care “varie[s] 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.”² Evidence of low-quality care appears in Medicare, the federal health program for the elderly and disabled.³ Quality of care does not appear to be higher in areas where Medicare spending is higher.⁴ In fact, some studies point to the somewhat paradoxical conclusion that

* Michael F. Cannon is the Cato Institute’s Director of Health Policy Studies. The author wishes to thank Susan Dudley of the Mercatus Center at George Mason University and Peter Van Doren of the Cato Institute for many helpful comments.

1. Elizabeth A. McGlynn et al., *The Quality of Health Care Delivered to Adults in the United States*, 348 NEW ENG. J. MED. 2635 (2003). See also Steven M. Asch et al., *Who Is at Greatest Risk for Receiving Poor-Quality Health Care?*, 354 NEW ENG. J. MED. 1147, 1147 (2006) (“Overall, participants received 54.9 percent of recommended care. . . . [T]here was only moderate variation in quality-of-care scores among sociodemographic subgroups.”).

2. McGlynn et al., *supra* note 1, at 2635.

3. See Elliott S. Fisher et al., *The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care*, 138 ANNALS INTERNAL MED. 273, 283 (2003) (reporting that the percentage of patients from select cohorts who received recommended care ranged from 19.7% to 87.7%) [hereinafter Fisher et al., *Part 1*]. See also Asch et al., *supra* note 1, at 1150; Priscilla Hollander et al., *Quality of Care of Medicare Patients with Diabetes in a Metropolitan Fee-for-Service Primary Care Integrated Delivery System*, 20 AM. J. MED. QUALITY 344 (2005).

4. Fisher et al., *Part 1, supra* note 3, at 273. Patients in high-spending regions received sixty percent more care, but those higher Medicare expenditures did not translate into higher-quality care. “The increased utilization was explained by more frequent physician visits, especially in the inpatient setting[.]. . . more frequent tests and minor (but not major) procedures, and increased use

Medicare patients are often less likely to receive recommended care in regions where Medicare expenditures are highest.⁵

Third-party payment is a potential contributor to the under-provision of quality health care. Most health care payments in the United States are made by third parties, usually employers, insurers, or government.⁶ 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.⁷ The result is a financing system akin to paying academics on the basis of the volume and intensity of footnotes.⁸

Medicare is the most obvious 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.”⁹ 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.¹⁰

of specialists and hospitals.” *Id.* Nor did higher spending translate into decreased mortality, better functional status, or higher patient satisfaction. Elliott S. Fisher et al., *The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care*, 138 ANNALS INTERNAL MED. 288, 288 (2003).

5. Katherine Baicker & Amitabh Chandra, Medicare Spending, *The Physician Workforce, And Beneficiaries’ Quality of Care*, W4 Health Aff. - Web Exclusive, April 7, 2004, at W4-184, W4-192, <http://content.healthaffairs.org/cgi/reprint/hlthaff.w4.184v1.pdf> (“States that spend more per Medicare beneficiary are not states that provide higher quality care. In fact, additional spending is positively correlated with end-of-life care but negatively correlated with the use of effective care.”). See also Fisher et al., *Part 1, supra* note 3, at 273 (“Quality of care in higher-spending regions was no better on most measures and was worse for several preventive care measures.”); *id.* at 283 (reporting that for seven out of ten types of recommended care, Medicare spending within a hospital referral region was inversely related to the quality of care).

6. See, e.g., Ctrs. for Medicare & Medicaid Servs., Office of the Actuary, Nat’l Health Statistics Group, National Health Expenditures Web Tables 8, <http://www.cms.hhs.gov/NationalHealthExpendData/downloads/tables.pdf> (last visited Dec. 10, 2006).

7. See, e.g., U.S. Fed. Trade Comm’n & U.S. Dep’t of Justice, *Improving Health Care: A Dose of Competition* 5, 13 (Executive Summary) (2004), available at <http://www.ftc.gov/reports/healthcare/040723healthcarerpt.pdf>; DAVID M. CUTLER, *YOUR MONEY OR YOUR LIFE* 101 (2004).

8. Such a payment system may or may not improve the overall quality of analysis, but assuredly it would result in overuse of low-quality footnotes.

9. U.S. Fed. Trade Comm’n & U.S. Dep’t of Justice, *supra* note 7, ch. 2, at 30.

10. MEDICARE PAYMENT ADVISORY COMM’N, REPORT TO THE CONGRESS: VARIATION AND INNOVATION IN MEDICARE 108 (2003), available at <http://www.medpac.gov/publications/>

Medicare's massive influence on the health care system has made it a major focus of efforts to solve the problems caused by quality-blind third-party purchasing.

Reform efforts have led third-party payors to experiment with financial incentives that encourage physicians and hospitals to provide recommended care. Such initiatives are termed "quality-based purchasing" (QBP) or "pay-for-performance" (P4P). P4P is an outgrowth of the "evidence-based medicine" (EBM) movement, which argues that providers too often rely on their own judgment, because scientific evidence on the effectiveness of medical interventions is either unavailable or ignored.¹¹ 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."¹² 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 such as Aetna, PacifiCare, and WellPoint have been leaders in the field; Highmark Blue Cross Blue Shield has experimented with P4P since 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.¹³

congressional_reports/June03_Entire_Report.pdf. See also Reed Abelson, *Medicare Says Bonuses Can Improve Hospital Care*, N.Y. TIMES, Nov. 15, 2005, at C3 ("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."); John E. Wennberg, *Variation in Use of Medicare Services Among Regions and Selected Academic Medical Centers: Is More Better?*, Duncan W. Clark Lecture at New York Academy of Medicine 2 (Jan. 24, 2005), available at http://www.dartmouthatlas.org/atlas/nyam_lecture.pdf (discussing under-use of high-quality care and overuse of low-quality care in Medicare).

11. See Evidence-Based Med. Working Group, *Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine*, 268 JAMA 2420 (1992); R. Brian Haynes, *What Kind of Evidence Is It That Evidence-Based Medicine Advocates Want Health Care Providers and Consumers to Pay Attention to?*, 2 BMC HEALTH SERVICES RES. 1 (2002), available at <http://www.biomedcentral.com/content/pdf/1472-6963-2-3.pdf>; Jack Hitt, *Evidence-Based Medicine*, N.Y. TIMES MAG., Dec. 9, 2001, at 68.

12. CUTLER, *supra* note 7, at 101.

13. *Medicare Value-Based Purchasing for Physicians' Services Act of 2005: Hearing on H.R. 3617 Before the Subcomm. on Health of the H. Comm. on Ways and Means*, 109th Cong. (2005) (statement of Karen Ignagni, President and Chief Executive Officer, America's Health Insurance Plans), available at <http://waysandmeans.house.gov/hearings.asp?formmode=view&id=3820> [hereinafter *Medicare Hearing*].

Medicare currently has ten demonstration programs underway that 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.¹⁴

Provider-focused financial incentives for high-quality care have the potential to improve quality in many instances. However, caution is in order. Creating a P4P program – where third-party purchasers create financial incentives for providers to deliver quality care – is an immensely difficult task. Quality has multiple dimensions and is often highly subjective, making “quality care” impossible to define uniformly for diverse populations. Even when it is possible to settle on a reasonable definition of quality, measures can be difficult to translate into financial incentives.

These difficulties suggest two approaches that would maximize the potential of P4P while minimizing any 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. Competition to improve the quality of care in a cost-effective manner encourages private purchasers to experiment with P4P, and private control gives purchasers flexibility in designing and altering those experiments. As important, private P4P experiments confine any harmful failures to smaller populations. As discussed below, the politics of Medicare all but guarantees that any potential harm resulting from a P4P scheme would, in the context of Medicare, be more likely to occur, harm more patients, and take longer to correct. Therefore, 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. Congress should resist the temptation to expand P4P into traditional Medicare.

Second, employers and insurers should experiment not only with provider-focused financial incentives but with patient-focused financial incentives as well. For example, private insurers “are offering consumers reduced co-payments, deductibles, and/or premiums in exchange for using providers deemed to be of higher quality, based on specific performance measures.”¹⁵ 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 patients and their doctors to deviate from “best practices” if doing so fits the patient’s needs.

This Article proceeds as follows. Part I discusses the practical difficulties

14. Press Release, Ctrs. for Medicare & Medicaid Servs., Medicare “Pay for Performance (P4P)” Initiatives (Jan. 31, 2005), available at <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=1343>.

15. *Medicare Hearing*, *supra* note 13.

PAY-FOR-PERFORMANCE

involved in designing a system that enables third-party purchasers to reward providers for high-quality medical care, particularly given the absence of research on the cost-effectiveness of such efforts. Part II considers how health care providers might respond to a P4P scheme, and how those responses might affect the cost and quality of medical care and insurance. Part III examines what the evidence says about whether P4P improves quality. Part IV begins with a brief sketch of Medicare's experiments with P4P and efforts to expand P4P within that program. Part IV then argues that the pitfalls of P4P are more likely to occur in traditional Medicare, and offers proposals on how to bring the benefits of P4P to seniors while minimizing their exposure to harm.

I. PITFALLS OF THIRD-PARTY P4P

Identifying and rewarding quality are difficult tasks for any third-party purchaser. As one study of P4P measures notes, “[e]xperience in other industries has shown that developing performance measures for complex phenomena is difficult and that inappropriate measures can have significant negative consequences.”¹⁶ 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 Part outlines the challenges faced by third-party purchasers, whether public or private, when attempting to improve quality through provider-focused financial incentives.

A. 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: patient outcomes, processes, structural factors, and patient satisfaction.¹⁷ 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.

Patient outcomes are the most obvious measure of health care quality. For example, outcome measures for heart attack patients could include patients' post-

16. R. ADAMS DUDLEY ET AL., U.S. DEP'T OF HEALTH & HUMAN SERVS., STRATEGIES TO SUPPORT QUALITY-BASED PURCHASING 68 (2004), available at <http://www.ahrq.gov/downloads/pub/evidence/pdf/qbpurch/qbpurch.pdf> (citing C. D. Ittner & D. F. Larcker, *Coming Up Short on Nonfinancial Performance Measurement*, 81 HARV. BUS. REV. 88, 88-95, 139 (2003)).

17. For example, a PacifiCare “quality incentive program” rewards physicians based on measures of patient satisfaction, as well as process measures such as “rates of cervical cancer screening, mammography, and hemoglobin . . . testing for diabetic patients.” Meredith B. Rosenthal et al., *Early Experience with Pay-for-Performance: From Concept to Practice*, 294 JAMA 1788, 1789 (2005).

intervention cholesterol levels, readmission rates, or mortality rates. However, outcome measures have limitations. First, patients may differ in their desired outcomes.¹⁸ Second, patient outcomes can be influenced by factors other than the medical intervention. For example, 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 providers' performance. As a result, providers are understandably reluctant to be rewarded or penalized 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."¹⁹ 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,"²⁰ 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 of heart attack patients, a process measure would reward providers based on how often they check cholesterol levels or prescribe beta-blockers for those patients. Process measures are the most often discussed type of P4P quality measure;²¹ thus their potential shortcomings will be discussed in more detail throughout Sections I.B and I.C 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."²² Examples include

18. See generally Avedis Donabedian, *Evaluating the Quality of Medical Care*, 83 MILBANK Q. 691, 694 (2005), reprinted from 44 MILBANK MEMORIAL FUND Q. 166 (1966), available at <http://www.milbank.org/quarterly/830416donabedian.pdf> (indicating that, 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").

19. *Id.*

20. U.S. Agency for Health Care Research & Quality, AHRQ Patient Safety Network Glossary: Structure-Process-Outcome Triad, <http://psnet.ahrq.gov/glossary.aspx> (last visited Dec. 10, 2006).

21. See generally DUDLEY ET AL., *supra* note 16, at 9, 69-70 (providing a survey of the literature on P4P experiments that finds more efforts based on process measures than on structural or outcomes measures).

22. Donabedian, *supra* note 18, at 695.

PAY-FOR-PERFORMANCE

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 costs and may undercut cost-effectiveness.

Table 1: Tradeoffs Presented by Measuring and Rewarding Different Dimensions of Quality

| <i>Quality Measure</i> | <i>Upside</i> | <i>Downside</i> |
|--|---|---|
| Patient outcomes | Captures patient health | Desired outcomes vary across patients Factors beyond control of providers affect outcomes Does not reveal how positive outcomes were achieved Providers can game outcome measures by selecting healthier patients Requires case-mix adjustment to demonstrate improved health Measurement lags |
| Processes | Captures provider actions that promote health | Can encourage inappropriate care for outliers Providers can game process measures through patient selection, data manipulation, etc. |
| Structural | Captures whether providers use human/physical capital known to improve health/convenience | Does not measure whether capital is used optimally Can require large investments by providers |
| Patient satisfaction | Measures whether providers meet patient expectations Captures intangible/subjective aspects of quality | Poor performers may score well if patients are ignorant of higher-quality options |
| Incorporating multiple types of quality measures | Captures benefits of each measure used | Adds complexity and cost Can discourage physician compliance |

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 present shortcomings because patients are not necessarily in the best position to evaluate the quality of care. For example, providers who deliver low-quality care may score well on patient satisfaction measures if patients are unaware that higher-quality care is available.

B. Collecting Reliable Data

A third-party payor'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 and applying accurate data. These include finding quality data that relate various inputs to outcomes, translating data into performance measures, making allowances for atypical patients, and targeting, calibrating, and continually adjusting performance measures and financial incentives in the face of uncertainty about the reliability of new findings.

1. Availability of Data

The success of pay-for-performance (also referred to as "quality-based purchasing," or 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.²³

At present, it is difficult or impossible to know for what share of health care expenditures useful data exist.²⁴ Where data are not available, third-party

23. DUDLEY ET AL., *supra* note 16, at 7.

24. The Centers for Medicare & Medicaid services provide this slippery description of the availability of such data:

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

purchasers have little ability to use financial incentives to drive quality improvements.

2. *Quality of Data*

Where data are available, purchasers must consider whether the data employed show a true relationship between a metric and a desired outcome. Even 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.”²⁵ 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 interventions²⁶ and that “most current published research findings are false.”²⁷

Concerns even exist 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, some researchers question whether clinical trials are too often “stopped early for benefit” – that is, when preliminary results appear positive and convincing. One study notes that “[p]rofessional organizations continue to issue recommendations on the basis of trials stopped early for benefit, including those . . . that seem most likely to overestimate effects.”²⁸ One expert notes that CPGs “have been reported to be variably flawed

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.

Value-Based Purchasing for Physicians Under Medicare: Hearing Before the Subcomm. on Health of the H. Comm. on Ways and Means, 109th Cong. (2005) (statement of Mark B. McClellan, Dep’t of Health & Human Servs.) (emphasis added), available at <http://waysandmeans.house.gov/hearings.asp?formmode=view&id=4678>.

25. Haynes, *supra* note 11, at 1.

26. See, e.g., John P. A. Ioannidis, *Contradicted and Initially Stronger Effects in Highly Cited Clinical Research*, 294 JAMA 218 (2005).

27. John P. A. Ioannidis, *Why Most Published Research Findings Are False*, 2 PUB. LIBR. SCI. MED. 696, 696 (2005).

28. Victor M. Montori et al., *Randomized Trials Stopped Early for Benefit: A Systematic Review*, 294 JAMA 2203, 2208 (2005) (“Such recommendations include the use of perioperative β-

in terms of conflict of interest, specialty turf battles, endorsement of new or relatively unproven pharmaceutical agents, and focus on a single condition compared with a broader clinical focus.²⁹

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.³⁰

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 do not 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 providers' perceptions of the reliability of the data, will influence how providers respond to the financial incentives and thus the effectiveness of those incentives.

3. Outliers

Even when high-quality, timely data are available, translating those findings into performance measures is complicated by 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 it does for most patients. P4P schemes that encourage providers to treat outliers like the average patient can therefore create perverse incentives that encourage low-quality care. Because

blockers in patients undergoing vascular surgery” made by the American College of Cardiologists and the American Heart Association.).

29. Patrick J. O'Connor, *Adding Value to Evidence-Based Clinical Guidelines*, 294 JAMA 741, 741 (2005).

30. Elliott M. Antman et al., *A Comparison of Results of Meta-Analyses of Randomized Control Trials and Recommendations of Clinical Experts: Treatments for Myocardial Infarction*, 268 JAMA 240, 240 (1992).

PAY-FOR-PERFORMANCE

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 across the thousands of patients within the trial, much less across the millions of patients in the general population.³¹ A treatment's overall beneficial effects may hide different effects on subgroups, including no effect or even harmful effects. For example, patients may respond differently to a given intervention as a result of multiple illnesses or interactions with treatment regimens for such co-morbidities.³² 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.

For example, the administration of beta-blockers to patients with cardiovascular disease is a common quality measure.³³ However, a recent study found that among acute coronary syndrome patients prescribed beta-blockers, certain genotype groups had lower rates of survival. The authors caution, "[f]urther studies of the efficacy of β -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."³⁴ One could infer from this study that compliance with a widely used quality measure could actually increase mortality

31. As O'Connor has stated:

[A]ll evidence-based recommendations are not of equal clinical benefit to a patient. Benefits documented in clinical trials are 'average' benefits and even within the trials the degree of benefit received from an intervention depends on many patient-specific factors. Practicing physicians care for patients with even greater patient-specific variation (because of restrictive eligibility criteria in most clinical trials), so it is not surprising to find wide variation in the benefits obtained. When treating elderly patients with multiple comorbid conditions, the complexity of care is compounded by the need to simultaneously address multiple clinical domains.

O'Connor, *supra* note 29, at 742.

32. A co-morbidity is a coexistent but unrelated disease or disorder.

33. See, e.g., CTRS. FOR MEDICARE & MEDICAID SERVS., HOSPITAL QUALITY INITIATIVE OVERVIEW 2 (2005), available at <http://www.cms.hhs.gov/HospitalQualityInits/downloads/HospitalOverview200512.pdf> (noting that two of the ten quality measures in the Hospital Quality Initiative demonstration program are administration of beta-blockers at (1) arrival and (2) discharge for acute myocardial infarction (AMI)); PREMIER, INC., CENTERS FOR MEDICARE AND MEDICAID SERVICES (CMS)/PREMIER HOSPITAL QUALITY INCENTIVE DEMONSTRATION PROJECT: PROJECT OVERVIEW AND FINDINGS FROM YEAR ONE 6 (2006), available at <http://www.premierinc.com/quality-safety/tools-services/p4p/hqi/hqi-whitepaper041306.pdf> (same).

34. David E. Lanfear et al., *β_2 -Adrenergic Receptor Genotype and Survival Among Patients Receiving β -Blocker Therapy After an Acute Coronary Syndrome*, 294 JAMA 1526, 1532 (2005).

among some patients.³⁵

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.³⁶ However, many CPGs lack guidance specific to the elderly and patients with co-morbidities. One study examining leading CPGs for nine chronic illnesses found that only four of the nine “addressed older individuals with multiple comorbidities.”³⁷

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.³⁸ 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 a provider who complies with P4P guidelines for treating each of a patient’s chronic illnesses would deliver lower-quality care relative to a provider who makes more individualized prescribing decisions.³⁹ The prospect of suffering financial penalties for providing individualized care to such patients could discourage providers from caring for patients with co-morbidities altogether.⁴⁰

35. Not all acute coronary syndromes are AMIs. 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.

36. See Cynthia M. Boyd et al., *Clinical Practice Guidelines and Quality of Care for Older Patients with Multiple Comorbid Diseases: Implications for Pay for Performance*, 294 JAMA 716, 716 (2005).

37. *Id.* at 718.

38. See *id.* at 716.

39. Mary E. Tinetti et al., *Potential Pitfalls of Disease-Specific Guidelines for Patients with Multiple Conditions*, 315 NEW ENG. J. MED. 2870, 2872 (2004) (“Individual medications that impart disease-specific benefits may be less beneficial, or even harmful, when taken along with other medications by patients with multiple coexisting conditions and variable health outcomes.”). See also Boyd et al., *supra* note 36, at 716 (“Basing standards for quality of care and pay for performance on existing CPGs could lead to inappropriate judgment of the care provided to older individuals with complex comorbidities and could create perverse incentives that . . . diminish the quality of their care.”).

40. Boyd et al., *supra* note 36, at 722. The author is aware of no definitive evidence that providers have gamed P4P schemes by avoiding particular patients. However, research suggests that providers do frequently game payment systems. See, e.g., Leemore S. Dafny, *How Do Hospitals Respond to Price Changes?*, 95 AM. ECON. REV. 1525, 1545 (2005) (noting that hospitals are “quite sophisticated” in their ability to increase reimbursements by gaming changes in Medicare payments); Matthew K. Wynia et al., *Physician Manipulation of Reimbursement Rules for Patients: Between a Rock and a Hard Place*, 283 JAMA 1858 (2000) (discussing physicians manipulating reimbursement rules). Results of a P4P program in Britain’s National Health Service raised concerns about gaming. Tim Doran et al., *Pay-for-Performance Programs in Family Practices in the United Kingdom*, 355 NEW ENG. J. MED. 375, 383 (2006) (suggesting providers may have improved their scores on performance measures by arbitrarily excluding from those measurements

PAY-FOR-PERFORMANCE

Other patients deviate from the mean in their preferences for particular health outcomes.⁴¹ Older patients and those with numerous health problems often have treatment goals that conflict with P4P measures:

Is a statin or 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?⁴²

For reasons of practicality, a P4P scheme might take account of easily measurable 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 would be penalized for treating patients according to their preferences.

For all of these reasons, outliers raise issues of equity between physicians compensated according to a P4P framework. Although P4P schemes are meant to correct some of the inequities of existing quality-blind payment systems,⁴³ they can create similar inequities by penalizing physicians who provide quality care by correctly treating a patient as an outlier.

The existence of outliers points to the limited usefulness of aggregate data in promoting quality.⁴⁴ Much medical practice relies on the use of “unorganized knowledge”⁴⁵ about the circumstances of each patient and her preferences. As

patients who would have lowered the providers’ scores); Laura A. Peterson et al., *Does Pay-for-Performance Improve the Quality of Health Care?*, 145 ANNALS INTERNAL MED. 265, 268 (2006) (citing four studies that found evidence of “gaming behavior”).

41. See Tinetti et al., *supra* note 39, at 2870.

42. *Id.* at 2872. See generally U.S. Food & Drug Admin., U.S. Dep’t of Health & Human Servs., *Updates: New Drug to Lower Cholesterol*, 37 FDA CONSUMER MAG. 6 (2003), available at http://www.fda.gov/fdac/departs/2003/603_upd.html (“In rare instances, severe muscle pain and muscle weakness resulting in kidney damage have been associated with statin drugs.”); U.S. Food & Drug Admin., U.S. Dep’t of Health & Human Servs., *High Blood Pressure Medicines to Help You*, <http://www.fda.gov/womens/MedicineCharts/highbloodpressure.pdf> (indicating that “[c]ommon [s]ide [e]ffects” of beta-blockers include “[f]eeling tired . . . [d]izziness . . . [f]eeling lightheaded.”)

43. See *Pay for Performance in Medicare: Hearing Before the S. Comm. on Finance*, 109th Cong. 9 (2005) (statement of Mark E. Miller, Executive Director, Medicare Payment Advisory Commission), available at http://www.medpac.gov/publications/congressional_testimony/Testimony_P4P.pdf (“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.”).

44. Though theoretically possible to gather quality data specific to outliers, accounting for all the possible combinations of physiological differences, comorbidities, and varied preferences of patients quickly yields a staggering number of outlier categories. Beyond a certain point, collecting quality data for outliers becomes cost-ineffective, and ultimately statistically prohibitive (as it becomes difficult to find enough subjects to conduct clinical trials).

45. F. A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519 (1945). In a

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.”⁴⁶ 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.

C. Tradeoffs

Beyond the challenges involved in defining quality and collecting useful data, third-party purchasers face a third 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 about the optimal target of incentives, the most effective types of performance targets, and the size of financial incentives. Poorly calibrated incentives can result in no effect, higher expenditures, inequities, reduced access to care, or even low-quality and inappropriate care.

1. 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 over-prescribe, under-prescribe, or unnecessarily compartmentalize care. For most performance measures, the question is resolved if the patient receives care from an integrated health care system responsible for all aspects of treatment. In those cases, the incentive would be targeted at the institution, such as Kaiser Permanente or the Veterans Health Administration.⁴⁷

However, patients typically receive care in non-integrated settings. According to Stanford University’s Alan Garber:

[T]he conceptual basis for assigning responsibility is unclear when a patient is treated by multiple physicians, some of whom the patient selects without the concurrence or even knowledge of the others. 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

medical context, “unorganized knowledge” would include the particular financial and physical circumstances of individual patients and each patient’s preferences for specific health outcomes, such as longevity, alertness, strength, etc. Physicians, who interact with individual patients, are in a better position to collect and use this knowledge than third-party purchasers who are farther removed from the patient.

46. Haynes, *supra* note 11, at 4.

47. See Alan M. Garber, *Evidence-Based Guidelines as a Foundation for Performance Incentives*, 24 HEALTH AFF. 174, 175 (2005).

PAY-FOR-PERFORMANCE

monitoring of disease complications and the side effects of treatment.⁴⁸

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 over-prescribing and even less coordination of care. Holding only one responsible (say, the cardiologist) could also lead to over-prescribing but also could lead to unnecessary compartmentalization of certain aspects of care (e.g., only cardiologists prescribing ACE inhibitors).

Table 2: Tradeoffs Presented by Different Targets of Financial Incentives

| <i>Target of Financial Incentives</i> | <i>Upside</i> | <i>Downside</i> |
|---|---|---|
| Individual physicians/ provider groups | Affects behavior of individual physicians/ groups Allows purchasers to evaluate individual physicians/groups | Difficult to assign responsibility for outcomes or compliance with process measures Can encourage duplicative efforts, poor coordination of care |
| Integrated health care system | Avoids problems of duplication, poor coordination of care | Many patients do not receive care in integrated settings |

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?”⁴⁹

2. Selecting 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.

48. *Id.* at 176.

49. *Id.*

Each option presents tradeoffs that will affect providers' responses 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 rewards all providers if they prescribe beta-blockers to ninety 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, seventy-five percent of bonuses went to providers who were already exceeding the performance target.⁵⁰ 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."⁵¹ 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 ten percent of providers who have the highest rates of prescribing beta-blockers to AMI patients in a given year. For instance, a Medicare P4P demonstration program (discussed in Part IV) awards bonuses to hospitals in the top two deciles in each of a number of metrics.⁵² A relative performance target is a moving target that depends on the behavior of other providers. Because the threshold for compliance 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.⁵³

50. Rosenthal et al., *supra* note 17, at 1792.

51. *Id.* at 1788.

52. Press Release, Ctrs. for Medicare & Medicaid Servs., Medicare Demonstration Shows Hospital Quality of Care Improves with Payments Tied to Quality (Nov. 14, 2005), available at <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=1729> [hereinafter Medicare Demonstration].

53. See, e.g., DUDLEY ET AL., *supra* note 16, at 28. This survey of P4P experiments (discussed further in Part III) found only two randomized, controlled studies that examined relative performance measures. "The two studies in which the provider faced significant uncertainty about whether they 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." *Id.*

PAY-FOR-PERFORMANCE

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 may motivate even poorly performing providers to improve. Such incentives, for example, could reward providers for every ten percentage point improvement on a given metric. But such a performance target would do little to encourage improvement among providers in the top decile. Whether increments of improvement are judged along an absolute scale or a relative scale, providers already above the ninetieth 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.

Table 3: Tradeoffs Presented by Different Types of Performance Targets

| <i>Type of Performance Target</i> | <i>Upside</i> | <i>Downside</i> |
|--|---|--|
| Absolute achievement | Clear expectations reduce uncertainty Allow providers to plan | Cost-ineffective; most bonuses go to already-high performers No incentive to improve beyond the upper-most target Can discourage improvement among poor performers |
| Relative performance | Can increase competition among high performers | Less certainty that compliance efforts will be rewarded Can discourage compliance among poor performers |
| Improvement | Encourage low-performers to improve Targeting absolute improvement reduces uncertainty | Already-high performers have less room for improvement Poor performers could receive larger bonuses than high performers |
| Combining two or more types of performance targets | Encourages compliance among all providers | Adds complexity and cost Poor performers could receive larger bonuses than high performers |

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, and separate rewards for improvement. However, including multiple performance targets and rewards adds complexity and gives all providers an opportunity to increase their incomes, creating affordability problems. One solution to the affordability problem is 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.

3. Sizing and Timing 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.⁵⁴ 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 different providers will respond to the same incentive in different ways.

Some observers suggest that financial incentives must account for at least ten percent of a physician's income.⁵⁵ 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 a \$100,000 electronic patient registry. If that bonus plus other benefits of implementing the registry do not at least match the cost of implementing the registry, the incentive would have no effect on the provider's behavior. To cause this provider to change his behavior, the bonus (plus other benefits) 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. Yet a survey of randomized controlled trials of P4P schemes found that none reported data on the cost of complying with the performance measures.⁵⁶

54. See, Rosenthal et al., *supra* note 17, at 1792-93.

55. Gary J. Young et al., *Conceptual Issues in the Design and Implementation of Pay-for-Quality Programs*, 20 AM. J. MED. QUALITY 144, 146 (2005).

56. DUDLEY ET AL., *supra* note 16, at 23.

PAY-FOR-PERFORMANCE

Table 4: Tradeoffs – Size of Financial Incentives

| <i>Size of Financial Incentives</i> | <i>Upside</i> | <i>Downside</i> |
|-------------------------------------|---|--|
| Modest | Makes P4P scheme affordable Encourages compliance among providers who can do so at lowest cost | May fail to induce compliance Rewards may go to already-high performers |
| Large | More likely to induce broad compliance | Makes P4P scheme less affordable May be cost-ineffective |

Another factor complicating the calibration of financial incentives is providers’ income goals. As R. Adams Dudley, a professor of medicine at the University of California San Francisco and a leading researcher of P4P efforts, notes, “[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.”⁵⁷

How a P4P scheme treats compliant versus noncompliant providers will affect both quality and costs. One way to encourage greater compliance is to ensure that the disincentives 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 could be held harmless but paid less than compliant providers. A more controversial option is to reduce payments to noncompliant providers. Financial penalties can improve overall affordability, but at the cost of provider resistance (as discussed in Section II.B). Nonetheless, without financial penalties, P4P can only increase health expenditures.

Table 5: Tradeoffs – Rewards and Penalties

| <i>Type of Financial Incentives</i> | <i>Upside</i> | <i>Downside</i> |
|-------------------------------------|--|--|
| Rewards-only | Less provider resistance Rewards high performers Encourages poor performers to invest in improvement | Increases costs Holds poor performers harmless Creates incentives to game system |

57. *Id.* at 11 (citing Allan Krasnik et al., *Changing Remuneration Systems: Effects on Activity in General Practice*, 300 BRIT. MED. J. 1698 (1990)).

| | | |
|-----------------------|---|--|
| Rewards and penalties | Penalizes poor performers Rewards high performers Encourages poor performers to invest in improvement | Increases likelihood of provider resistance Creates incentives to game system |
|-----------------------|---|--|

Other important factors are 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.⁵⁸ Likewise, third-party payors 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 net value, of those rewards.

Table 6: Tradeoffs – Updating P4P Features

| <i>Frequency of Updates</i> | <i>Upside</i> | <i>Downside</i> |
|-----------------------------|--|--|
| Frequent | Makes use of latest scientific evidence Allows cost-effective calibration of incentives | Increases compliance costs for providers Encourages provider resistance |
| Infrequent | Predictability allows providers to plan | Rewards suboptimal care Reduces cost-effectiveness |

D. Cost-Effectiveness

Even if P4P delivered demonstrable improvements in health care quality, that would not demonstrate that P4P is worthwhile. Designing a P4P program that is cost-effective is also a significant challenge. Despite significant health gains, many or all P4P designs could impose costs that outweigh those gains. 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.⁵⁹

Another important cost dimension is the hidden costs that P4P might impose

58. See Rosenthal et al., *supra* note 17, at 1789.

59. *E.g., id.* at 1788 (noting that a “prototypical physician pay-for-performance program” made bonus payments of \$3.4 million over a one-year period); Doran et al., *supra* note 40 at 376, 383 (“In 2004, the National Health Service committed £1.8 billion (\$3.2 billion) in additional funding over a period of three years for a new pay-for-performance program for family practitioners.”).

PAY-FOR-PERFORMANCE

by encouraging inappropriate care or reducing access. 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 non-elderly to afford coverage. The health gains that a P4P scheme “purchases” could also be outweighed by the health losses that result from encouraging inappropriate outlier care. The costs of P4P have yet to be quantified, much less compared to the potential benefits.⁶⁰

II. HOW WILL PROVIDERS RESPOND?

Quality-based purchasing is designed to affect the behavior of providers for the benefit of patients. The problems posed by P4P in a health care setting are similar to those of other principal-agent settings, where the principal (here, purchasers) face difficulties creating financial incentives that encourage their agents (providers) to behave in the desired manner.⁶¹

Health care providers are highly suspicious of P4P efforts,⁶² 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, increase costs, and even leave some patients worse off. This Part discusses the factors that will affect providers’ receptivity to P4P schemes, what can happen if providers choose to undermine P4P efforts, and how P4P could affect individual providers’ contributions to medical knowledge.

60. E-mail from R. Adams Dudley, Associate Professor of Medicine & Health Policy, University of California, San Francisco to author (Aug. 3, 2006, 12:51:00 PST) (on file with author) [hereinafter Dudley E-mail] (indicating that “[t]here is no one who has published about cost or cost-effectiveness” of P4P efforts). See generally R. Adams Dudley, *Pay-for-Performance Research: How To Learn what Clinicians and Policy Makers Need To Know*, 294 JAMA 1821, 1822 (2005) (commenting on Rosenthal et al., *supra* note 17 and noting that “no prior pay-for-performance research has reported the cost of improving quality or how that compares with the incentives offered in the pay-for-performance program.”). But see Peterson et al., *supra* note 40, at 267 (citing a study of one P4P experiment where “the author asserted that the incentive and administrative costs were small compared with potential gains in improved health and lower overall health care expenditures.”).

61. For discussions of how monetary incentives fail to solve principal-agent problems in other areas, see, for example, Bruno S. Frey & Margit Osterloh, *Yes, Managers Should Be Paid Like Bureaucrats*, 14 J. MGMT. INQUIRY 96 (2005), and Michael C. Jensen & Kevin J. Murphy, *Performance Pay and Top-Management Incentives*, 98 J. POL. ECON. 225 (1990).

62. E.g., Jim Molpus, *Pay for Performance: Is the Payoff Worth the Effort?*, HEALTHLEADERS ROUNDTABLE, Aug. 2005, at RT9 (2005), available at <http://www.healthleadersmedia.com/pdf/roundpdf/roundtable-Aug-2005.pdf> (noting that “[p]hysicians are still very, very suspicious of the motivation of these programs.”).

A. 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."⁶³ Nonetheless, some hospital executives believe that "not in the near future, nor possibly ever, will we develop a reliable severity adjustment system."⁶⁴ 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.⁶⁵ Providers may have limited ability to persuade smokers to enroll in such programs.

Other factors will affect providers' receptivity to P4P schemes. If there are too many schemes, "[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."⁶⁶ Providers' willingness to comply with P4P standards is also likely to decline if the number and complexity of schemes increase,⁶⁷ providers perceive the standards to be based on unreliable data, or

63. Garber, *supra* note 47, at 179.

64. Ateev Mehrotra et al., *Employers' Efforts to Measure and Improve Hospital Quality: Determinants of Success*, 22 HEALTH AFF. 60, 65 (2003) (noting that in interviews, some hospital executives expressed agreement with the statement).

65. See DUDLEY ET AL., *supra* note 16, at 29.

66. CUTLER, *supra* note 7, at 102.

67. See Molpus, *supra* note 62, at RT4 (noting that, according to one private P4P administrator, "[w]e 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

PAY-FOR-PERFORMANCE

payors are not transparent about the quality criteria.⁶⁸ Providers' reactions to any of those factors will be aggravated by the size of the financial disincentives involved, especially if those 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, stating that “[t]he 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.”⁶⁹ According to the AMA, all P4P programs must be completely voluntary; reimburse physicians the cost of their participation; finance rewards with supplemental funds; use “the best-available risk adjustment”; keep program features stable for at least two years;⁷⁰ and allow for deviation from guidelines when clinically appropriate with “minimal, but appropriate, documentation.”⁷¹ In addition, P4P programs must not employ financial penalties, judge individual physicians relative to one another, “threaten the economic viability of physician practices” that do not participate, judge physicians on the basis of factors beyond their control, or limit patient access to care.⁷² Opposition from the AMA is one reason a P4P proposal was dropped from the fiscal year 2006 budget resolution.⁷³

B. 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.

measures, the methodology and the application of the measurement sets differ.”).

68. See, e.g., Philip Betze, *Pay for Performance Tipping Point*, HEALTH LEADERS NEWS, Sept. 15, 2005, <http://www.healthleadersmedia.com/viewcontent/72244.html> (noting that the Medical Group Management Association, which represents physician group practices, has criticized 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.’”).

69. Am. Med. Ass’n, H-450.947 Pay-for-Performance Principles and Guidelines, Health and Ethics Policies of the AMA House of Delegates, *available at* http://www.ama-assn.org/apps/pf_new/pf_online?f_n=browse&doc=policyfiles/HnE/H-450.947.HTM (last visited Dec. 10, 2006).

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

71. Am. Med. Ass’n, *supra* note 69. The conditions listed in the text are not exhaustive. *Id.*

72. *Id.*

73. See *Congress Locks AMA into Pay-for-Performance Program in 2007*, INSIDE CMS, Jan. 12, 2006; Michael Romano, *AMA Deal Rankles Specialty Docs: Quality Pact with Feds Could Widen Professional Rift*, MOD. HEALTHCARE, Feb. 27, 2006, at 7.

Alternatively, providers could refuse to do business with third-party payors who tie payment to “unreasonable” performance measures. That response would disrupt many patients’ access to care. Finally, providers could comply with a P4P scheme, but do so 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 for targeted patients.

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 on only the healthiest people.”⁷⁴ If third-party payors reward providers on the basis of their patients’ cholesterol levels, some providers may select patients who are most likely to stick to a cholesterol-lowering treatment regimen. And they may 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” who would find it increasingly difficult to obtain care and could be relegated to low-quality providers.⁷⁵

Another method is data manipulation.⁷⁶ As many as fifty 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).⁷⁷ As many as seventy percent of physicians state they would be willing to do so under certain circumstances.⁷⁸ Physicians report a number of tactics for manipulating data in order to obtain reimbursement that could also be used to game and defeat P4P measures.⁷⁹

74. CUTLER, *supra* note 7, at 109.

75. Boyd et al., *supra* note 36, at 722.

76. See, e.g., CUTLER, *supra* note 7, at 108-09 (“If insurers are paid for controlling blood pressure in hypertensives, for example, they cannot be allowed to call everyone a hypertensive, knowing that most will be ‘controlled’ when blood pressure is actually tested.”).

77. Sidney T. Bogardus, Jr. et al., *Physicians’ Interactions with Third-Party Payers: Is Deception Necessary?*, 164 ARCHIVES INTERNAL MED. 1841, 1842 (2004). See also Wynia et al., *supra* note 40, at 1858. (“A sizable minority of physicians report manipulating reimbursement rules so patients can receive care that physicians perceive is necessary. Unless novel strategies are developed to address this, greater utilization restrictions in the health care system are likely to increase physicians’ use of such manipulative ‘covert advocacy’ tactics.”)

78. Bogardus, Jr. et al., *supra* note 77, at 1842.

79. Bogardus et al. stated:

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 use *rule out cancer* as the indication for a test rather

PAY-FOR-PERFORMANCE

An experiment in the United Kingdom's National Health Service (NHS) provides an example of how physicians could game P4P performance measures. In 2004, the NHS rewarded family practitioners based on the proportion of the practitioner's patients who received recommended care.⁸⁰ To avoid encouraging inappropriate care for outliers, the NHS permitted physicians to exempt patients from the denominator when calculating that proportion. Researchers found the use of such exemptions to be the strongest predictor of whether a physician reached the performance targets and concluded, "[m]ore research is needed to determine whether these practices are excluding patients for sound clinical reasons or in order to increase income."⁸¹

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.⁸² Much of this spending is the result of greater intensity of care, such as more frequent hospital admissions and specialist consultations.⁸³ The history of managed care illustrates how difficult it is for third-party purchasers to reduce overuse. Where third-party purchasers are blind to or unable to control 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.⁸⁴

C. 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."⁸⁵ However, as discussed earlier, clinical trials report average effects

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).

Id. at 1842 (citations omitted). Strategies such as altering the severity of a condition or changing diagnoses could be employed to game performance measures just as they are used to game reimbursement rules.

80. Doran et al., *supra* note 40, at 375.

81. *Id.* at 375. "The generally low levels of exception reporting suggest that large-scale gaming was uncommon." *Id.* at 383.

82. Wennberg, *supra* note 10, at 2.

83. Fisher et al., *Part 1*, *supra* note 3, at 273.

84. *See generally id.* at 273.

85. David M. Eddy, *Evidence-Based Medicine: A Unified Approach*, 24 HEALTH AFF. 9, 10 (2005). *See also* Evidence-Based Medicine Working Group, *Evidence-Based Medicine: A New*

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 ninety percent of AMI patients are prescribed beta-blockers allows physicians to deviate from the standard in ten percent of cases. But payors and providers will differ over whether ten percent is sufficient flexibility or too little.

III. EVIDENCE OF 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 evidence-based medicine:

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.⁸⁶

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 data on the effectiveness of P4P to be “sparse.”⁸⁷ 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

Approach to Teaching the Practice of Medicine, 268 JAMA 2420 (1992); Simon R. J. Maxwell, *Evidence Based Prescribing*, 331 BRIT. MED. J. 247 (2005) (editorial); David L. Sackett et al., *Evidence-Based Medicine: What It Is and What It Isn't*, 312 BRIT. MED. J. 71 (1996) (editorial).

86. Haynes, *supra* note 11, at 2.

87. DUDLEY ET AL., *supra* note 16, at 63.

PAY-FOR-PERFORMANCE

results when financial incentives were targeted to individual physicians, individual providers, and provider groups. According to the authors, “[t]here 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.”⁸⁸ 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.⁸⁹ Two types of financial incentives, fee-for-service payment enhancements and bonuses, showed mixed results.⁹⁰ Results were also mixed for studies that measured whether financial incentives encouraged the use of preventive care.⁹¹ 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).”⁹² 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. Purchasers incur the expense of P4P 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, both because the costs of P4P schemes have yet to be systematically measured⁹³ and also because effectiveness must be established before cost-effectiveness can be established.

The P4P movement proceeds from two premises: first, that clinicians tend to under-use 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, let alone doing so in a cost-effective manner.

IV. P4P IN MEDICARE

Prominent health policy scholars,⁹⁴ most recently the Institute of Medicine,⁹⁵

88. *Id.* at 28.

89. *Id.*

90. *Id.* at 28-29.

91. *Id.* at 29.

92. *Id.*

93. Dudley E-mail, *supra* note 60.

94. E.g., Donald M. Berwick et al., *Open Letter: Paying For Performance: Medicare Should Lead*, 22 HEALTH AFF. 8 (2003).

95. INST. OF MED., NAT’L ACADEMIES, REWARDING PROVIDER PERFORMANCE: ALIGNING

have argued that Medicare should take the lead in paying for performance. Medicare has launched limited demonstration programs to test the concept. In 2005, Medicare released the first quality-based bonus payments in the program's history, following promising results from one P4P demonstration, the Premier Hospital Quality Incentive program.⁹⁶ That program, launched in 2003, collects data on about thirty quality indicators for joint replacements, coronary artery bypass grafts, heart attacks, heart failure, and pneumonia.⁹⁷ For each clinical area, hospitals that score in the first and second deciles receive bonus payments from Medicare of two percent and one percent of Medicare payments for those services, respectively.⁹⁸ After the first year, the demonstration used the bottom two deciles in each area of care to set baselines for poor performers.⁹⁹ In the third and subsequent years, Medicare will reduce payments by up to two percentage points to hospitals that score below those baselines in the clinical areas involved.¹⁰⁰ Medicare predicts that most hospitals will improve and that "few, if any, hospitals would get a payment reduction."¹⁰¹ 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.¹⁰²

Congress is considering proposals to expand on those initiatives within Medicare. In the 109th Congress, the former chairman of the Health Subcommittee of the House Committee on Ways and Means, Representative Nancy Johnson (R-CT), introduced legislation that would give larger payment increases to physicians who meet administratively specified performance targets or who make significant progress toward meeting them.¹⁰³ Senator Chuck Grassley (R-IA) introduced even more expansive legislation, which would create P4P incentives for hospitals, physicians, Medicare Advantage plans, home health agencies, and other providers.¹⁰⁴ 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.¹⁰⁵ In late 2006, Congress passed legislation tying higher physician payments to quality reporting requirements, a move intended to facilitate broader P4P initiatives.¹⁰⁶

INCENTIVES IN MEDICARE (2006).

96. Medicare Demonstration, *supra* note 52.

97. *Id.*

98. *Id.*

99. PREMIER, INC., *supra* note 33.

100. Medicare Demonstration, *supra* note 52.

101. *Id.*

102. Abelson, *supra* note 10, at C3.

103. Medicare Value-Based Purchasing for Physicians' Services Act of 2005, H.R. 3617, 109th Cong. (2005).

104. Medicare Value Purchasing Act of 2005, S.1356, 109th Cong. (2005).

105. *Budget Reconciliation Remains a Loose End for Congress; Senate-Approved Package Must Return to the House*, WASH. OUTLOOK, Dec. 23, 2005 (on file with author).

106. Robert Pear, *Medicare Links Doctors' Pay to Practices*, N.Y. TIMES, Dec. 12, 2006, at 1.

A. Special Challenges Posed by Medicare

Introducing P4P into the traditional Medicare program presents a number of unique problems. First, Medicare's elderly patient population is more likely to suffer from co-morbidities, which makes it more likely that a P4P scheme would encourage inappropriate care for some patients. Second, the political forces that govern Medicare make it less likely that those perverse incentives would be corrected, and more likely that a P4P scheme would increase costs for taxpayers who finance the program. Finally, because Medicare greatly influences the behavior of private purchasers, any harm caused by a Medicare P4P scheme likely would spill over into the private sector as well.

1. Greater Potential for Error

Any harm that a P4P system could conceivably cause is more likely to appear in Medicare, for two reasons. First, Medicare's patient base is more susceptible than those of private insurers to the unintended harms that can result from P4P programs. This is due to the fact that Medicare enrollees are older, less healthy, and more likely to have low incomes¹⁰⁷ and to consume more medical care¹⁰⁸ than non-elderly Americans with private health insurance. Close to one-third of Medicare beneficiaries have four or more chronic conditions, and those patients account for nearly eighty percent of Medicare spending.¹⁰⁹ The large number of beneficiaries with chronic conditions increases the likelihood that a

107. See, e.g., Karen Davis et al., *Medicare Versus Private Insurance: Rhetoric And Reality*, W2 HEALTH AFF. - WEB EXCLUSIVE, Oct. 9, 2002, at W311, <http://content.healthaffairs.org/cgi/reprint/hlthaff.w2.311v1.pdf>. As Davis et al. write:

Medicare beneficiaries are more likely than the privately insured are to be in poor health and have low incomes. In the survey, two-thirds of persons under age sixty-five with private health insurance rated their health status as excellent or very good, compared with two-fifths of elderly Medicare beneficiaries. The proportion of elderly Medicare beneficiaries rating their health as fair or poor was three times higher than that of privately insured adults. Four of five Medicare beneficiaries had a chronic condition, compared with just over one-third of the privately insured. Medicare beneficiaries were four times as likely as the privately insured were to report having two or more chronic conditions.

Id. at W313.

108. See generally Ctrs. for Medicare & Medicaid Servs., National Health Expenditure Data: Age Tables 1-6, <http://www.cms.hhs.gov/NationalHealthExpendData/downloads/agetables.pdf> (last visited Dec. 10, 2006) (showing that per capita personal health care expenditures for Americans age sixty-five and older are roughly four times those for Americans under age sixty-five in the table entitled, "Personal Health Care Spending by Type of Service, Age Group, and Source of Payment Distribution, Calendar Year 1996").

109. Robert A. Berenson & Jane Horvath, *Confronting the Barriers to Chronic Care Management in Medicare*, W3 HEALTH AFF. - WEB EXCLUSIVE, Jan. 22, 2003, at W3-37, W3-38, <http://content.healthaffairs.org/cgi/reprint/hlthaff.w3.37v1.pdf>.

P4P scheme would create incentives to mistreat such patients and turn them into “medical hot potatoes”¹¹⁰ that providers would try 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 and P4P measures.¹¹¹ Moreover, Medicare patients are more likely to have treatment goals that differ from those assumed by CPGs and P4P measures.¹¹² As a result, a Medicare P4P effort is more likely to create harmful financial incentives than efforts focused on non-elderly patients.

Compounding this challenge is the fact that Medicare is a creature of the political process. The political forces that govern Medicare would 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 “700,000 physicians, 6,000 hospitals and thousands of other providers and suppliers,”¹¹³ who 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 42 million seniors,¹¹⁴ 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 and Medicaid Services (CMS), and whatever other bodies make or influence Medicare policy. For nearly a decade, the health care industry has led other sectors of the economy in terms of dollars spent lobbying Congress,¹¹⁵ and employing P4P financial incentives in Medicare would only increase such lobbying. The unavoidable political pressure would reduce Medicare’s flexibility to make timely, focused, and evidence-based adjustments

110. Boyd et al., *supra* note 36, at 722 (“Current pay-for-performance initiatives can create financial incentives for physicians to focus on certain diseases and younger or healthier Medicare patients.”).

111. *Id.* at 720.

112. See Tinetti et al., *supra* note 39 at 2870. (“[E]vidence is emerging that patients, particularly elderly patients and those with multiple conditions, vary in regard to the amount of importance they place on health outcomes such as longer survival, the prevention of specific disease events, and physical and cognitive functioning, as well as in the amount of inconvenience and risk of adverse effects they are willing to tolerate.”)

113. Press Release, Ctrs. for Medicare & Medicaid Servs., Fact Sheet: HCFA Management Reforms (May 1, 2000), available at <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=379>.

114. EARL DIRK HOFFMAN, JR. ET AL., BRIEF SUMMARIES OF MEDICARE & MEDICAID: TITLE XVIII AND TITLE XIX OF THE SOCIAL SECURITY ACT AS OF NOVEMBER 1, 2005 6 (2006), available at <http://www.cms.hhs.gov/MedicareProgramRatesStats/downloads/MedicareMedicaidSummaries2005.pdf>.

115. See PoliticalMoneyLine, Leading Sector Spending for Federal Lobbying 7/1/05-12/31/05, Money in Politics Databases, http://www.tray.com/cgi-win/lp_sector.exe?DoFn=my&Year=05 (last visited Dec. 10, 2006). For more information, see previous reports dating back to Leading Sector Spending for Federal Lobbying 1/1/99–6/30/99, Money in Politics Databases, http://www.tray.com/cgi-win/lp_sector.exe?DoFn=my&Year=99 (last visited Dec. 10, 2006).

PAY-FOR-PERFORMANCE

to its payment structure and would put upward pressure on Medicare outlays. The politicization of quality-based financial incentives could also be expected to politicize the search for data to guide those incentives, as interested parties seek federally financed research on modes of care that they believe should be rewarded for being of higher quality.

Indeed, rent-seeking behavior would attend every aspect of a Medicare P4P system. As do Medicare's payment systems broadly, a P4P system would spur congressional and administrative lobbying by providers who seek to protect or increase their incomes, who fear being penalized for factors beyond their control, who do not 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 that performance measures can be gamed, who do not 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 come from those who see P4P as a way to reduce Medicare outlays. Given the future financial pressures facing Medicare¹¹⁶ and the growing scarcity of federal resources for other congressional priorities, politicians and interest groups could 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 groups that are relatively easy to organize for political action.¹¹⁷ By comparison, the per capita benefits of using P4P to constrain Medicare spending are spread out among a larger group of individuals (i.e., taxpayers) that is more difficult to organize. The likely result is that Medicare would be able to create P4P 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. Given the influence that Medicare's P4P decisions would have on providers' incomes and the overall tax burden, those interest groups can be expected to use political

116. See generally GOV'T PRINTING OFFICE (GPO), THE 2006 ANNUAL REPORT OF THE BOARD OF TRUSTEES OF THE FEDERAL HOSPITAL INSURANCE AND FEDERAL SUPPLEMENTARY MEDICAL INSURANCE TRUST FUNDS (2006), available at <http://www.cms.hhs.gov/ReportsTrustFunds/downloads/tr2006.pdf>.

117. See, e.g., Alan Garber, *Cost-Effectiveness and Evidence Evaluation as Criteria for Coverage Policy*, W4 HEALTH AFF. - WEB EXCLUSIVE, May 19, 2004, at W4-284, W4-293 (2004), <http://content.healthaffairs.org/cgi/reprint/hlthaff.w4.284v1.pdf> ("Because they are a large and politically powerful constituency, Medicare beneficiaries have a powerful voice in deliberations over any major change.").

pressure to block changes that they expect would adversely affect their interests. In contrast, private purchasers have greater flexibility when designing and adjusting their P4P schemes.¹¹⁸ As a result, Medicare would take longer to correct errors than private third-party purchasers do.

Medicare would have to overlay a P4P scheme on its already complex administrative pricing responsibilities. Medicare currently operates sixteen different payment systems for various types of providers and health plans. The physician payment system alone must set prices for more than 7000 distinct services¹¹⁹ in each of eighty-nine payment localities.¹²⁰ P4P would pile even more complexity on top of that system. 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 over-investment in some areas of care and under-investment in other areas¹²¹ and for being slow to fix such errors.¹²² For example, 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 procedures. Moreover, ASC payments will not be adjusted for these factors until 2008.¹²³

The same rigidity would afflict Medicare's administration of performance-

118. Providers have less power to block P4P experiments by private purchasers. Though providers can refuse to participate in a private P4P program, the insurer can still survive if its customers value the improvements in technical quality more than any reduced access that comes from providers being excluded from a plan's network or refusing to accept reimbursements offered to non-compliant providers as payment in full. If the insurer loses customers, on the other hand, that would signal to the insurer that its customers perceive that the costs of its P4P program outweigh the benefits.

119. MEDICARE PAYMENT ADVISORY COMM'N, PAYMENT BASICS: PHYSICIAN SERVICES PAYMENT SYSTEM 1 (2006), available at http://medpac.gov/publications/other_reports/Sept06_MedPAC_Payment_Basics_Physician.pdf.

120. MEDICARE PAYMENT ADVISORY COMM'N, REPORT TO THE CONGRESS: MEDICARE PAYMENT POLICY 11 (1999), available at http://www.medpac.gov/publications/congressional_reports/Mar99%20Ch1.pdf.

121. See, e.g., U.S. FED. TRADE COMM'N & U.S. DEP'T OF JUSTICE, *supra* note 7, at 9 (executive summary).

122. See, e.g., *Medicare: Private Payer Strategies Suggest Options To Reduce Rapid Spending Growth: Hearing Before H. Subcomm. on Health, Comm. on Ways and Means*, 109th Cong. 7 (2006) (statement of Janet L. Shikles, U.S. General Accounting Office), available at <http://www.gao.gov/archive/1996/he96138t.pdf> ("Because of strict statutory constraints and its own burdensome regulatory and administrative procedures, [CMS] is slow to address overpricing and overutilization problems.").

123. U.S. Fed. Trade Comm'n & U.S. Dep't of Justice, *supra* note 7, ch. 3, at 25.

PAY-FOR-PERFORMANCE

based payments. Inaccurate data, mis-targeted or mis-calibrated 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 unwarranted financial bonuses).

2. *A Large Wake*

Not only would errors be more likely in a Medicare-administered P4P program, but the resulting harms would be far more widespread than those of similar errors by private payors. In addition to having a higher *proportion* of patients who are in poor health¹²⁴ and 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 thirty-seven million elderly and disabled enrollees in traditional Medicare and another five million enrolled in private plans.¹²⁵ Medicare beneficiaries also consume more medical care than the non-elderly. Finally, for reasons discussed below, they also would be less able than non-Medicare patients (or those patients' employers) to switch insurers if a P4P program causes unintended harm.

Any harms resulting from a Medicare P4P system would likely spill over into private insurance. Private insurers' payments are often heavily influenced by Medicare's payment rates.¹²⁶ If the federal government creates one P4P system for the traditional Medicare program, other third-party purchasers would face strong incentives to adopt that system as well. Many purchasers (e.g., state governments and private insurers, including Medicare Advantage plans) likely would rather have the federal government incur the costs of creating and updating their 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 creates could spread beyond the Medicare population.

B. How To Address the Unique Pitfalls of Medicare P4P

P4P gives third-party payors considerable power to influence – for good or

124. See Davis et al., *supra* note 107, at 313.

125. Milt Freudenheim, *UnitedHealth to Buy PacifiCare in Push into Medicare*, N.Y. TIMES, July 7, 2005, at C1 (explaining that the nation's largest private insurer in 2004 was WellPoint, with 27.7 million enrollees).

126. Uwe E. Reinhardt, *The Medicare World from Both Sides: A Conversation with Tom Scully*, 22 HEALTH AFF. 167, 168 (2003). Tom Scully, former administrator of CMS, stated that "Medicare and Medicaid are such dominant players that the private sector has been forced to follow along – shadow pricing [those programs] in recent years." *Id.*

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.

1. Confine P4P to Private Medicare Plans

One option that would enable Medicare beneficiaries to reap the benefits of P4P while minimizing any harm 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 continue the learning process that allows for testing and refining P4P strategies. Instead of creating a single set of P4P measures and incentives, private Medicare plans would be able to experiment with different, 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, Medicare Advantage plans would gravitate toward whatever successful P4P strategies emerged, while keeping unintended harms to a minimum. For example, private insurers adopted prescription drug coverage years before Medicare did so, and private plans enjoy more flexibility to innovate and adjust that coverage than Medicare Part D plans.

Preventing CMS from developing P4P incentives for traditional Medicare would be necessary to catalyze this learning process. A P4P scheme in traditional Medicare would apply to thirty-seven 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 innovative alternative or supplemental P4P incentives, those efforts would be less likely to influence providers. According to Tom Scully, former administrator of the Centers for Medicare and Medicaid Services, “[i]n many markets Medicare and Medicaid comprise over 65 percent of the payments to hospitals, and more than 80 percent in some physician

PAY-FOR-PERFORMANCE

specialties.”¹²⁷ Necessity would force providers to give highest consideration to Medicare’s performance measures. Providers might ignore financial incentives offered by other payors, 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 already requires). Even where providers complied with alternative private P4P programs, private payors 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 therefore 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 costly 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 payor is trying to treat an outlier as though she were the average patient. The solution to such screening is for Medicare to adjust its payments to reflect the expected health care costs of the individual patient. In fact, the Centers for Medicare and Medicaid Services are currently refining Medicare’s risk-adjustment capabilities.¹²⁸ 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.

Another potential criticism of confining P4P to Medicare Advantage plans is that without traditional Medicare’s purchasing power, those plans would be less able to change providers’ behavior. Although this is true, the fact that Medicare has the market power to influence providers’ behavior for the good also means it has the power to influence providers’ behavior in ways that harm beneficiaries. Ultimately, P4P schemes will be more effective if they focus on precision first and market power second. Smaller experiments by private insurers are better positioned to deliver that precision, and could build market power by establishing a reputation for quality. Though the evidence is mixed, private P4P efforts have shown some ability to influence provider behavior.¹²⁹ Moreover, Medicare’s market power derives from the political power of providers and seniors. Though Medicare theoretically has the market power to change providers’ behavior, providers often have the political power to change Medicare’s behavior. The fact that private plans are not as easily influenced by providers offsets their relatively

127. *Id.* at 169-70.

128. See generally Robert A. Berenson, *Medicare Disadvantaged and the Search for the Elusive ‘Level Playing Field’*, W4 HEALTH AFF. - WEB EXCLUSIVE W4-572, W4-585 (2004), <http://content.healthaffairs.org/cgi/reprint/hlthaff.w4.572v1.pdf>.

129. See generally DUDLEY ET AL., *supra* note 16.

weaker market position, and leaves open the question of whether private plans or Medicare would have more influence over providers' behavior.

Allowing P4P only in private Medicare plans would confine P4P to a maximum of five million Medicare beneficiaries at present, or twelve 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 advanced in the late 1990s¹³⁰ 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.

2. Use Patient-Focused Financial Incentives

Provider-focused financial incentives remain an unproven method for improving quality. Much more research is necessary before payors can know whether and where P4P will change providers' behavior and improve patient outcomes. Thus, it is important that payors not focus solely on provider-focused financial incentives. Other measures may also induce providers to improve the quality of care.

One possibility is the use of incentives that increase patients' financial 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, could conceivably have a greater influence on provider behavior, and would allow patients to avoid the harms that may result from hidden financial incentives targeted to providers.

The same sort of data that third-party payors 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 (or who use providers known for delivering recommended care) would face lower out-of-pocket costs, while those who do not would face higher out-of-pocket costs. Patients would know sooner whether a provider was not adhering to the plan's quality guidelines because that deviation would affect their pocketbooks. In such cases, a dialogue between the patient and provider (and perhaps the health plan) could ensue. Both the price signals offered by the plan and the subsequent dialogue would lead to better-educated patients who would help drive quality improvements.

130. See Nat'l Bipartisan Comm'n on the Future of Medicare, Building a Better Medicare for Today and Tomorrow (Mar. 16, 1999), available at <http://thomas.loc.gov/medicare/bbmtt31599.html>.

PAY-FOR-PERFORMANCE

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 decision making 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 might 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 patients 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 payors – 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. As Mary E. Tinetti of the Yale School of Medicine and her colleagues write:

[O]ne of the hallmarks of quality-assurance programs is a reduction in the variation of practice patterns among providers. It is difficult to separate inappropriate variation due to neglect or ignorance on the part of providers from appropriate variation due to the total disease burden and the preferences of patients.¹³¹

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 mis-targeted or mis-calibrated 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

131. Tinetti et al., *supra* note 39, at 2870.

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 if incorporated into 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 P4P efforts. 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 decision making would be at the level of the individual patient.

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 responsibility 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.