



RESEARCH REPORT

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

Payment Methods: How They Work

Robert A. Berenson
URBAN INSTITUTE

Divvy K. Upadhyay
URBAN INSTITUTE

Suzanne F. Delbanco
CATALYST FOR
PAYMENT REFORM

Roslyn Murray
CATALYST FOR
PAYMENT REFORM

April 2016

Updated June 10, 2016





ABOUT THE URBAN INSTITUTE

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

Contents

Acknowledgments	vi
Introduction	1
Context, Design, and Operational Issues Affect Payment Method Impact	2
Payment Method Attributes	3
Methods and Analysis	4
Selected Payment Methods	6
Bibliography	7
Fee Schedules for Physicians and Other Health Professionals	8
Key Objectives	9
Strengths	9
Weaknesses	10
Design Choices to Mitigate Weaknesses	11
Compatibility with Other Payment Methods and Benefit Design Options	12
The Focus on Performance Measurement	13
Potential Impact on Provider Prices and Price Increases	13
Primary Care Capitation	14
Key Objectives	14
Strengths	15
Weaknesses	15
Design Choices to Mitigate Weaknesses	16
Compatibility with Other Payment Methods and Benefit Designs	17
The Focus of Performance Measurement	18
Potential Impact on Provider Prices and Price Increases	18
Per Diem Payment to Hospitals for Inpatient Stays	19
Key Objectives	20
Strengths	20
Weaknesses	20
Design Choices to Mitigate Weaknesses	21
Compatibility with Other Payment Methods and Benefit Designs	21
The Focus of Performance Measurement	22
Potential Impact on Providers' Prices and Price Increases	23
Diagnosis Related Groups–Based Payment to Hospitals for Inpatient Stays	24
Background	24

Key Objectives	25
Strengths	25
Weaknesses	26
Design Choices to Mitigate Weaknesses	27
Compatibility with Other Payment Methods and Benefit Designs	28
The Focus of Performance Measurement	29
Potential Impact on Provider Prices and Price Increases	30
Global Budgets for Hospitals	31
Key Objectives	32
Strengths	33
Weaknesses	33
Design Choices to Mitigate Weaknesses	34
Compatibility with Other Payment Methods and Benefit Designs	35
The Focus of Performance Measurement	35
Potential Impact on Provider Prices and Price Increases	36
Bundled Episode Payment	37
Key Objectives	38
Procedure-Based Bundled Episodes	39
Strengths	39
Weaknesses	40
Design Choices to Mitigate Weaknesses	41
Condition-Specific Bundled Episodes	42
Strengths	42
Weaknesses	42
Design Choices to Mitigate Weaknesses	43
Compatibility with Other Payment Methods and Benefit Designs	44
The Focus of Performance Measurement	44
Potential Impact on Provider Prices and Price Increases	45
Global Capitation to an Organization	46
Key Objectives	47
Strengths	47
Weaknesses	48
Design Choices to Mitigate Weaknesses	49
Compatibility with Other Payment Methods and Benefits Designs	50
The Focus of Performance Measurement	51
Potential Impact on Provider Prices and Price Increases	51
Shared Savings	53

Key Objectives	54
Strengths	54
Weaknesses	55
Design Choices to Mitigate Weaknesses	56
Compatibility with Other Payment Methods and Benefit Designs	57
The Focus of Performance Measurement	58
Potential Impact on Provider Prices and Price Increases	58
Pay-for-Performance	60
Background	60
Key Objectives	61
Strengths	61
Weaknesses	62
Design Choices to Mitigate Weaknesses	63
Compatibility with Other Payment Methods and Benefit Designs	64
Focus of Performance Measurement	65
Potential Impact on Provider Prices and Price Increases	66
Notes	67
References	68
About the Authors	69
Statement of Independence	71

Acknowledgments

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

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

A technical expert panel advised the project team and reviewed the reports at different stages. This team consists of **Michael E. Chernew**, Leonard D. Schaeffer professor of health care policy and director of Healthcare Markets and Regulation Lab, Harvard Medical School; **Francois de Brantes**, executive director, Health Care Incentives Improvement Institute; **Anna Fallieras**, program leader, Health Care Initiatives and Policy, General Electric; **Kate Farley**, executive director, Pennsylvania Employees Benefit Trust Fund; **Joseph J. Fifer**, president and chief executive officer, Healthcare Financial Management Association; **Robert Galvin**, chief executive officer, Equity Healthcare, operating partner, Blackstone, and former chief medical officer, General Electric; **Paul Ginsburg**, professor and director of public policy, Schaeffer Center for Health Policy and Economics, University of Southern California, and senior fellow and director, Center for Health Policy, Brookings Institution; **Stuart Guterman**, senior scholar in residence, AcademyHealth, and former vice president, Medicare and Cost Control, The Commonwealth Fund; **Vincent E. Kerr**, president, Care Solutions, National Accounts, UnitedHealthcare, and former chief medical officer, Ford Motor Company; **Peter Kongstvedt**, principal, P.R. Kongstvedt Company, LLC, and senior health policy faculty member, George Mason University; **Jeff Levin-Scherz**, assistant professor, Department of Health Policy and Management, Harvard University, and national coleader, Willis Towers Watson; **Robert Murray**, president and consultant, Global Health Payment LLC, and former executive director, Maryland Health Services Cost Review Commission; **Dave Prugh**, independent adviser and consultant, and former vice president of Reimbursement and Contracting Strategy, WellPoint, Inc.; **Simeon Schwartz**, founding president and chief executive officer, WESTMED Medical Group; and **Lisa Woods**, senior director, US health care, Walmart Stores Inc.

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

Introduction

The broad policy consensus that payment methods for physicians and hospitals need to evolve from volume based to value based often implicitly assumes clear dividing lines between the two categories. However, most of what are considered value-based payment reform models are being implemented on top of current, volume-based payment approaches, or as HHS calls it, “fee-for-service architecture.” This points to our need to understand the attributes of all common payment approaches—those long in use and more recent reforms—to better judge not only their strengths and weaknesses as stand-alone payment methods but also how they likely interact with other payment methods. With this knowledge, we can adopt designs that improve the effectiveness of payment reform models.

Accordingly, to gain a better understanding of payment reform opportunities, we explore not only the attributes of reform approaches but also payment methods that constitute their underlying architecture. Our review demonstrates that, in fact, every payment method has strengths and weaknesses. By understanding them, it might be possible to implement payment reform designs that take advantage of their strengths and mitigate their weaknesses. Often the best way is to develop mixed or hybrid payment models that accentuate the strengths of each method while mitigating the negative attributes.

Busse and Quentin (2011) make this conclusion on the broad adoption of diagnosis related groups (DRGs) in most European countries:

The payment of hospitals in all countries ... consists of a highly sophisticated mix of different payment mechanisms that aim to modify the type and strength of the incentives in DRG-based hospital payment. The resulting intricately blended payment systems—incorporating elements of fee-for-service payment, per diem payment and global budgets—are more likely to contribute to achieving the societal objectives of securing high-quality hospital care at affordable costs than any other hospital payment mechanism alone.(p. 164)

Our primary considerations in describing the attributes of payment systems are how payment methods can be designed to maximize their potential and mitigate their weaknesses and how adoption of complementary payment and benefit designs can enhance their strengths. In addition, payment attributes include other considerations that round out the core elements to be considered when deciding which payment methods to adopt and in what combinations.

Context, Design, and Operational Issues Affect Payment Method Impact

Too often, analyses of payment methods are based on idealized versions and focus on the incentives the payment method embodies while ignoring practical issues that influence how it will behave when adopted and implemented.

The context of a payment methods' adoption often matters crucially to its impact. For example, traditional Medicare sets payment rates, whereas private payers have to negotiate rates. Pricing power resulting from some forms of consolidation may therefore have differential impacts on the success of payment methods, such as population-based payments, designed for large provider organizations. Similarly, private payers have more flexibility than traditional Medicare to design benefits that complement particular payment approaches, such as tiered or narrow networks. To pay hospitals through global budgets requires an all-payer system that addresses payments across the board—no individual payer, even one as important as Medicare, can itself pay hospitals through global budgeting. The context matters.

The specific design of the payment method, including the relative generosity of the payments, can also strongly influence the effect on providers' behavior. A fee schedule inherently contains incentives to provide more services, often more than needed or appropriate. But misvaluation of fees (i.e., payments far more or less than cost of production) will favor certain services more than others. Under

population-based payment approaches, such as those for accountable-care organizations, payments are accompanied by measurements of quality for preventive services. Such a policy may reduce the temptation inherent in the payment method to stint on these services. Evaluating the impact of the payment method without factoring in the policy design can lead to inaccurate conclusions.

Finally, in describing the theoretical incentives any payment method produces, analysts may miss the substantial operational challenges of implementation. These include administrative feasibility and the potential for perverse, unintended provider responses that can defeat the method's purpose. Anticipating and addressing operational challenges in design through accompanying policies and oversight may resolve the concerns. Yet, sometimes, implementation challenges may make a conceptually logical payment method too difficult to actually put into place.

Payment Method Attributes

Despite “it all depends” caveats that offer cautions before definitive conclusions about the growing array of payment options in use or proposed, policymakers should consider payment methods' attributes to decide how (or whether) to proceed with payment reform. Identifying payment methods' attributes can also instruct consumers, patients, providers, payers, and policymakers about their potential benefits and harms, informing how monitoring and oversight might proceed. Further, too much of the discussion of payment reform has focused on payment models' theoretical effects rather than on their interactions with other payment methods. We must also consider interaction with an array of benefit designs that either encourage or frustrate the opportunities for payment reform to improve value.

Advised by a technical expert panel of payment and benefit design experts, we selected the nine provider payment methods used most commonly by third-party payers or insurers—public and otherwise—to pay physicians and hospitals. Understanding how each payment method works, with its strengths, weaknesses, and other attributes identified, will help us find complementary payment and benefit design approaches that combine the strengths and mitigate weaknesses inherent in each payment method.

Methods and Analysis

All payment methods reviewed here have been peer-reviewed in the literature. However, we do not consider the available research-based evidence definitive, largely because research on payment methods depends on the specific payment design, including the generosity of payment, the context in which it would be applied, and the ability to manage attendant operational challenges. To generalize from the available, somewhat limited literature would be misleading. At the same time, others have reviewed payment approaches, crafting their own assessments similar to ours. What makes our review unique is our concise summaries of the payment methods' most salient attributes. We list the primary sources we relied on for much of our information and judgment, but we do not attempt to reference literature for every observation made. Our review is not intended for an academic audience, but rather as a practical guide for stakeholders interested in learning more about payment and its intricacies.

In addition, we have largely relied on informed, expert opinion, not only from the authors but also from a technical expert panel of payment experts who collectively represent the views of informed payers, purchasers, providers, payment administrators, and academic economists and policy analysts. The payment attributes listed, then, reflect the peer-reviewed evidence with its limitations, the authors' experiences, and the panel's the wide and deep expertise, producing consensus judgments as well as informed speculation. We made a special effort to consider not only the effect of payment incentives but also actual implementation, with identification of operational issues and challenges.

The nine payment methods reviewed are a subset of the payment models presented in *A Typology of Payment Methods* (Berenson et al. 2016). As noted in that document, different labels are often applied to one payment method. And there is certainly no best way to organize their presentation. Our payment typology de-emphasizes the sorting of payment methods according to provider type. Yet, given the move toward integrating services across traditional provider silos, some payment methods, in fact, do apply specifically to particular provider types. Accordingly, our nine payment method chapters include provider-specific and generic methods. We also consider payment methods that make base payments to providers and those that provide incremental bonuses and penalties on top of a base payment.

We recognize that listing payment methods as distinct is also somewhat arbitrary. Payment methods can be viewed as falling on a continuum rather than with a clean line separating them. Even a fee schedule for health professionals, often viewed as the prototype of fee-for-service, can include payment codes with elements of both population-based payment and episode-based payment. As we consider designs that attempt to mitigate the weaknesses each payment method exhibits, it becomes clearer that practical application of payment methods often blurs the lines between them. But to

discuss fine distinctions between payment methods, we have elected to review their “purer” forms as the starting point.

A few of the nine payment methods reviewed are long-standing approaches that have been used in different countries and in different markets. Others are more recently proposed approaches, currently undergoing active testing by Medicare, Medicaid, and private payers. Some, such as fee schedules for physicians, are widely, almost ubiquitously, used in the United States and many other countries. Others, such as global budgets for hospitals, are rare in the United States (although global budgets are now being implemented in an all-payer demonstration in Maryland) but have long been in broad use internationally.

Two methods reviewed—per diem payment and payment for hospital stays using DRGs—are by now classic ways of paying hospitals. We include them because their merits can vary in relation to other payment reforms that might be adopted and in relation to benefit designs that affect their operational feasibility. We also include “value-based” payment models being actively tested by Medicare and private payers, including bundled episodes, population-based payment, shared savings, and pay-for-performance. Finally, we revisit primary care capitation, which is being rediscovered as a potential payment reform approach either on its own (with performance reporting) or as a hybrid in conjunction with a reduced price fee schedule.

The impact of any particular payment method will vary based on source of payment (such as private insurance, Medicare, Medicaid, or direct payment by consumers and patients). The clearest example is the discussion of the effect of the payment method on prices. Medicare sets administrative prices whereas private insurance negotiates rates with providers, so the latter is much more dependent on market factors in which the payment methods are adopted. As we review the various attributes of payment methods, where relevant we attempt to distinguish how the method applies to different payers.

We organize the discussion of core attributes of payment methods in the following way:

- **Background information.** An explanation of how the payment method works and relevant experience with the approach
- **Key objectives.** What the payment method is designed primarily, sometimes uniquely, to achieve
- **Strengths.** Both theoretical, incentive-related likely advantages and practical, operational ones

- **Weaknesses.** Both theoretical, incentive-related likely disadvantages and practical, operational ones
- **Design choices to mitigate weaknesses.** Opportunities in actual implementation, largely based on the weaknesses identified, to reduce potential detrimental effects
- **Compatibility with other payment methods and with benefit design options.** Given that any payment method will be strongly interdependent with (1) concurrent methods for the same or related providers and (2) variations in benefit designs, we identify common interactions, both positive and negative. In this section, also, we suggest payment hybrid approaches that are either theoretically appealing based on incentives or are operating in limited areas of the United States or other countries.
- **Focus of performance measurement.** Policymakers and payers have broad interest in being able to measure many aspects of care, perhaps best summarized in the Institute for Healthcare Improvement's Triple Aim goal of simultaneously improving population health, patients' experience of care, and per capita cost. However, we consider measures of these domains of care common for all payment methods here, so we emphasize the vulnerabilities for which performance measurement would be particularly desirable.
- **Potential impact on providers' prices.** Most discussions of payment reform focus on their likely impact on health care costs, not on the impact on prices per se, prices being a major determinant of costs. Often, discussion of costs tends to be dominated by impact on service use; the equally important issue of transaction prices that determine payment amounts is largely neglected. Prices are often unrelated to payment method. For example, a market-dominant health care system can demand high prices whatever the form of payment. However, certain payment approaches have intrinsic features that could affect providers' prices. Note that Medicare sets prices and generally does not negotiate. So we identify and briefly discuss the features that may affect prices for private payers and whether particular market-related features are likely to influence them.

Selected Payment Methods

- Base payments
 1. Fee schedules for physicians and other health professionals

2. Primary care capitation
 3. Per diem payment to hospitals for inpatient stays
 4. Diagnosis related groups-based payment to hospitals for inpatient stays
 5. Global budgets for hospitals
 6. Bundled episode payments
 7. Population-based payments, including capitation
- Incremental payments
8. Shared savings
 9. Pay-for-performance

Bibliography

The following sources informed the analysis in the nine chapters of this report.

- Berenson, Robert A., Jonathan H. Sunshine, Arkaprava Deb, Julia A. Doherty, Ellen T. Kurtzman, Elizabeth S. Richardson, Noah S. Kalman, et al. 2012. *The Effect of Provider Payment Systems on Quality, Cost and Efficiency, and Access: A Systematic Literature Review*. Warsaw, PL: InterQuality Research Project.
- Dredge, Robert. "Hospital Global Budgeting." In *How-To Manuals: Designing and Implementing Health Care Provider Payment Systems*, edited by John C. Langenbrunner, Cheryl Cashin, and Sheila O'Dougherty. Washington, DC: World Bank.
- Hall, Mark A., and Robert A. Berenson. 1998. "Ethical Practice in Managed Care: A Dose of Realism." *Annals of Internal Medicine* 128 (5): 395–402.
- Physician Payment Review Commission. 1989. *Physician Payment Review Commission, Annual Report to Congress*. Washington, DC: Physician Payment Review Commission.
- Kongstvedt, Peter R. (ed). 2012. *Essentials of Managed Health Care, 6th ed*. Burlington, MA: Jones & Bartlett.
- Langenbrunner, John C., Cheryl Cashin, and Sheila O'Dougherty (eds). 2009. *How-To Manuals: Designing and Implementing Health Care Provider Payment Systems*. Washington, DC: World Bank.
- Orentlicher David. 1996. "Paying Physicians More to Do Less: Financial Incentives to Limit Care." *University of Richmond Law Review* 30: 155–198. papers.ssrn.com/sol3/papers.cfm?abstract_id=2020793.
- Quentin, Wilm, David Scheller-Kreinsen, Miriam Blümel, Alexander Geissler, and Reinhard Busse. 2013. "Hospital Payment Based on Diagnosis-Related Groups Differs in Europe and Holds Lessons for the United States." *Health Affairs* 32 (4): 713–723.
- UnitedHealth Center for Health Reform & Modernization. 2012. *Farewell to Fee-For-Service? A "Real World" Strategy for Health Care Payment Reform*. Hopkins, MN: UnitedHealth Group.
- Office of the Assistant Secretary for Planning and Evaluation. 1990. *Incentive Arrangements Offered by Health Maintenance Organizations and Competitive Medical Plans to Physicians. Report to Congress. Volume 1*. Washington, DC: US Department of Health and Human Services.

Fee Schedules for Physicians and Other Health Professionals

A fee schedule is a list of the maximum rate a payer will allow for services, with the definition of services based on code sets such as CPT (Current Procedural Terminology) in the United States and ICD-10 PCS (International Classification of Diseases, tenth revision, Procedure Coding System) in some other countries. Typically, the payment is the lower of the provider's actual charge or the fee schedule allowance. Most payers determine fee schedules first by establishing relative weights (also referred to as relative value units) for the list of service codes and then by using a dollar conversion factor to establish the fee schedule.

Before payers used fee schedules, they used variations what is referred to as the usual, customary, reasonable (UCR) method. This approach, modeled after the method most private payers used at the time, was enacted into law as Medicare's method for compensating physicians in 1965. Medicare's version was referred to as CPR—customary, prevailing, and reasonable—representing the lowest of (1) the physician's billed charge for the service, (2) the physician's customary charge or the physician's median charge for the service over 12 months, or (3) the prevailing charge for that service in the geographic community. CPR was criticized as inherently inflationary, inciting physicians to continually increase their charges. Moreover, CPR perpetuated distortions in charges by providing better insurance coverage for tests and procedures than for evaluation and management services such as office visits.

Eventually, payers came to view predetermined payment maximums as a preferred approach. Initially, from the 1960s through the 1980s, payers based relative value units on prevailing charges in various markets, as with the California Relative Value Scale. Rather than rely on charges that may not reflect the underlying resource costs of providing services, Medicare's physician fee schedule, introduced in 1992, is based on estimates of covered services' relative resource costs, the value of physicians' work as measured by time and service intensity, and professional liability costs. These resource costs are adjusted for differences in input prices for goods and services in different markets, then the total is multiplied by a standard dollar amount—the conversion factor—to arrive at the payment allowance. Most U.S. payers base their own fee schedules on Medicare's, although they generally use different conversion factors; payers then sometimes modify actual fees based on price negotiations with individual practices.

Typically, fee schedules pay retrospectively for one-time services—a procedure, a test, an office visit. However, some fee schedule codes are forms of capitation (e.g., payment for a month of dialysis-related professional services) or are episode based (e.g., payment for a 90-day “global” period of postsurgery routine care, a month of complex chronic care coordination).

Key Objectives

Fee schedules for professionals, including physicians, promote professional activity in general and specific professional activities in particular by providing generous payments for services payers intended to encourage. In many national health systems and throughout the United States, fee schedules are the foundational approach on which other payment methods are based.

Strengths

- In contrast to payments based on physician charges, a fee schedule gives payers more control over payment, offers predictable payments, and counters the inevitable inflationary effect of UCR-based payment methods.
- Fee schedules reward activity and industriousness and promote patients’ access to care because providers get paid more for doing more.
- The approach is consistent with how transactions are conducted in retail markets, so payers can rely on consumers’ and patients’ discipline with cost-sharing to affect service use and prices.
- Fee schedules are well established, with well-described impacts; specific reform proposals have been made to improve fee schedule functioning and performance.
- Theoretically, the approach can encourage desired behavior by paying more to encourage or less to discourage provision of particular services.
- A fee schedule implicitly adjusts for the different case mixes different clinicians and practices experience, thereby paying comparatively more for sicker patients that need more services.
- The approach provides payers with data about patient care, which can then be analyzed to establish performance measures or used for other purposes.

- Fee schedules can accommodate elements from other payment reform approaches that are similar to capitation or episode-based payments while also permitting targeting of particular services. The approach does not require adoption of a full, fee schedule replacement approach.

Weaknesses

- Fee schedules encourage overprovision of services, because clinicians often determine the need for services and can induce patient demand.
- The method ignores whether the service was appropriate or performed well; payment is provided for activities, not for outcomes. Indeed, even inappropriate or poorly performed services that generate need for additional services are paid.
- Fee schedules can contribute to care fragmentation, as fee schedules provide no inherent incentive for providers to coordinate care.
- Fee schedule payments generate a large number of billable transactions; this in turn generates high administrative costs for health professionals.
- Activities not codified and covered for payment in a fee schedule may be marginalized. In fact, many activities clinician practices perform are not recognized for payment because transaction costs exceed the value of the services or because the payer has difficulties assuring the services were actually performed.
- Coding complexity, with U.S. payers relying on more than 8,000 codes, makes fee schedules susceptible to “gaming” or outright fraud.
- Payers must make major effort to keep the list of recognized services and their associated fees current, reflecting technological changes and work process improvements that alter relative resource costs. Without that effort, relative fee levels distort professionals’ use of time and the mix of services they provide.
- No data are currently available from which to determine relative values for services; current fees rely on flawed estimates of work and practice expenses that somewhat reflect clinicians’ self-interest. Clinicians who help payers set relative values seem to overvalue tests and procedures, while undervaluing time spent with patients in office visits and other so-called cognitive activities.

Design Choices to Mitigate Weaknesses

Most payers using fee schedules must decide how to counteract the inherent incentives for providers to continually increase service production, thereby increasing costs. Medicare has attempted to establish macro-level expenditure limits that would reduce pro rata fees when a target level is exceeded.

Medicare's sustainable growth rate (SGR) mechanism (in place for more than 15 years) seemed to give clinicians a perverse incentive to increase volume of services, even as the collective interest would have been to restrain service production. In the face of what would have been major formula-driven fee reductions, Congress repealed the SGR in 2015.

An alternative approach would be for payers to modify individual fees to more closely approximate underlying resource costs. In the past, private payers have relied on Medicare's relative value scale to set fees. Yet, Medicare's relative fees are generally thought to exhibit payment distortions, overvaluing tests and some procedures and underpaying activities provided by primary care physicians and so-called cognitive specialties. Other payers can more actively participate in the rule-making process that determines Medicare fee schedule payment rates or can on their own attempt to modify relative values, although these payers would be negotiating with practices that have a financial interest in resisting such modifications. The market area's particular practice environment would likely affect how successful such a strategy would be.

Payers, including Medicare, have recently recognized they can create new fee schedule codes to reward evaluation and management activities that had never been specifically paid, including complex chronic care management and activities related to patients' transitions from hospitals to community-based or other postacute settings. Paying for some important services (e.g., routine phone calls and e-mail communications) on a fee schedule is challenging, because the transaction costs of billing and receiving might be more costly than the service itself. However, a range of other activities might be amenable for inclusion on a fee schedule.

Some elements of value-based payment can actually be included on a fee schedule. To address problems such as "upcoding" or outright fraud, payers can consider reducing the granularity inherent in the CPT coding system, which unintentionally promotes complexity and encourages providers' gaming to achieve higher payment, by not fully recognizing current fee differentials for marginally more complex services. Payers could also reduce gaming by "packaging" some low-cost ancillary services into the other services for payment purposes, thereby reducing providers' incentives to perform overvalued tests, as is done in the Medicare outpatient hospital payment system.

In sum, fee schedules could better recognize “value” through consideration of coding changes, greater accuracy in establishing relative values, reduced coding granularity, clearer coding rules, and other improvements.

Finally, some have proposed that a more direct approach to fee schedule design would add value. Instead of basing relative value units—and fees—only on resource costs, fees could be based on policy judgment. In other words, fees would be modified so that health professionals would change the mix of services they provide, with the goals of producing high value mix of services and altering how clinicians spend their time. However, Medicare would find changing fee levels to produce higher value politically challenging, with difficulty achieving consensus. Private payers modifying fees to accomplish a higher-value service mix would still be subject to market negotiations, with physicians sometimes able to prevent changes that would alter their fees.

Compatibility with Other Payment Methods and Benefit Design Options

Fee schedules are commonly a foundation for other payment methods because they are in such broad use. For physician payment, only capitation approaches represent a rejection of fee schedules as the base payment (even though an organization receiving global capitation may itself distribute payment to its constituent members through fee-schedule-based productivity metrics).

Fee schedules could be combined with capitation and pay-for-performance or included in other hybrid approaches, as adopted in other countries such as Denmark and the Netherlands. The hybrid fee schedule/capitation approach attempts to balance overuse and underuse incentives to approach payment neutrality, while still paying physicians their rough variable costs for additional fee-schedule services. For example, a hybrid payment system could pay primary care physicians 70 percent of a revalued, more accurate fee schedule and 30 percent capitation—with some element of public reporting and possibly payment for performance.

Fee-for-service is compatible with many benefit design options that rely on greater or variable cost-sharing. These designs provide consumers the choice to forego services they think unnecessary, with all the potential strengths and weaknesses of cost-sharing as a cost-containment strategy. Indeed, benefit designs that encourage consumers to shop prudently for physician services assume fee schedule payments—patients’ cost-sharing obligations are based on the prices associated with fee schedule services.

The Focus on Performance Measurement

Measures of clinical appropriateness are desirable but few are available, largely because claims data (i.e., the data payers generally rely on to construct performance measures) lack the clinical nuance needed to assess appropriateness. This is especially true for the many services for which appropriateness relies on individual patients' characteristics, including their personal preferences. Given that basic fee schedule payments are agnostic about quality—they pay regardless of how well the service was provided—quality measures could well complement fee schedules, such as with the recently enacted Merit-Based Incentive Payment System (MIPS) for physicians in traditional Medicare. The pay-for-performance (P4P) strategy for improving quality and value for physicians (and other providers) remains controversial, with evidence of its effectiveness still unclear—although it is being adopted by many payers, fostered by Medicare's initiative.

Potential Impact on Provider Prices and Price Increases

The existence of fee schedules does not mean that prices in commercial insurance markets are necessarily consistent across either payers or individual providers. In fact, evidence suggests fee-schedule prices vary widely both across and within markets, from as little as 70 percent of the Medicare rate in some markets to more than 500 percent for some large practices in other markets. Analysts believe the variation occurs because different physicians and insurers having different leverage in their negotiations, which in turn may be attributed to factors such as an area's level of competition and a hospital's or physician practice's reputation.

Primary Care Capitation

Capitation is a prospective unit of payment per patient, per month or year, in which a payer makes a fixed payment for a defined set of services, regardless of the quantity of services actually provided. This payment approach can be used for an individual health professional, for a group of health professionals for their collective professional services (“professional capitation”), or for provider organizations to assume risk for most health services (“global capitation”). Primary care capitation was a commonly used payment method in the 1980s and 1990s and still persists (although less commonly) today, associated with the rise and decline of health maintenance organizations (HMOs) as a common form of managed care. Primary care capitation required insured individuals to select one primary care physician (known variously as a “gatekeeper” or “primary care case manager”) both to provide of routine care and to approve referrals for other nonemergency health services.

Actuarial concerns complicated the early capitation models that estimated rates based on an actuarial analysis of fee-for-service claims; arguably, serving as a gatekeeper requires providers to expend additional effort that was not recognized under fee schedules and was therefore not included in their payments. Also, primary care capitation methods typically adjusted payments for a patient’s age and gender but not for health status, leading to a mismatch between patients’ needs and capitation payments.

Many states, concerned about the incentive to deny needed services under primary care capitation, consider the method to constitute risk bearing. These states thus restrict primary care capitation for use within an HMO structure, but not in other products, including the more common and growing preferred provider organization (PPO) insurance model.

Key Objectives

The theoretical virtue of primary care capitation is that it permits primary care physicians themselves to decide what mix of activities best serves each patient, rather than rely on third-party payers to approve payment codes and payment levels to influence how clinicians spend their time. Primary care capitation places decision-making in the hands of health professionals who may be in a better position than distant insurers to act in patients’ best interests. In addition, in direct contrast to fee schedules, primary care capitation in effect establishes spending limits for the patients a physician is responsible for, thereby creating financial incentives in favor of activities that reduce spending.

Strengths

- Primary care capitation places “performance risk” on clinicians, providing them financial incentives to limit provision of unneeded services.
- This payment approach internalizes to the primary care physician decisions over the allocation of activity and costs, permitting more flexibility in individualizing care to meet patients’ needs.
- Payers can support newer forms of communication and care delivery that substitute for the traditional office visit, such as e-mail and telehealth, much more easily and prudently through capitation than under fee schedules.
- The approach gives payers predictable and capped costs, while providing the recipient clinician a predictable cash flow.
- Primary care capitation is administratively straightforward (although design approaches to address its weaknesses can add substantial complexity). At its simplest, providers receive payment every month for a roster of patients, obviating the need for billing and paying for each service provided (as under fee-for-service).

Weaknesses

- Primary care capitation has traditionally been adopted in HMOs, which can restrict patients’ choice via the primary care physician route.
- HMOs are subject to state regulatory oversight and therefore can apply primary care capitation. Self-funded employer plans are not subject to state regulatory oversight so they could use primary care capitation, but as a practical matter they typically contract with available state-regulated HMOs (capitation permitted) and PPOs (not permitted) and go along with the payment method the plan administrator uses.
- Primary care capitation may lead to stinting on care—particularly care that can be avoided without compromising the patient’s well-being in the short term (e.g., disease screening and prevention services).

- In the absence of risk adjustment for health status, primary care physicians can “cream-skim”—that is, shun sicker, costlier patients that would take up more time and resources in favor of healthier ones for whom payment would be the same.
- The approach creates an incentive for primary care physicians to refer their patients to other physicians for services outside the scope of the capitated payment. For example, primary care clinicians under a capitated payment may choose to refer patients to specialists, who may be paid fee-for-service, rather than caring for them directly, fragmenting care and raising total costs to the payer.
- In a pure primary care capitation model, payers can’t use a fee schedule or P4P to promote activities or services they want to encourage.
- A single monthly capitation payment has little transparency to reveal clinicians’ activities, making performance measurement, performance assessment, and risk adjustment for health status virtually impossible.
- Payment per capita provides clinicians a financial incentive to take on too many patients, exacerbating concerns about stinting and overreferral.
- Conversely, primary care capitation assumes statistical averaging of patients with different health care needs, so a minimum number of patients is needed for it to work correctly. Payers may need to maintain a fee-for-service program in parallel with capitation, adding administrative complexity.
- Direct payments to primary care providers represent a small percentage of health care spending—5 to 6 percent. Unless the design includes a strong incentive system for rewarding or penalizing total health care spending attributable to the physicians’ patients, this approach might not be worth the implementation effort.

Design Choices to Mitigate Weaknesses

Payers can design primary care capitation to mitigate stinting by measuring performance of preventive services. Yet that requires physicians to submit information typically contained in fee schedule claims, counteracting the simplicity capitation offers. Encounter data would allow payers to perform risk adjustment of capitation payment levels, as well as to assess quality and access to care.

Primary care capitation gives physicians strong incentives to refer patients to providers outside the purview of their capitation payments. To address this, payers could impose financial penalties for excessive referrals and downstream health care spending. Such an approach would help justify the administrative effort to introduce a capitation method only for a subset of physicians. In the past, primary physicians under capitation were accountable for health spending for patients in what were called risk pools. Capitation “withholds (e.g., 20 percent of the monthly capitation amount) could be returned if the risk pool showed a surplus, a form of the shared savings approach now being used for accountable care organizations (ACOs).

Payers may impose ceilings on the number of individuals physicians can have on their rosters, to counteract the incentive to overexpand the physician’s patients. Cream-skimming can be addressed through risk adjustment, but that would require physicians to submit encounter forms with claims-like detail; moreover, experience suggests encounter data that is not part of a payment claim may be less reliable.

Payers can encourage performance of particular services by making them targets of a complementary P4P program, or “carving out” services from the capitation package (e.g., immunizations, for fee schedule payment). Most generally, primary care capitation might be placed with some amount of fee schedule payments in a mixed payment model to balance the incentives of the two approaches.

Compatibility with Other Payment Methods and Benefit Designs

Capitation is typically used only by HMOs because only HMOs can use a primary care physician system in which a patient selects a single physician or group for services and for access to specialty care.

As discussed under fee-for-service, a hybrid of primary care capitation and fee schedule payment, as well as incremental payments such as shared savings and P4P, are all compatible—and in some contexts, probably desirable. This hybrid approach softens the polar financial incentives of capitation and fee-for-service but adds complexity to the payment approach.

Flat co-payments for office visits can be accommodated under primary care capitation, but high-deductible plans and prepayment through capitation would seem incompatible, at least from the standpoint of the primary care physician trying to manage total costs of care. However, high patient

cost-sharing reduces patients' demand for services, which would actually result in less demand on physicians' time and practice resources. But the payers' actuaries would realize this reduced spending and would likely seek to lower capitation levels. Value-based insurance design (V-BID), which reduces or eliminates cost-sharing for certain high-value primary care services, would counter the incentive for physicians to stint on these particular services.

Prior authorization by primary care clinicians is a natural complement to this payment approach and, indeed, is the clinicians' responsibility when they have the formal role of gatekeeper to elective services provided by other providers. Capitated primary care physicians might welcome precertification by health plans for specialty-generated care if they have risk pools or other two-sided risk incentives.

The Focus of Performance Measurement

HEDIS-type performance measures are strong in the areas of screening and primary and secondary prevention services, and they can be useful to identify stinting on these services.¹ Because concerns about appropriate referrals are central to assessing performance, measures related to referral rates would be desirable, as would rates of emergency room and hospital admission and readmission.

Potential Impact on Provider Prices and Price Increases

Capitation rates are typically calculated based on actuarial analysis of past experience, so that the rates represent a community average rather than historic costs for individual practitioners or practices. In this way, capitation payments do not reflect provider-specific pricing differentials. However, physician practices with negotiating leverage can, nevertheless, achieve higher capitated amounts that deviate from the community average, effectively passing through higher prices and desires for higher-than-average price increases in their capitated rates.

Per Diem Payment to Hospitals for Inpatient Stays

Per diem payment for inpatient services provides a fixed amount for a patient day in the hospital, regardless of a hospital's charges or costs incurred for caring for that particular patient. In the most common arrangement in the United States, the payer negotiates per diem rates with the hospital and pays that rate without adjustment. If the payer and hospital can accurately predict the number and mix of cases, they can accurately calculate a per diem rate. All else equal, the larger the volume of cases applicable to a payer, the more predictable the average daily cost—and the per diem level—will be.

Often, however, hospitals want to exclude days in an intensive care unit or another specialized unit, unless there is a sufficient volume of regular medical-surgical cases to make the reimbursable costs predictable. Accordingly, multiple per diems are often negotiated on the basis of service type (e.g., medical-surgical, obstetrics, intensive care, heart surgery). Service-specific per diems diminish payers' need to formulate outlier provisions for unusually costly patients.

Per diem payment is often subject to carve-outs for particularly high-cost items and services, such as surgical implants and expensive drugs. The costs for these items can be passed through, sometimes with a markup for the hospital. Payers may also offer differential per diems for different days in the hospital; for example, the first day of a surgical stay would be paid at a higher rate than subsequent days, such that the later days are paid closer to the variable cost of the day rather than the average cost of all days.

Subject to a negotiated contract is whether the payer is obligated to pay for all days, regardless of their medical necessity. Private insurers are typically able to deny payment for days they deem unnecessary, using length of stay guidelines and performing either concurrent or retrospective review of clinical records to determine whether the billed days are medically necessary. Until recently, most U.S. insurers preferred per diems to DRG-based case rates because of their ability to deny days at the end of a hospital stay.

Recently, payers and hospitals have found DRG-based payment methods attractive because of their much stronger incentives and rewards for shorter stays and reduced costs. Yet some providers, especially ACO-like organizations that actively manage both who gets into the hospital and how they receive care, prefer to use per diems because they can directly control length of stay and do not need to provide hospitals payment for the average length of stay on which DRGs are calculated.

In some countries with health systems that rely on global hospital budgets, payers have combined budget targets with per diems as the billing units to provide cash flow.

Key Objectives

Per diems represent an administratively straightforward way of modifying the inherently complex and inflationary approach of paying for each individual service hospitals provide. As a readily calculated metric, per diems provide straightforward payment negotiations between payers and hospitals. Per diems offer consumers the potential for cross-hospital cost comparisons, if such information is made transparent to the public (although hospital-specific variations for different service lines compromise that potential).

Strengths

- Per diems, over more than 30 years, have led to straightforward administration and contracting. This payment method has facilitated administrative standardization, with supporting software to facilitate coding and billing.
- Per diems provide some constraints on cost-generating hospital behavior, because the payment amount per day is prospectively set (while the total actual payment is retrospective). Although hospitals have an incentive to generate longer stays to secure additional paid days, they should want to hold down costs per day.
- In contrast to both itemized individual services on the one hand and DRGs on the other, per diems can provide greater transparency for consumers to compare prices and lengths of stay among hospitals, as a surrogate for overall hospital costs.

Weaknesses

- Hospitals have no incentive to avoid unnecessary days during a hospitalization.

- Per diems do not provide much transparency about hospitals' actual clinical activities, in contrast to DRGs. Thus, although per diems may facilitate length of stay comparisons, they do not permit comparisons among hospitals on activities or on outputs produced.
- Efforts to control costs may require third parties that monitor per diems to determine medical necessity through aggressive "continued stay" medical review. This introduces administrative complexity and sometimes inappropriate intrusion into clinical care, especially if the length-of-stay criteria are not supported by strong evidence.

Design Choices to Mitigate Weaknesses

The most common approach to addressing the incentive for unnecessary admissions and longer-than-needed stays is through precertification of hospitalization and concurrent or retrospective medical review to deny medically unnecessary days. Payers will have to weigh the benefits of this regulatory approach and perhaps vary its application depending on their assessment of a hospital's proclivities toward excessive stays.

Operationally, whether per diems are calculated for different service units or service lines can vary. Payers can adopt a sliding-scale per diem approach: they will either pay a lump-sum settlement at the end of the year or withhold an amount from the final payment for the year to effectively reduce the per diem, depending on total bed days or admissions in the year. These reconciliations could be made more often than annually if hospital volume is high enough. And payers can attempt by contract to reduce how much hospitals benefit financially from payments for pass-through costs and from outlier payments.

Compatibility with Other Payment Methods and Benefit Designs

For many decades, per diems had become the dominant approach for payer-hospital contracting for inpatient services because of relative administrative simplicity and modest incentives to reduce hospital costs per day, if not per stay. Starting in the mid-1980s with an inpatient prospective payment system, however, Medicare introduced an alternative with stronger cost-containing incentives because it provided hospitals with a fixed payment for the entire hospital stay. Payers able to deny days through

continued stay oversight tended to prefer per diems, although some evidence shows that contracting has moved more toward case rates using DRGs, to be compatible with Medicare.

With the current growing interest in global payment, there is a rationale for medical groups bearing risk to prefer per diems to DRGs. Medical groups at risk for inpatient hospital services now have direct interest not only in avoiding inpatient care through more vigilant and higher-quality ambulatory care, but also in using less-intensive and less-costly sites of service. In addition, medical group personnel can actively manage their patients as inpatients and accomplish early discharges supported by strong transition programs, rather than rely on the hospital and separate clinical staff, for a high-quality, “early” discharge, thereby addressing the incentive for unnecessarily long hospital stays.

Per diems are compatible with procedure-based, bundled episodes for care provided during an inpatient stay, because providers can benefit from reduced stays. In contrast, paying a full DRG amount even when length of stay is reduced makes it more difficult for the parties receiving the bundled episode payment to generate savings for themselves. As with global payment, providers receiving a condition-specific bundled episode payment might prefer per diems to DRGs. Per diems impart a direct interest in controlling the use of and the length of inpatient hospital stays, and so providers might spend less under per diems than under DRGs.

Per diem payment for inpatient services is compatible with benefit design approaches that limit consumers’ access to services through precertification and continued stay review, with the attendant strengths and weaknesses of that form of utilization management. Some commercial payers and hospitals have shown interest in instead moving toward DRGs, suggesting that they would rather rely on DRGs’ stronger cost-reducing incentives than on a “regulatory” approach of denying days they pay for.

The Focus of Performance Measurement

As noted, per diems give little information on clinical activities a hospital performs, particularly not the output per unit cost and the distribution of clinical services by condition. Separate data must be collected for measurement. Hospitals have incentive under per diem payment to generate both inpatient admissions and longer-than-necessary stays. Thus, measures of hospital-specific admissions for ambulatory care-sensitive conditions and of preventable readmissions would be particularly useful to assess—especially to see how hospitals are responding to perverse effects.

Potential Impact on Providers' Prices and Price Increases

Nothing intrinsic in the per diem payment approach affects hospitals' prices or their incentives to increase prices. Hospitals with pricing power in their negotiations can seek higher per diem rates beyond their actual costs and can sustain a high cost structure. They also are in a position to avoid the cost discipline imposed by per diems) by having service lines and outlier cases revert to payment of charges or discounts off of charges, rather than the negotiated per diems, resulting in higher payments.

Diagnosis Related Groups–Based Payment to Hospitals for Inpatient Stays

Background

Diagnosis related groups (DRGs) provide a flat per-discharge (or per-death) payment that varies based on diagnoses, severity, and whether and what procedures were performed. DRGs are used for two purposes: In some systems, DRGs are a measure for assessing hospitals' case mixes and activities. In other systems, including Medicare, DRGs are used as an additional payment method. The basic setup for DRG-based hospital payment includes the following elements:

- a patient classification system to group patients with similar clinical characteristics and relatively homogeneous resource consumption into hundreds of DRGs;
- hospital cost information used to determine DRG weights, usually based on relative average treatment costs of patients falling within each DRG;
- a standard monetary conversion factor, used to convert DRG weights into base payment rates for each DRG;
- actual payment rates, obtained by adjusting the DRG base rates for structural differences across hospitals (e.g., wage rates, teaching status, rural area designation) and further resource-consumption variables (e.g., length of stay, readmissions, use of high-cost drugs or services).

Hospitals are paid based on the number and the type of DRGs they produce. The approach assumes that hospitals treat a random variation of patients such that, on average, patients who are more costly than their DRG payment rate are offset by patients who are less costly. Creating more DRG categories to reflect severity differences would decrease perhaps-unrealistic assumptions about random variation, as this would more accurately account for the systematic variations in costs associated with different clinical conditions.

Different payers adopt different DRGs to affect the actual incentives hospitals face. These differences do not alter basic incentives but rather represent operational differences. For example, across countries and payers most DRG weights and payments are based on average costs. However, it is possible to introduce normative rather than empirically based standards for modifying empirically derived weights thought to distort behavior.

Most DRG payment systems include outlier payments as insurance against incentives to avoid or prematurely discharge costly or potentially costly patients (called “outlier cases,” based on length of stay or actual computed costs). Outlier payments also protect hospitals from losses related to a “bad draw” of exceptionally costly patients relative to their DRG payment rates. To prevent a skew in calculation of average DRG costs, most DRG systems exclude outlier cases from the determination of average costs and provide separate outlier payments—these payments usually kick in only after a cost or length of stay threshold, generally far higher than the average for the DRG, is reached. Outlier payment therefore reflects to a limited extent the actual cost incurred by the hospital for extreme cases, rather than the cost of an average case, to balance the cost-containing objectives of DRGs with practical concerns about payment fairness.

Key Objectives

Medicare adopted DRGs as an alternative to so-called cost-based payment to fundamentally change hospitals’ incentives to reduce costs associated with an inpatient stay. Given that a prospective payment based on a patient’s principal diagnosis, the hospital has an incentive to eliminate unnecessary services and to reduce the length of stay. In contrast to the United States, many developed countries introduced DRGs not as a replacement for cost-based reimbursement, but rather as a substitute for hospital global budgets to promote and reward hospital activity. Under some forms of global budgeting, hospitals with a guaranteed budget could adopt a complacent attitude about attracting patients, thereby producing queuing or waiting periods for elective services.

Strengths

- Because the payment amount per principal diagnosis is fixed, hospitals have strong incentives to reduce costs per stay.

- Payers can achieve savings over time because hospitals' responses to DRG incentives lower average costs per case, which in turn permits lower DRG payment levels.
- Hospitals may improve care quality because they will typically improve internal care pathways and reduce lengths of stay (longer stays can be associated with greater iatrogenic harm and hospital-acquired infections).
- DRGs may be more market-oriented than other hospital payment systems because hospitals may improve quality and efficiency by treating patients for which the hospital has a competitive advantage.
- Having a uniform, standard classification system facilitates transparency and permits interhospital comparisons by payers and consumers.
- DRGs eliminate the need to review the appropriateness of every service provided during a patient's stay, so monitoring can focus on the appropriateness of the stay.
- Most health systems and an increasing number of U.S. payers now use DRGs; new approaches to promoting quality and cost containment can be transferred into improved DRG model designs.
- Hospitals paid under DRGs by Medicare would see a common payment model if private payers adopted the same approach, thereby eliminating conflicting incentives with per diems (the predominant method of hospital payment used by insurers).

Weaknesses

- With a fixed payment per case, hospitals retain an incentive to increase the number of patients hospitalized, even when outpatient management is acceptable or preferred.
- Hospitals benefit from increasing revenues per patient, most easily achieved by changing coding practices of diagnoses and procedures ("DRG creep") or by providing services that lead to reclassification of patients into higher-paying DRGs.
- In comparison to other methods for paying hospitals, DRGs are more complex, requiring coding expertise, data systems, and active oversight of coding by payers.

- In commonly used DRG designs, performing a surgical procedure produces a substantially higher payment net of cost for the same diagnosis without a procedure. Thereby, clinical decision-making is potentially skewed to favor procedures when medical management might suffice.
- Hospitals have an incentive to select profitable, low-cost patients (“cream-skimming”) in each DRG and transfer or avoid unprofitable, higher-cost patients.
- Hospitals may discharge prematurely, compromising quality yet rewarding hospitals if the patients are readmitted (unless the DRG design does not permit a new payment for readmission within a specified time period, e.g., 30 days).
- Hospitals may transfer patients to other hospitals or postacute care facilities, generating overpayments from the artificially low length of stay. Payers can follow Medicare’s lead by reducing the payment when such a transfer occurs early in a hospital stay, though that adds yet more administrative complexity.

Design Choices to Mitigate Weaknesses

Many design issues determine the precise incentives hospitals experience under DRG payment systems. Innovative approaches have been developed over time and can be adapted for broader use:

- A transfer policy, under which short-stay discharges to another facility result only in a partial DRG payment, with the amount prorated length of stay. Similarly, DRG payments might not be made for very short stays, called “observation” stays in the United States, with the patients considered as outpatients. (However, the U.S. experience with observation days suggests this approach has its own problems and remains a work in progress.)
- Separate payments might be made for certain services that should not be assigned to specific DRGs because doing so would discourage their provision (e.g., costs associated with teaching).
- As a response to the broad incentive under DRGs to increase admissions, payers can set quasi-hospital budgets or volume thresholds that put predetermined limits on payments. Or, for volume increases above the baseline, payments can be reduced to approximate variable, rather than average, hospital costs.

- DRGs impart incentives for premature early discharge, often compromising quality and increasing readmissions. Accordingly, payers can forgo making a new DRG payment for patients readmitted for the same problem within a DRG-specific duration after discharge (as Germany). This form of warranty policy has a few potential variations.
- Payers can continually recalculate both DRG weights and monetary conversion factors to prevent DRG payments diverging substantially from underlying costs of production; such cases distort behavior, producing DRG “winners” and “losers.” One proposal would adjust the payment rate for a DRG when volume increases sharply—an indicator that the payment rate for that DRG may be excessive.
- Quality might be improved if payers deny placement into the higher paying DRGs for hospital-acquired conditions. Medicare uses this approach with mixed results.
- DRG weights might be modified to encourage desired behavior. That is, payments can be adjusted to provide a small “penalty” for providers performing a procedure associated with patterns of inappropriate care.
- To promote greater efficiency and enhance quality, payers might base DRG weights not simply on actual cost allocation in the different DRGs. Payers also might consider normative standards of care, that is, the cost of care for efficiently produced evidence-based care.

Compatibility with Other Payment Methods and Benefit Designs

DRGs can readily be used in the calculation of bundled episode payments. If a payer defines an episode as the DRG does, which may be the case with hospital-based procedures, the DRG could represent the hospital portion of the bundled payment. In that way, DRGs are more useful than per diems in calculating condition- or procedure-specific bundled episode payments.

DRGs provide hospitals with stronger incentives than per diems to decrease provision of unneeded services and to promote more internal collaboration for efficiency. Yet an organization receiving population-based payments is at risk for the costs of hospitalization. Thus, payers may prefer to contract with hospitals using per diems rather than DRGs, if the payer can directly affect length of stay through its own efforts rather than relying on the hospital.

DRG payment would seem incompatible with various approaches to patient cost-sharing, because it is based on average cost for a diagnosis category rather than a patient's own experience. In some all-payer rate-setting states in the 1970s and 80s, the lack of correspondence between DRG payment amounts and individual patients' experiences created problems with cost-sharing obligations: patients with short stays or low costs, understandably, did not want to pay an averaged amount (while more costly patients benefited).

Variation in individuals' costs based on DRG case averages was considered inequitable. For that reason, Maryland's all-payer system has continued to use payment based on unit charges, in which DRGs serve as a "unit of constraint" but not the actual "unit of payment." The unit payment—approved charges—permits a closer relationship between patients' obligations and their own costs incurred, a particularly significant issue for those without insurance (i.e., "self-responsible" patients).

Many commercial payers, nevertheless, have adopted DRG payments while still using high-deductible benefit designs. For most subscribers, the cost of a hospitalization exceeds their out-of-pocket maximum. Even a short stay will blow through the deductible, whether the subscriber's portion is calculated based on the DRG's average cost or on the patient's actual experience. In practice, insured subscribers have seemingly accepted their cost-sharing portion of an average-priced DRG without much objection, perhaps because hospital prices' complexity and lack of transparency obscure the inequity in patient's obligations with DRG payments. DRGs remain problematic for self-responsible patients because of the mismatch between average case payments and the patient's actual experience.

The Focus of Performance Measurement

Many other countries have experienced an increase in hospital admissions in response to DRG incentives. This is likely because most other countries are all-payer systems, so they have no "safety valve" ability to make up revenue shortfalls through other payers, as is common in the United States. Accordingly, measures of appropriateness of hospitalization would help payers monitor hospitals' responses to DRGs, but these are not broadly available. More simply, payers can measure readmission rates (as is being done under Medicare) with financial penalties applied to hospitals with high readmission rates.

Although readmission rates would seem a straightforward measure to use, experience suggests even the validity of this measure is open to question. For one, reducing readmissions may also reduce admissions, such that the calculated readmission rate may miss the reductions in both the numerator

and the denominator. What otherwise would have been a readmission can be redesignated as an observation stay to keep readmission rates low.

It also would be desirable for payers to measure premature hospital discharges, but such measures are not available currently.

Potential Impact on Provider Prices and Price Increases

DRG-based payments have no inherent incentives that counter the market power of “must-have” hospitals. Medicare has the power to set DRG payment rates. Yet for private payers, hospitals with pricing power can demand higher-dollar conversions with standard DRG weights to achieve higher payments and to demand generous terms for outlier payments. Hospitals also frequently carve out particular high-volume specialized service lines from DRGs, with payment based on discounts off charges to produce greater margins.

Global Budgets for Hospitals

A global budget provides a fixed amount of funding for fixed period of time (typically one year) for a specified population, rather than fixed rates for individual services or cases. The main objective is to constrain the amount a hospital can spend in order to limit the total amount of money spent on health care within the system. This approach contrasts with “line-item budgeting,” which breaks down the amount into specific line items, such as salaries, drugs, equipment, and maintenance. Hospital managers often cannot change line-item allocations without approval from funders (usually a government agency). Essentially, a global budget represents a one-line budget and provides the hospital more management flexibility to allocate resources. Over time, global budgeting has replaced line-item budgeting in developed countries that rely on regulation more than on market forces to control health care spending.

Global budgets for hospitals can be “soft” or “hard.” Under a soft global budget, the purchaser or payer assumes part or all of any overruns. But consensus has emerged that soft budgets are ineffectual. Under a hard global budget, the hospital’s payment is limited to the prospectively set global budget amount, transferring financial risk to the hospital. A core concept is that hospitals would have an incentive to reduce the unit cost of output, so they are able to increase the volume of services provided through the budget. Conversely, if unit costs rise volume must fall, although the payer can adjust the global budget for reasonable, overall volume growth or declines.

A global hospital budget implies that all payers participate and thus is simpler to operationalize in a single-payer or all-payer environment, such as Maryland. Although a global budget concept may be applied to some but not all payers in a jurisdiction, guaranteeing a budget across all payers, changes in volume for nonparticipating payers would put participating payers at risk. The Maryland all-payer payment reform demonstration, which began in 2014, ties payments for individual services to hospital-specific global budgets. This is intended modify prior incentives in the Maryland all-payer rate-setting program, which limited prices but encouraged increased volume.

A global budget can have many forms, but it must be explicit about the services it funds. The global budget needs a precise method for specifying which patients are included and to which hospital’s budget they are to be assigned. Administratively, global budgets sometimes are managed through a contractor positioned between hospitals and purchasers or payers. Spending for items outside the operating budget (e.g., capital, teaching, and research and development) is generally funded outside the global budget. Hospitals may also receive an up-front payment or a temporary addition to their rate

base to support infrastructure enhancements for care management and IT capabilities needed to operate under the global budget.

Hospital budgets are generally set through one of three approaches—historical, capitated, and normative—or some combination of the three:

- In the historical approach, which is the most commonly used, the first-year budget becomes the base and subsequent years' budgets are pegged off that experience. The historical cost approach is the easiest operationally and the most common approach.
- Capitation aims to distribute resources based on relative needs of the populations providers serve. However, capitation requires sophisticated data to perform health-status risk adjustment and advanced modeling capabilities and is not widely used for setting global budgets.
- There are many variants of the normative approach to setting budgets. In essence, they use an external rate-setting approach to set a unit price for services, which is then multiplied by the anticipated or desired volume of services. This approach does not necessarily take account of historical levels of activity, access, or provider costs.

Payers can allocate the global budget directly to hospitals, but also can produce cash flow by making unit payments (as in Maryland), per diem payments, or case-based payments. Subsequently, payers would reconcile the resulting payment totals with the agreed-upon budget amount. This payment approach must allow payments for patients who are not part of the hospital's global budget, such as out-of-area individuals seeking urgent care.

In Maryland, once a hospital's global budget is set (based on the hospital's historical revenue), the payers' revenue contributions are apportioned retrospectively based on each payer's proportion of revenue. (Proportion of revenue, in turn, is based on each payer's charges during the year.) The unit prices adjust up or down to all payers based on the aggregate volumes of patients the hospital serves, to assure the hospital complies with the global budget amount.

Key Objectives

Global budgets give hospitals clear incentives to manage provision of care within a defined budget constraint, emphasizing the policy objective of cost containment. One of the clearest incentives is to

reduce the number of admissions that the global budget must cover; the volume of admissions is an important approach to reducing hospitals' variable costs.

Strengths

- Similar to capitation, global budgeting fundamentally changes the incentives hospitals face, providing a direct incentive to improve operating efficiency and reduce volume of cases, outpatient encounters, and services per patient.
- A hard cap global budget rigorously enforces limits on spending and provides spending predictability for payers and health care policymakers.
- A global budget, in contrast to a line-item budget, provides hospital management with more autonomy and flexibility to improve efficient production of health services.
- A global budget is relatively straightforward for the hospital to administer, and it is seemingly less susceptible to the fraud associated with false or inflated claims for services. Patient cost-sharing obligations can be included under a hospital global budget if cash flow is based units of service or per diems.

Weaknesses

- Global budgeting does not apply readily outside of an all-payer or single-payer environment.
- Global budgets do not promote competition among hospitals or reward hospitals for growth in market share, unless they include a mechanism to adjust a hospital's budget for shifts in volume due to desirable changes in referral patterns. (In Maryland, this mechanism is referred to as a market-shift adjustment.)
- Without specific performance incentives and assessments, hospitals under global budgeting can operate within their budgets by limiting spending, even if the spending reduction approach might negatively affect access and quality. In Europe, some hospitals have responded to global budgets by producing queues for elective services.
- The common, historical-basis approach to budget setting reinforces existing resource flows, which may not accurately reflect need or market value.

- Payers may base allowances for annual budget increases on factors unrelated to health, such as the growth in inflation or GDP, or on budgetary constraints outside of the health care sector, thereby eroding the global budget's purchasing power.
- A normative approach presumes the global budget is based on highly granular and accurate data, which may be more aspirational than real in many situations. Too much divergence from historical spending may cause real financial hardship for affected hospitals, which can compromise quality and access to care.

Design Choices to Mitigate Weaknesses

The global budget's fixed spending limit must contain a framework to ensure that volume levels are achieved or adjusted for, quality standards are met, and changes in service delivery are consistent with payers' wishes. Without such mechanisms, a global budget can become, in effect, a block grant to be spent at the hospital's will, with no guarantee patients will receive needed services. In particular, payers must assess changes in volume of services provided; hospitals operating under fixed or guaranteed budgets may become complacent about competing for patients or providing needed services. This approach to assuring acceptable performance would substitute oversight of hospital activity for costly and cumbersome claims-based oversight. To assure adequacy, budgets can be adjusted fairly easily for shifts in market share. For example, the Maryland demonstration approach rewards hospitals for increased market share, at a rate meant to approximate each hospital's variable costs. At the same time, the rate is limited by the extent of decrease in admissions in a hospital's area.

Payers may incorporate elements from historical, capitation, and normative approaches to setting global budgets to maximize the advantages of each, although starting with historical budgets will minimize initial disruptions. Mixed approaches can include some element of P4P based on achievement of a few high-level performance targets, such as reductions in average length of stay, readmissions, hospital-acquired conditions, and patient safety measures, to ensure providers do not simply reduce quality and access to adhere to their budgets. Several countries have adopted DRGs not as a payment mechanism, but rather to increase transparency of hospital activity and to permit comparison across hospitals under global budgets.

Compatibility with Other Payment Methods and Benefit Designs

Global budgets are theoretically compatible with population-based payment approaches, such as shared savings, because they remove hospitals' incentive to increase the volume of services.

Hospitals with global budgets should have cost-sharing approaches compatible with their cash flow approaches. For example, when cash flow is provided through DRGs, first-day deductibles are compatible; however, co-insurance amounts would not be calculable because the average case might be different from an individual patient's experience. In contrast, when cash flow is based on permitted service unit charges or per diems, co-insurance or co-payments for services are possible. Similarly, payers can implement high-deductible plans more readily when cash flow is based on per diems or unit charges than when it is based on DRGs.

Generally, the primary mechanism for cost containment under global budgets is on the provider side, not on consumer demand, with the strong incentive providers have to reduce spending. On the other hand, payers can use patient cost-sharing to split the cost of hospital services with consumers rather than to discipline hospital spending. If anything, absent complementary DRGs and quality measures to monitor and assess hospital activity, under global budgets hospitals already have a strong incentive to admit only patients with real needs for inpatient care. In contrast, other approaches to hospital payment, one way or another, reward hospitals for increasing their volume of cases or services.

Global budgets may be less compatible with certain insurance benefit designs, such as tiered networks. Higher-cost hospitals with static global budgets will realize benefits as patients in tiered network products self-refer to lower-cost hospitals. Given the already strong incentives for cost reduction under global budgets, prior certification and continued stay review would seem to have little role as a benefits management approach.

The Focus of Performance Measurement

The global budget is oriented toward reducing service volumes and costs, but explicit measures of quality and access to services are needed. Whereas measures for ambulatory care-sensitive conditions, emergency room use, and hospitalization rates may be needed under volume-based payment methods, with global budgets the concern would be overaggressive reductions in rates on these measures to decrease workload. That is, it would be useful but operationally challenging to measure inappropriately

denied hospitalizations. But there are measures that could be used to indicate access problems; in Europe, for example, wait times for elective admissions are routinely measured under hospital global budgeting.

Potential Impact on Provider Prices and Price Increases

As emphasized earlier, a global budget can only work as envisioned in an all-payer context such that the actual payments conform to the budget. In an all-payer system, prices are subject to administrative price setting or limits – ceilings and floors -- on the prices hospitals can charge. The impact of global budgets on commercial insurers' prices will vary based on whether the budget-setting approach is historical, normative, or capitation based. Historically based global budgets incorporate providers' current price structure in the base (absent any dramatic changes in their mix of services). Applying an allowed percentage increase, rather than a fixed dollar amount, for all hospitals in subsequent years, also would incorporate pricing differences already in place.

Bundled Episode Payment

With the bundled episode approach, a prospective payment is made for all care a patient receives over the course of a defined clinical episode or period of management, instead of for discrete services (as with a fee schedule) or for all care a patient receives (as in global capitation). The episode of care has two dimensions: a clinical dimension, which can represent either the set of services or the clinical conditions that compose the episode, and a time dimension that reflects the beginning and the end of the episode. In essence, the approach is designed to transfer financial responsibility for the technical risk (i.e., risk related to care production) that is under the included providers' control, but not the probability (or insurance) risk that relates to the burden of illness and injury in any large patient population. The bundled providers—clinicians and facilities—have common financial incentives to control the cost of the bundle, because they keep the savings or bear the cost of overruns if costs differ from the fixed payment.

Here we distinguish bundled episode payment from episode payment. The former refers to payment that covers all care for a defined clinical condition *across* various providers of patient care, whereas the latter refers to the duration of service the payment covers, whether or not provided by a single provider or by providers working together. Even in the United States, with its reliance on fee schedules, individual physicians sometimes receive payment for a care episode, for example, for costs associated with professional services for a pregnancy. However, in this example, hospitals would be paid separately for the facility costs associated with the actual delivery.

Bundling separate payment streams into a single one is a unique challenge with bundled episodes. Extending the length of the episode, for example, beyond a hospital discharge, also can be an important strategy to promote care coordination, depending on the providers included in the episode.

Many, though not all, proposals for bundled episode payments involve the care around a hospitalization. For example, in the Medicare Acute Care Episode demonstration, payment was made for bundled episodes triggered by joint replacements and particular cardiac procedures. Medicare is currently testing other approaches under the Bundled Payments for Care Improvement (BPCI) initiative, which includes models that bundle various combinations of physician and hospital inpatient services during the hospital stay, as well as postacute care services within 30 to 90 days after discharge. Preliminary evidence suggests that extending the episode to the postacute care period produces savings, as home health care is substituted for more costly institutional postacute care and readmission rates decrease.

Recently, the Centers for Medicare & Medicaid Services Innovation Center has launched two additional bundled payment models: comprehensive care for joint replacement, mandatory for all hospitals in the demonstration areas, and an oncology care model that bundles services for patients receiving chemotherapy.

These demonstrations, as well as many private sector efforts, focus on more efficient production of procedures—usually performed on an inpatient basis—but there is also interest in condition-based bundled episodes. In this approach, an episode for a reasonably well-defined chronic condition would be the focus of the payment model; for example, the bundle would include services for patients with ischemic heart disease or diabetes for a period of time, perhaps as long as a year. Compared to procedure-based episodes, bundled episodes for conditions could affect much more health care spending and could create much stronger incentives for care coordination across health professionals and providers. Condition-based bundled episodes also could counter the volume-inducing incentives of procedure-based episodes, as discussed below. However, particular challenges associated with chronic condition-based episodes must be addressed—particularly for patients with multiple chronic conditions.

Key Objectives

The primary objective of a bundled episode model is to promote better coordination among clinicians, hospitals, and other providers. These entities respond to the efficiency incentive of the fixed prospective amount they receive for providing all the patient's services during the episode of care.

In addition, whereas most other payment and delivery reform approaches—including patient-centered medical homes and ACOs—prominently emphasize primary care physicians' role in spending reductions, bundled episodes rely more on specialists (who provide the most costly services). *Procedure-based* bundled episodes provide specialists with an opportunity to assume primary, risk-based responsibility for producing lower-cost care with high quality. *Condition-based* bundled episodes can permit physicians to manage care over extended periods, without creating a bias toward performing procedures, by rewarding managing physicians for avoiding costly complications.

Because the incentives, designs, and operational mechanics of procedure-based and condition-based bundled episodes differ substantially, we consider their strengths, weaknesses, and design features separately.

Procedure-Based Bundled Episodes

Strengths

- Procedure-based bundled episodes internalize the incentive for efficiency to affected providers. The contracted party receiving the bundled payment, often the hospital, earns a higher margin if patients are discharged earlier. Yet the party also bears the financial risk of both readmissions and the cost of postacute care for episodes extended beyond hospital discharge, providing a counter-balance to premature discharge.
- Procedure-based bundled episodes provide an incentive for acute and postacute care providers to communicate and coordinate, to both improve patient outcomes and reduce costs. For common inpatient procedures, a typical area for such cooperation is purchasing equipment and supplies, such as expensive artificial joints, using combined market leverage to negotiate better prices from manufacturers.
- Hospitals would likely develop close relationships with physicians and postacute-care facilities more willing to participate in a cooperative venture, and follow care guidelines to achieve quality and cost targets (again, assuming there is a posthospital portion of the defined episode). Providers in acute and postacute settings are encouraged to communicate about ensuring continuity of care for the patient.
- Bundled episodes can be viewed as partway between volume-based payment and true population-based payment (e.g., global capitation), allowing clinicians and organizations to ease into broad payment reform with increased accountability for quality and costs.
- Bundled episodes are more consistent with the service-line strategies hospitals have adopted over the past decade than with population-based payment approaches. Thus, the approach could be more readily undertaken without major change in business models and complex organizational cultures.
- Bundled episodes are seen as providing an “on-ramp” toward value-based payment for some procedural specialties, comparable to medical homes for primary care physicians.
- Because the approach requires providers to cooperate but not to integrate, it can reduce the potential for provider consolidation that can raise transaction prices.

Weaknesses

- Procedure-based bundling remains firmly a volume-based payment method in that it rewards providers for initiating more episodes. The approach might result in high-quality, efficiently produced, but unneeded procedures.
- Arguably, combining various providers' separate payments into the same bundled payment promotes—and perhaps even heightens—opportunities for providers to increase volume of services. That is, providers together could brand and market their now jointly produced services.
- Incentives to skimp on care or to avoid sicker patients are inherent in any fixed-episode payment approach, if there is no risk-adjustment mechanism that pays for additional services provided to sicker individuals.
- Hospitals and other providers have a logical impulse to narrow their referrals to favored postacute care providers, which might compromise patients' choice of provider.
- Although procedure-based bundling is, in some ways, less of a change from the usual payment approaches, it may lack the simplicity of a fixed capitation payment, generating additional administrative expense to adjudicate claims at scale.
- Determining which specific claims for payment belong in a bundled episode is operationally challenging, especially for posthospital services for patients with multiple conditions. Ensuring that services in the bundled episode are paid one time, and one time only, can be difficult. In Medicare's BPCI initiative, recipients are accountable for all spending by the patient addressing claims-related challenges but raising other concerns.
- Where health professionals and other providers, such as hospitals, remain legally independent, sustaining relationships among multiple providers may be difficult. Of particular concern is that a hospital, typically the dominant cost center in a procedure-based bundled episode, may dominate the collaboration and act in its own interests, which may not be congruent with the others'.
- Relatively few procedures may be amenable to a bundled episode approach—payment for less common and for complex procedures be difficult to bundle, thereby limiting the potential impact of the payment method.

- Even assuming procedure-based bundling is successful, hospitals and physicians in noncompetitive markets may be able to increase volumes and prices for other services to make up for reduced revenues on the bundles. This is more likely if basic payment, except for the bundled episodes, remains volume-based through fee schedules for physicians and either per diems or DRGs for hospitals.
- Only by obtaining discounts off the legacy payment equivalent can payers (and possibly consumers) benefit financially from more efficiently produced episodes. Such discounting may be difficult to negotiate where provider partners have market power, especially with integrated hospital-physician entities.
- A potential barrier to providers' participation is their assumption of risk for large losses. These can result from catastrophic medical events or from patients covered under bundled payment requiring additional services.

Design Choices to Mitigate Weaknesses

Several operational policy options would address both the problem of excessive financial risk and the incentive to stint on care. These include (1) an outlier policy that exempts some amount of payments from the bundle, (2) risk corridors in which payers would share both upside and downside risk with providers, and (3) requirements that payment recipient organizations purchase reinsurance for spending above a certain threshold.

The strongest candidates for bundled payment are episodes for which current costs vary substantially and for which well-established practice guidelines (that can form the basis for normative standards) determine reasonable costs. Payers, however, should want to ensure each procedure is appropriate. Appropriateness may be better ensured by a centers of excellence approach than by broad application of the bundled episode payment approach to participating providers. Other approaches might include reconsideration of second-opinion programs and external precertification determination by the payer. Yet, there is often a large “grey zone” of appropriateness for common procedures.

Giving an entity funds that otherwise would flow directly to a different provider via a prospective payment raises concerns. One approach is to continue paying individual providers in a bundle separately and retrospectively, using standard payment methods that reconcile payments at the end of the episode (and perhaps withholding a portion of payments to cover overspending, if that occurs). In this case, the collaborative group would only have to agree on how to distribute surpluses or pay back

deficits, while a flow of core funding for each entity would be assured. However, the approach of maintaining separate payment streams may to some extent undermine the goal of true collaboration across health professionals and other providers.

Condition-Specific Bundled Episodes

Strengths

- Condition-specific bundling could involve a larger percentage of health care spending than procedure-specific bundled episodes.
- This approach directly counters the possible bias in procedure-based bundled episodes toward unnecessary procedures, by focusing on all components of care rather than each procedure.
- Condition-based episodes could provide a significant role for specialists, who may be functioning as principal physicians for patients with chronic health conditions, without encouraging them to perform procedures or refer their patients to other providers. .
- Evidence-based, clinical guidelines are typically condition based and may be more practical to use with condition-specific episodes than with capitation for all services under a provider's control.

Weaknesses

- Many conditions—even common ones—are not well defined, offering providers an incentive to “find” conditions in order to receive a prolonged payment for a condition-specific episode. Current variations in ICD diagnosis coding, even for common conditions such as congestive heart failure, suggests a lack of standardization with the potential for gaming (although definitions of conditions for episodes are improving).
- Although condition-specific bundling is not likely as complex as the risk adjustment needed for global capitation, there would still be a need for case-mix adjustment for chronic conditions.

- Many patients, especially in Medicare, have multiple chronic conditions. Paying on a condition basis, perhaps to different groups of providers, would not be holistic, possibly counteracting the goal of better care coordination.
- Alternatively, holding a particular provider who has accepted payment for a particular condition-based episode accountable for total health spending as in Medicare’s BPCI initiative can generate conflicts among different physicians caring for different conditions.
- In managing a chronic condition, the cost of a procedure typically dwarfs the cost of medical management absent the procedure. A single condition-specific payment, then, would perhaps create a powerful incentive for its primary recipient to not refer the patient for necessary procedures.

Design Choices to Mitigate Weaknesses

To address the potential for patients with coexisting conditions to receive “nonholistic,” separate episodes of care, payers could combine conditions that often “travel together” (e.g., hypertension, congestive heart failure, diabetes, chronic renal failure) into a single payment that providers would be responsible for managing collaboratively. However, this approach is akin to capitation and undermines the simplicity condition-specific episode payment is meant to achieve.

To address the concern that clinicians will make questionable diagnoses to trigger an episode payment, strict criteria can be required. These criteria can include demonstrating positive test results or providing patients with multiple encounters and documenting diagnoses via claims forms (thus demonstrating a minimum level of persistence or verification of the diagnosis by one or more other clinicians). Although the testing requirement will likely help confirm diagnoses, it might also generate unnecessary, sometimes invasive, tests. This perverse result would see tests ordered to qualify the patient for payment instead of to meet clinical need. Relying on persistence of the same diagnosis on multiple claims is less intrusive for patients, but may be gamed.

The challenges described here are not unique to condition-specific episodes; they also are present in diagnosis coding as the basis for risk-adjusting capitation payments. For private insurance bundles, the common approach to avoiding false-positive diagnoses is to consider newly diagnosed conditions as being in an observation period until the next benefit year begins. At that point, the diagnosis would trigger a condition episode, assuming confirming claims document the condition.

Compatibility with Other Payment Methods and Benefit Designs

Procedure-specific bundled episodes are compatible with reference pricing approaches, as both rely on a fixed, predetermined price for complete procedures and applicable postprocedure days, rather than rely on fee-for-service physician billing. Reference pricing might also improve the likelihood that procedure-based episodes are appropriate, if the payer's approach requires that providers able to perform an intervention at its reference price also meet basic standards, including commitment to evidence-based guidelines and other basic quality standards. Variable cost-sharing, such as with V-BID, varies by individual services and so would seem inapplicable to procedure-specific episode payments; the procedures for which this payment method is being tested do not generally fall within the V-BID ambit.

Narrow and tiered provider networks are typically determined by hospitals and employed or affiliated professionals. Yet, bundled episodes are determined at the hospital and specialty level. The provider collaboration best able to manage a bundled episode might not correspond to the providers selected in a narrow or favored tier. At the same time, for selected or favored providers within the network, a bundled episode approach might be complementary.

Payment for a condition-specific inpatient treatment episode is quite compatible with DRGs, but not with per diem payment. Indeed, some jurisdictions have expanded the duration of a DRG case to extend beyond hospital discharge, so that the hospital takes responsibility for improving discharge planning and transitioning patients back to other facility-based or community-based providers. The approach may also complement various approaches to capitation or population-based payment: the primary care physician at risk or, especially, the ACO-like group would have reason to better assure a procedure is appropriate.

The Focus of Performance Measurement

Procedure-specific and condition-specific bundled episodes present different challenges for measurement. Procedure-specific bundles remain volume based, with incentives to generate perhaps unneeded services. As with other approaches to addressing concerns about provision of unnecessary bundles, it would be useful to measure rates of inappropriate services. At this time, such measures are in a formative state.

Condition-specific episodes, on the other hand, have inherent, strong incentives for providers to avoid costly procedures, which could lead to stinting on care that would benefit patients. We may count on professionalism to protect against denial of referral in clear-cut situations when a procedural intervention is needed. Yet, there is usually a “grey zone” of discretion in which financial incentives may have a large influence. Here, measures of denial of appropriate services would be useful to monitor performance. However, without a population base as the denominator for calculation of procedure rates, it is hard to determine whether needed procedures are being denied

For both versions of bundled episodes, a fixed payment raises concerns about short cuts in quality. Payers could develop specific procedure and condition measures for common bundled episodes, such as joint replacements and deliveries. These measures would not be broad-based core metrics but, rather, would be relevant to the specific procedures and conditions paid under the episode approach. Metrics might include patient-reported outcomes that are currently being developed, which would inform consumers’ and payers’ choice of providers that are to be offered bundled episodes. Indeed, narrowly focusing the episode to a procedure or condition makes available quality metrics more relevant for facilitating individual patient choice. Further, given the incentives inherent in bundled episodes, the Centers for Medicare & Medicaid Services are including measures of shared decision-making, patient-reported outcomes, and clinical appropriateness.

Potential Impact on Provider Prices and Price Increases

As with some payment approaches, if the bundled episode fee is based on historic costs, the payer will build in current pricing disparities that result from variable negotiation leverage. Basing payment on normative pricing, such as the community average, would directly penalize higher-priced providers and make their voluntary participation less likely. A specific concern is that putting hospitals, clinicians, and, perhaps, postacute care facilities together into a recognized “focused factory” could produce, in effect, a “bargaining unit.” This could raise prices higher than they would be if the parties were negotiating separately, without their ongoing joint participation in providing services.

Global Capitation to an Organization

Global capitation is a payment model specifically for integrated health care delivery. In this model, capitation payment for services delivered by different providers or at different levels of care is combined into a single prospective payment to an integrated care organization or a large physician group. The provider is then responsible for delivering all needed care for a defined population and for distributing payments to its constituent providers out of the capitation pool. The core concept is that total payment does not vary based on the actual services provided to individuals in the population served.

The services included in global capitation typically include at least physician, hospital, and postacute care facility-based services, and may include additional services, such as prescription drugs. For health plan enrollees, capitated payments are expressed as per member per month payments² and are usually adjusted at least for age and sex. Payers also have pushed to risk-adjust payments by enrollees' health status, usually calculated based on the diagnoses clinicians and hospitals specify on encounter forms. A similar approach that effectively provides global capitation is for insurers to pay a specified "percentage of premium"—an actuarially based amount that represents the percentage of the market-specific insurance premium (which also is made per member per month) that supports the clinical services for which the capitated provider organization is responsible.

By accepting a defined fixed payment to provide contracted services, providers assume the financial risk for their patients, usually including both insurance risk and technical risk. The former is financial risk caused by the likelihood of a random event occurring that is not under the control of providers. The latter relates directly to how care is produced, and therefore is under the providers' control. Because of insurance risk, organizations accepting global capitation typically buy reinsurance to help protect against losses from unanticipated high-cost cases.

For global capitation contracting to work well, it should apply over a sufficient number of members to spread insurance risk, thereby reducing volatility and the impact of bad financial experience resulting from random occurrences. Capitation rates are usually determined based on actuarial analysis, taking into account prior spending by the population to be served, the benefits covered in the capitation payment, and assumptions about medical management and other cost controls available to the integrated group.

Although the global capitation model had been in limited existence for many years, it spread substantially in the 1990s with sharp growth in the number of HMOs. The failure of physician practice

management companies (which were hired to manage risk) and some integrated delivery systems (that were taking risk), as well as the negative reactions from providers and patients about perceived restrictions on choice and access to care, dampened this trend. Over the past decade, there has been a shift from global capitation toward capitation for professional services only. Also, self-funded employers are generally precluded from shifting risk to provider organizations, as this would bring scrutiny from state insurance regulators.

Key Objectives

Because payment does not vary based on the actual services provided, global capitation encourages provider organizations to deploy the mix of activities and staffing they deem best able to serve the populations assigned to their care. If payers similarly provide global capitation payments, the recipient organization in effect must manage within a defined, known budget, with inherent incentives to increase efficiency and manage costs. In essence, global capitation transfers responsibility for managing medical care costs and quality from the third-party payer directly to the provider receiving the payment.

Strengths

- Global capitation is the most robust method for health care services across the spectrum to internalize incentives for improving efficiency and effectiveness. It is the prototypical population-based payment method and offers the recipient organization the greatest opportunity to change its business model and culture.
- The model promotes integration of services across what are often “siloes” independent clinicians and facilities.
- Global capitation is a form of prepayment that provides ongoing cash flow, which permits recipients to deploy capital for delivery system enhancements and to establish reserves that comply with regulatory requirements on risk bearing.
- The model offers providers broad flexibility to determine the best mix of services and to identify the particular professionals best able to meet the target population’s needs.

- Global capitation puts clinicians, rather than payers, directly in charge of patient care decisions, but offers incentives for more prudent expenditures.
- A global capitation payment is a relatively simple transaction, involving less administrative infrastructure for both payers and providers than fee-for-service does. Yet, the method becomes complex when payers require risk adjustment of payments and monitoring of quality.
- Approaches such as two-sided shared savings attempt to incent provider organizations to reduce total cost of care. In contrast, global capitation, which makes up-front base payments to the provider organization, permits the greatest flexibility in deployment of resources and payment of health professionals, other providers, and suppliers.

Weaknesses

- The method places both insurance and technical risk on providers, which can result in financial losses outside providers' control and lead to indiscriminate service reductions and stinting on care.
- Many organizations lack the capital and infrastructure, including administrative data systems, to manage substantial financial risk. Global capitation may not be a payment reform with broad applicability, especially given the cautionary experience of its wider use in the 1990s.
- Similarly, organizations may lack the managerial skills and commitment to manage patient care in ways consistent with global capitation.
- Risk-bearing provider organizations have greater potential to become insolvent or to compromise quality. This potential calls for strong regulatory oversight, which some jurisdictions may be reluctant to take on. Further, oversight of globally capitated entities can be challenging because data and reporting are lacking when there is no fee-for-service billing.
- In market-based health systems without regulated prices, consolidated and integrated groups capable of accepting global capitation can develop market power and use it in their price negotiations with payers, thereby raising prices and health care spending even if they are able to reduce service use.
- Common risk-adjustment approaches using recorded diagnoses are subject to “code creep”—when more extensive coding of diagnoses and procedures increases patients' apparent

costliness. Adjusting for and detecting code creep also makes the approach more administratively complex than the simple per person per month payment suggests.

- Established approaches to global capitation, in an HMO context, generally require individuals to “enroll” with the capitated organization, therefore limiting their freedom to select clinicians of choice at the point of service.
- Many clinical activities—for example, evidence-based recommendations for prevention services—don’t “pay off” for many years. The typical one-year enrollment does not reward globally capitated providers for investing in preventive activities when would be realized in the future, when the patient may no longer be enrolled with that organization.
- Consumers may be concerned that the incentives in global capitation (as well as in other forms of capitation) will compromise the doctor-patient relationship, at the core of which is the physician’s duty to act in the patient’s best interests, regardless of impact on the bottom line.

Design Choices to Mitigate Weaknesses

Various approaches to addressing insurance risk are possible, including reinsurance for high-cost individuals and for excessive aggregate spending—typically, these are available at a cost to the organization. Global capitation can be accompanied by risk corridors, in which a payer and an at-risk provider organization agree to share in both upside and downside risk. This reduces financial exposure (even for technical risk) and can be especially beneficial for organizations new to the approach or with limited reserves, while still providing financial incentives to prudently preserve resources.

Global capitation works best with implementation of sophisticated risk adjustment for patients’ health status. Risk adjustment attempts to mitigate providers’ incentives to avoid sick or complex patients and promotes integrated management models to care for such patients. However, the data needed for most risk adjustment models must be collected from encounter data with diagnoses, thereby reducing administrative simplicity, one theoretical advantage inherent in paying a fixed, prospectively set amount. Risk adjustment based on diagnostic coding in claims or encounter forms is subject to code creep that raises risk scores and, thus, payments. That has been the experience in Medicare Advantage and could be even a greater issue when provider organizations have a direct interest in “finding” all diagnoses to improve risk scores—and payment amounts.

How capitation rates are updated determines their impact on spending over time. Setting and updating capitation payments based on providers' incurred costs can reduce incentives for cost control. So, in distinct contrast to typical calculations in shared savings payments for ACOs, global capitation payments are often based on actuarial analysis of likely costs for the average patient across all providers, then risk-adjusted for the population served by the organization.

An appeal of percentage-of-premium payment is that it grounds payment to the average marketplace premiums (or the equivalent for public payers). However, percentage-of-premium payments may also reflect strategic considerations the payer uses in determining its premium rates (e.g., to gain market share by having aggressively priced premiums). When that occurs, the percentage-of-premium may be lower than the capitation rate that would otherwise have been determined through actuarial analysis.

Compatibility with Other Payment Methods and Benefits Designs

Globally capitated providers adopt their own provider payment methods rather than rely on the third-party payer to do so. Global capitation is flexible in permitting a range of options, assuming the globally capitated group can administer or contract out for the approach, which may involve sophisticated claims processing. The payment method may vary based on specific participants in the risk-bearing entity and the particular delivery culture. For example, a globally capitated, integrated health system based around a hospital might use DRGs to make hospital payments compatible with the dominant Medicare system, while providing incentives to reduce length of stay. Yet, a globally capitated medical group might shop aggressively for hospital care and use per diems, because the group, through its own physicians, could assure that length of stay is not extended unnecessarily.

Globally capitated provider organizations can pick a range of payment options to compensate health professionals, including salary, productivity analysis of fee-schedule-based relative value units produced, various forms of sub-capitation, including primary care, specialty, and contact capitation.

Because most states regulate risk-bearing provider entities out of concern providers could become insolvent or stint on care, global capitation requires an HMO product design, rather than a PPO. PPOs' market shares have been growing with a concomitant decrease in HMO share, because PPO products are consistent with employer self-funding and because PPOs assure patients more choice in providers.

Global capitation relies fundamentally on supply-side rather than demand-side attention to costs and cost control. Under global capitation, fee-for-service transactions between payer and provider do not exist, so some forms of patient cost-sharing do not work well. Co-payments for office visits and some other services are commonly used, but high deductibles, coinsurance, and value-based benefit designs would seem incompatible, because the patient's share of the cost would be tied to encounter codes, which may not be relevant under global capitation. At the same time, high deductibles and other forms of cost-sharing do discourage spending so have theoretical appeal to some.

The Focus of Performance Measurement

Payers can require quality-related performance measures to expose organizations that would stint on care, as well as to counter the disincentive to invest in prevention activities that might prevent illness. Providing the public with measures of primary and secondary prevention services, such as those available through HEDIS, may help allay consumers' concerns about health professionals compromising care for financial reasons.

However, provider organizations are responsible for the entire continuum of care, not simply prevention and routine services (for which there are a raft of available performance measures). Globally capitated providers want to keep most services within their own delivery system, even when a referral might produce a better clinical outcome, because of cost differentials and because continuity and coordination of care is easier within a narrower group of health professionals. Yet, currently, few measures of referral appropriateness for services for which differences in clinical expertise may be relevant to the resultant care outcomes. Where measures fall short, payers will have to consider other approaches to holding provider organizations accountable for quality.

Potential Impact on Provider Prices and Price Increases

Global capitation rates are typically calculated based on actuarial analysis of average individuals—the normative or community average—not the historic costs of care capitated providers have experienced. In short, paying the community average, adjusted for individual consumer risk, provides a basis for addressing pricing (in that the capitation amount would exclude the effect of prices on overall spending).

However, an insurer's ability to globally capitate a provider based on the community average varies with each party's negotiating leverage. Provider organizations with pricing power can achieve higher capitated payment levels that deviate up from the community average. When the payer must concede higher capitated rates to some provider organizations, it might have to reduce payments to other providers to meet actuarial targets for competitive premiums, thereby further increasing pricing variations across providers.

A globally capitated health care system or physician group offers different opportunities for addressing pricing differences in hospitals and, to a lesser extent, the physician specialty market. Health care systems typically grant preference to hospitals within their own systems, whatever their prices. In contrast, a physician group without commitment to a particular hospital or hospital system can aggressively shop on price, assuming quality is comparable. By actually moving or threatening to move patients from one hospital to another, a capitated physician group can achieve price concessions from hospitals seeking to preserve or increase their market share of bed days and outpatient services. Of course, again, physician groups can only negotiate lower prices in reasonably competitive markets, such that the threat of moving patients is a credible one.

Shared Savings

The shared savings method has been introduced partly to give providers an “on-ramp” or “training wheels” for moving away from fee-for-service payment. This form of incremental payment, which some consider a form of pay-for-performance focused primarily on spending reductions, is commonly associated with accountable care organizations (ACOs). Yet, the generic approach may be applied to any type of provider organization. In current Medicare and commercial insurance demonstrations, shared savings reward or possibly penalize (shared risk) ACOs for their spending performance (in relation to spending targets for a population of patients attributed to the ACO). In this method, essentially, base payments continue using established methods (typically, fee schedules for professionals, DRGs or per diems for hospitals), while ACO entities can receive additional payment if their spending for beneficiaries is lower than a target. When the ACO achieves savings, it can then distribute its share (possibly adjusted by performance on a set of quality measures) to its constituent members.

Shared savings programs can be characterized as “upside-only” (“one-sided”) risk or “upside-and-downside” (“two-sided”) risk which here we have labeled “shared risk”. Consistent with the training-wheels notion, most ACO arrangements start with upside-only risk and migrate to shared risk over time, once the ACOs have had some experience with shared savings. Some shared savings programs, including Medicare’s, require a “minimum savings threshold,” which separates spending reductions due to successful ACO efforts from spending reductions due to random variations; the more people assigned to the ACO, the lower the threshold for receiving shared savings. Shared risk models generally give ACO providers a larger percentage of savings bonuses in exchange for the financial risk they are willing to assume.

An essential part of the shared savings approach as applied to ACOs is reliance on quality performance measures. An ACO’s performance on quality measures determines how much of a bonus it is eligible to receive for reducing spending against its target. In contrast to many P4P programs that target individual providers, ACO shared savings programs use measures that reflect the quality the ACO provides its patient population.

A fundamental difference between global capitation and shared savings and shared risk approaches is that the former is a base spending method and includes most revenues a payer provides the recipient organization, whereas the latter provides incremental reward or penalty placed on top of other base payment methods—usually the legacy payment methods of fee schedules for physicians, DRGs or per

diems for hospitals, and so on. Therefore, the incremental shared savings approach does not offer providers the same opportunity to change their business models and cultures or furnish the same predictable cash flow to support change. (The Medicare Advance Payment model attempts to address this deficiency for small organizations participating in Medicare shared savings programs for ACOs.) Also, global capitation typically pays based on a marketwide average for all providers, whereas shared savings approaches typically rely on each ACO's historic spending as the base, at least initially.

The update factor needed to trend the ACO's historical spending for inflation can be determined in various ways. Payers can use a matched patient control group to determine the rate of increase applied to the target group's historic spending. More commonly, an external factor, such as an estimated trend in overall health care spending, is used to establish the applicable spending target. In most formulations, both historic target spending levels are risk-adjusted for age, sex, health status, and possibly other factors such as socioeconomic ones.

Key Objectives

Shared savings is generally viewed as a practical, transitional payment model to grant providers experience with incentives to spend more prudently on health care services. Many payers, including Medicare, propose an evolution from fee-for-service to upside-only to shared risk, and perhaps ultimately, to more robust forms of population-based payment such as global capitation. Taking their lead from Medicare, most payers' shared savings programs also place primacy on using the approach to assure quality improvement, as shared savings don't flow to the recipient provider organization unless it meets quality performance thresholds—an element that has not been central in capitation payment arrangements in the past.

Strengths

- One-sided shared savings does not require provider organizations to take on major financial risk, something many such organizations—especially small organizations—are not in a position to do. It establishes gentler, perhaps more realistic, positive incentives that can provide a reasonable entry for organizations that are new to risk-bearing and lack capital to manage global capitation.

- The approach to using historic spending as the base for spending targets may be more practical for many organizations whose spending exceeds the average. Shared savings presumably targets “low-hanging fruit”—savings that may be easier to achieve from a high base.
- One-sided shared savings can generally be adopted under PPO arrangements because of the absence of risk taking.
- The “on-ramp” concept makes sense. With the experience gained, ACOs that improve under upside-only shared savings should naturally evolve into working toward the stronger incentives of shared risk.
- Shared risk fundamentally changes volume-based incentives much as capitation does, but may be more practical to implement. Routine cash flow continues under established payment methods, so the ACO need not take on the challenging role of paying its constituent provider members (although it still needs to decide how to share any savings earned with its members.)
- The central role of population-based quality metrics provides some assurance that spending reductions will not compromise quality.
- Models that attribute individuals to ACOs typically do not lock patients in to a particular primary care physician responsible for approving referrals. Most programs either do not limit choice at all (e.g., Medicare) or are placed on a PPO product platform that has gentle benefit design incentives to influence provider choice.
- Even short of a fundamental reorientation to providing care, provider organizations under shared savings can adopt relatively straight-forward approaches (e.g., improving transitions of care from hospital to community, coordinating care for patients seeing many different providers, adopting evidence-based guidelines).

Weaknesses

- The dominant, base payment methods used in shared savings models remain volume based. Expecting the small incremental incentives placed on a separate or intermediary ACO organization to reduce spending (to counter the volume-inducing incentives of the underlying payment system) may be unrealistic.

- Using unadjusted historic spending to determine spending targets is unfair to organizations that have had above-average performance on spending, as they have less room to achieve additional spending reductions.
- Similarly, under shared savings, there is a law of diminishing returns after the “easy savings” have been achieved. The maintenance of volume-based payment models—especially under one-sided shared savings—could actually make it more difficult to achieve the major changes in providers’ business models and cultures that are the goals of stronger payment approaches such as global capitation.
- Operationally, determining whether and to what extent savings have actually been attained can be challenging. The Medicare ACO program has been subject to criticism for its retrospective attribution of patients for which the ACO is responsible and for its non-intuitive calculation of shared savings bonuses.
- As with global capitation, ACOs may need to consolidate and integrate to have sufficient size and scale to meet requirements under shared savings and shared risk methods. This may empower organizations to use their newfound organizational clout to negotiate higher prices for the base payments that determine most of constituent members’ revenue—as well as strengthen their negotiating position with other payers that don’t participate in the shared savings arrangement.

Design Choices to Mitigate Weaknesses

There are many operational challenges in implementing shared savings and shared risk for integrated care organizations accountable for cost and quality for a population:

- how patients will be attributed to an ACO and whether freedom to choose a provider will be limited, if granted at all
- whether patients should share in savings
- setting any minimum savings threshold ACOs will have to meet before they can share in savings
- the selection of applicable quality measures
- if and when to require two-sided risk

- whether to apply risk corridors to provide financial protection when ACOs do accept risk
- how to evolve the payment target from on an ACO's historic costs trended forward to some consideration of market-based, average costs
- providing the right balance between the percentage of savings providers keep (to provide incentives to participate) and the savings payers keep (to make shared savings worth the effort)

Often ignored in shared savings discussions are the underlying payment methods that provide ACO constituent providers with basic payments. Reforms targeted to these legacy payment models can complement the shared savings overlay. But, typically, these are given little attention because the marginal rewards (or penalties) associated with shared savings are the focus. For example, there is a broad perception that physician fee schedules overpay for tests and procedures, at the expense of communication with patients and other professionals, thereby making the ability to reduce spending more difficult. These distortions create conflict between the ACO and some constituent physicians. Correcting flaws in the design of these base payment methods could facilitate the savings the ACO is trying to achieve.

Compatibility with Other Payment Methods and Benefit Designs

As an incremental payment approach, shared savings provides bonuses and shared risk bonuses and penalties compatible with a range of base payment methods. The savings calculation is based on spending regardless of the form of payment, except for global capitation (which provides a fixed payment, making shared savings calculations unnecessary).

Because shared savings does not alter the basic payment methods and payment rates, it is not incompatible with various cost-sharing options (as global capitation is). At the same time, shared savings, like global capitation, is also a supply-side strategy: it offers provider organizations incentives to actively assess medical need and reduce avoidable services, rather than relying on consumers to become smarter shoppers for services.

The Focus of Performance Measurement

The most prominent use of shared savings is with ACOs. A significant goal of the Medicare approach, which has been largely adopted by private insurers and some Medicaid programs, is to introduce performance on quality measures as an essential element of the model; ACOs do not get to keep savings or have a smaller share of savings based on their quality performance. The rationale is straightforward: incentives in shared savings can produce stinting on care (even more so in shared risk and capitation), withholding important (usually nonurgent) services that largely prevent adverse health outcomes months or years later. Measures related to organizational performance for caring for population health are also desirable in the context of ACOs with shared savings or shared risk incentives. For both of these objectives (assuring provision of primary and secondary prevention services and measuring population health for specified individuals), measures are available and likely statistically valid if measuring at across ACO organizations rather than individual health professionals.

However, consumers are also concerned that payment methods that encourage more prudent health care spending will lead to a reduction in desirable referrals to physician specialists and other health professionals with specialized expertise, especially if those providers are not part of the ACO network. At this time, there are no performance measures related to appropriateness of referrals. Network adequacy requirements may help ensure that specialized expertise is available within a network such as those in an ACO. However, availability does not assure appropriate referrals are made or that patients are informed about their referral options.

Potential Impact on Provider Prices and Price Increases

The horizontal and vertical integration ACOs represent can be used to increase pricing power in negotiations with private payers over physician and hospital prices. This counters the cost savings objectives of the shared savings approach. Medicare-approved ACOs will still be subject to antitrust scrutiny for exceeding provider concentration benchmarks, and they can be prevented from entering commercial insurance markets even if they function as Medicare ACOs. However, arguably, antitrust enforcers will find it more difficult to prevent a successful Medicare ACO from participating as an ACO with private insurers. A successful Medicare ACO, even less likely to be subject to antitrust scrutiny, might be able to negotiate higher prices with private payers without attempting to become an ACO with those payers.

For commercial insurance, in contrast to public payers, shared savings is based on spending targets, which themselves are based on historic costs trended forward. This method actually accepts and “bakes in” pricing differentials that the constituent providers have been able to negotiate. The shared savings approach might moderate price increases, but not necessarily, because providers’ prices might well be higher than their share of savings from beating spending targets. Further, as under global capitation, a physician group without a commitment to a particular hospital or hospital system can aggressively shop on price, assuming quality is equal. By actually moving or threatening to move patients from one hospital to another, a capitated physician group can achieve price concessions from hospitals seeking to preserve or increase their market share of bed days and outpatient services. Of course, again, physician group shopping can occur only in reasonably competitive markets such that the threat of moving patients is credible.

Pay-for-Performance

Background

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

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

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

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

Key Objectives

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

Strengths

- Most payment methods' primary impact is on volume of services produced. P4P introduces into payment policy emphasis on the quality of care produced, a core element of care that has been missing in base payment methods.
- P4P permits payers and purchasers to emphasize which aspects of performance deserve priority—for example, aspects of care that might be compromised under the incentives of particular base payment models.
- P4P is complementary to public reporting of performance; together, public reporting and pay-for-performance impart transparency to better hold providers accountable for the large payments they receive, to help consumers and patients make informed choices of provider, and to support quality improvement efforts.
- P4P can be complement base payment methods without changing their basic structures. It offers payers the most practical approach to improving value with providers who are unwilling or unable to accept new forms of base payment.

- As an incremental payment method, P4P can be implemented with varying degrees of intensity, consistent with the context of application, the strength of the measures available for the clinical conditions to which it is being applied, or other relevant factors.
- Although there are clear gaps in what is accurately measureable, a commitment to P4P could create momentum to expand measure sets and approaches to achieving greater measurement accuracy.

Weaknesses

- P4P introduces significant administrative complexity associated with acquiring data and verifying it for accuracy.
- Behavioral economics suggests that, in professions that require high cognitive skill and high intrinsic motivation, associating better performance with financial incentives could be counterproductive because it might compromise commitment to quality. P4P incentives for organizations such as hospitals may or may not impart this “crowd out” of intrinsic motivation.
- For many health professionals and providers, there are major gaps in what aspects of care are measurable using current data sources. Therefore, erroneous judgments about a provider’s overall quality and value may occur.
- Most P4P programs tend to concentrate on clinical process measures rather than outcomes, which are what consumers or payers are most interested in achieving. Moreover, research indicates that process measures are not strongly related to significant healthcare outcomes.
- Apparent improvement in performance may simply reflect more extensive documentation of what was already being done for patients, thus reflecting improvement in reporting rather than actual performance.
- It is administratively easy and practical for payers to base measurement on objective administrative data, usually from claims such as laboratory test results, but that limits the choice of measures. Clinical data from records, self-reporting by providers, or patient-reported outcomes is more costly to obtain and not necessarily reliable (although this may change with improvements in electronic medical records).

- The small incremental reward and penalty payments common in P4P programs may not be sufficient to counter the much stronger incentives in the base payment methods that produce a larger share of provider payment.
- In a multipayer health care system, different P4P regimes may cause providers dissonance in responding to different measures, different measurement requirements, and different approaches to rewards and penalties.
- Attainment approaches in public reporting and P4P that compare providers' performance may not be fair to providers with more challenging patient populations, perhaps because of socioeconomic factors or unmeasured case-mix differences. Yet, improvement approaches compromise the goals of providing consumers with information for choice and for accountability—as well as making it easier for initially poor performers to receive rewards for their improvement.
- Clinicians might respond to P4P incentives by altering their professional activities so they perform better on the P4P measures, as they sometimes do in public reporting programs—for example, by not caring for high-acuity patients with greater likelihood of experiencing a poor outcome.

Design Choices to Mitigate Weaknesses

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

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

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

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

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

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

Compatibility with Other Payment Methods and Benefit Designs

As an incremental payment based on measured performance, P4P is compatible with all base payment methods. P4P is also compatible with shared savings incremental rewards and penalties, with the former typically focused more on quality and the latter on costs. However, measurement of costs and cost increases could be included as part of P4P without a separate shared savings program. A prime example is the Medicare MIPS program, under which one of the four components of a physician’s value

score will be their resource use performance as measured by spending for their patients. More typically, P4P would include utilization data—for example, readmission rates—whereas shared savings by definition would focus on expenditures.

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

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

Focus of Performance Measurement

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

Potential Impact on Provider Prices and Price Increases

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

Notes

1. The Healthcare Effectiveness Data and Information Set (HEDIS) is one of the most widely used health care performance measures in the United States and is managed under the auspices of the National Committee for Quality Assurance.
www.ncqa.org/HEDISQualityMeasurement/HEDISMeasures.aspx#sthash.B74KCSM4.dpuf.
2. Where consumers are not enrolled as members in a health plan, capitated payments might be referred to as per person per month.

References

- Berenson, Robert A., Divvy K. Upadhyay, Suzanne F. Delbanco, and Roslyn Murray. 2016. *A Typology of Payment Methods: Payment Methods and Benefit Designs: How They Work and How They Work Together to Improve Health Care*. Washington, DC: Urban Institute.
- Busse, Reinhard, and Wilm Quentin. 2011 "Moving Towards Transparency, Efficiency, and Quality in Hospitals: Conclusions and Recommendations." In *Diagnosis-Related Groups in Europe*, edited by Reinhard Busse, Alexander Geissler, Wilm Quentin, and Miriam Wiley. Berkshire, UK: McGraw Hill Open University Press. eurodrg.projects.tu-berlin.de/publications/DRGbook/Ch10_Busse.pdf.

About the Authors

Robert A. Berenson joined Urban as an Institute fellow in 2003. In this position he conducts research and provides policy analysis primarily on health care delivery issues, particularly related to Medicare payment policy, pricing power in commercial insurance markets, and new forms of health delivery based on reinvigorated primary care practices. In 2015, Berenson was appointed to the Physician-Focused Payment Model Technical Advisory Committee set up by Congress through the Medicare and CHIP Reauthorization Act of 2015 to advise the Secretary of the Department of Health and Human Resources. In 2012, Berenson completed a three-year term on the Medicare Payment Advisory Commission, the last two years as vice chair. From 1998 to 2000, he was in charge of Medicare payment policy and private health plan contracting in the Centers for Medicare and Medicaid Services. Previously, he served as an assistant director of the White House Domestic Policy Staff under President Carter. Berenson is a board-certified internist who practiced for 20 years, the last 12 years in a Washington, DC, group practice. He is a graduate of the Mount Sinai School of Medicine and a fellow of the American College of Physicians.

Divvy K. Upadhyay is a research associate in the Health Policy Center at the Urban Institute, where he focuses on qualitative research on health reform measures at the federal and state levels. His policy research interests include reforms related to physician and hospital payment; health care delivery; the Affordable Care Act; Medicare and Medicaid; and issues in medicine related to primary care and diagnosis errors. Before joining Urban, Upadhyay worked with the editorial team of *Health Affairs*. Trained as a physician, Upadhyay holds an MPH with a focus on health policy and organization from the University of Alabama at Birmingham.

Suzanne F. Delbanco is the executive director of Catalyst for Payment Reform (CPR), an independent, nonprofit corporation working on behalf of large health care purchasers to catalyze improvements to how we pay for health services and to promote better and higher value care in the United States. In addition to her duties at CPR, Delbanco serves the board of the Health Care Incentives Improvement Institute. Previously, Delbanco was the founding chief executive officer of The Leapfrog Group. Delbanco holds a PhD in public policy from the Goldman School of Public Policy and an MPH from the University of California, Berkeley School of Public Health.

Roslyn Murray is the project and research assistant for CPR. In this role, Murray provides background support and conducts research for CPR on policy and health care payment reform issues. She has also

held research assistant positions at Stanford School of Medicine, the National Institutes of Health's National Heart, Lung and Blood Institute, and San Francisco General Hospital. A graduate of Stanford University, Roslyn holds a bachelor's degree in human biology, health, and health policy.

STATEMENT OF INDEPENDENCE

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



2100 M Street NW
Washington, DC 20037

www.urban.org