# Rewarding Results Pay-for-Performance: Lessons for Medicare

By Amanda Folsom, Cyanne Demchak, and Sharon B. Arnold

Ever since 2001, when the Institute of Medicine recommended that public and private purchasers of health care build stronger incentives to enhance quality, there has been a proliferation of pay-for-performance (P4P) demonstrations.[1] But despite this growth in new initiatives for quality-based incentive programs for hospitals, physicians, nursing homes and home health providers, there is limited evidence of the effectiveness of early private-sector demonstration projects.[2] As efforts to link reimbursement and performance continue to increase, a careful review of the lessons learned from early initiatives should be an important component of the design, implementation and evaluation of any new P4P initiatives.

On December 15, 2006, the Rewarding Results (RR) National Evaluation Team convened leaders from the demonstration projects to meet with Centers for Medicare and Medicaid Services (CMS) staff and demonstration implementation and evaluation contractors to discuss the extent to which lessons learned from their early demonstrations were transferable to the Medicare environment. The Robert Wood Johnson Foundation and the Agency for Healthcare Research and Quality funded the event. The meeting goals and objectives were to:

- Identify major design, implementation and evaluation challenges and lessons learned in P4P
- Discuss what could be done differently with respect to design, implementation and evaluation strategies based on what has been learned
- Discuss implications and design consideration in implementing P4P systems for the Medicare program and its beneficiaries.

CMS participants included 44 CMS contractors and CMS staff responsible for Medicare P4P initiatives. Twelve experts from the Rewarding Results demonstrations were assembled by the National Evaluation Team to identify and report crosscutting themes from the three years of demonstrations.

## History of Rewarding Results

In 2002, the Robert Wood Johnson Foundation, California HealthCare Foundation and the Commonwealth Fund sponsored seven Rewarding Results demonstration projects. These demonstrations provided early, hands-on experience in the design, implementation and evaluation of P4P initiatives. The seven initiatives were selected from 151 applications submitted by health plans, employers, unions and state agencies, as well as collaborative efforts between employers, providers and plans. These locally designed demonstrations made providers eligible for financial and nonfinancial rewards upon achievement of specific quality goals linked to clinical performance and medical outcomes. The Rewarding Results demonstration projects offered a variety of approaches to employing incentives to achieve higher quality health care. They included the following:

- Blue Cross of California (BCC) piloted financial and nonfinancial incentives for preferred provider organization (PPO) network physicians practicing in family practice, general practice, internal medicine, obstetrics/gynecology, pediatrics, cardiology and endocrinology in the San Francisco area. Physician incentives included Web-based performance scorecards and annual increases in the standard PPO fee schedule. Measures included evidence-based preventive care, chronic care management, formulary compliance, timely electronic claims submission and continuity of access.
- Blue Cross Blue Shield of Michigan (BCBSM) offered opportunities for diagnostic related group (DRG) payment increases to all of its 86 contracted acute care hospitals. Hospital incentives for the care of plan members amounted to an up to 4 percent DRG increase in the following year for scoring at or above threshold standards for selected all-payer, self-reported Joint Commission (formerly JCAHO) measures; medication safety; a community health project; and efficient utilization.
- Bridges to Excellence (BTE) designated per member, per year bonus payments to physicians for meeting the requirements of its Diabetic Care Link, Cardiac Care Link and Physician Office Link programs. The demonstration occurred in Boston, Cincinnati, Louisville and upstate New York. Clinical performance scores for the clinical measures were based on National Committee for Quality Assurance (NCQA) measures. In addition, bonuses were linked to the innovative use of information technology. Bonus payments were designed to share the actuarial expected savings to employers with accredited physicians.

Robert Wood Johnson Foundation

Changes in Healthcare Financing and Organization is a national program of the Robert Wood Johnson Foundation administered by AcademyHealth.

AcademyHealth
Advancing Research, Policy and Practice

- Integrated Healthcare Association (IHA) created a single, multi-health plan set of measures which included clinical quality (50 percent), patient satisfaction (30 percent) and information technology investment (20 percent) measures for physician groups. Financial incentive payments were paid directly from seven health plans to contracted groups in accordance with individually designated and independently operated health plan P4P programs. The demonstration included a widely disseminated public scorecard displaying comparable physician group performance, which served as a major nonfinancial incentive.

- The Center for Health Care Strategies' (CHCS) Local Initiative Rewarding Results (LIRR) used incentives to improve child and adolescent access, service, and Health Plan Employer Data and Information Set (HEDIS®) scores in seven Medicaid health plans in California. The health plans also sought to improve the collection and transmission of encounter data from capitated provider groups. Individual health plans' financial incentives included performance-based risk pool distributions, tiered capitation increases and bonus payments for guideline adherence. Incentives were designed to provide specific dollar amounts per child that met the measure. Several plans offered nonfinancial, age-appropriate incentives for adolescents upon completing a well visit.

- Massachusetts Health Quality Partners (MHQP) developed a statewide data infrastructure to enable six major health plans to aggregate HEDIS clinical effectiveness measures across health plans at the physician, group and integrated delivery system levels. Aggregated profiles were used to create a new program of performance feedback and public release of comparable, trended quality performance information. These profiles were designed to overcome the fragmented, multiple, incomplete and sometimes conflicting performance profiles provided by individual health plans. The aggregated data could also be used to evaluate the six health plans' uniquely designed and operated financial and nonfinancial incentives.

- Rochester Independent Practice Association (RIPA) and Excellus Health Plan collaborated to develop a physician reimbursement program based on locally developed care quality guidelines for chronic conditions, cost efficiency measures and patient satisfaction. Patient registries and patient-specific physician notification systems were developed. Each physician's adherence to guidelines was ranked to determine the annual distribution ranging from 50 percent to 150 percent of a physician's 10 percent withhold of HMO fees.

## Lessons Learned from Rewarding Results and Crosscutting Themes

The design and implementation challenges expressed in these seven demonstrations provide an opportunity to identify lessons learned and implications for Medicare P4P. Six crosscutting programmatic themes emerged from discussions during the December 2006 learning session with CMS: stakeholder engagement, infrastructure, incentive methodology, continuous quality improvement, transparency, and measuring and evaluating impact. The Rewarding Results lessons and implications for Medicare P4P in each of these thematic areas are discussed below.

**"Pay for performance is not simply a mechanism to reward those who perform well; rather, its purpose is to encourage redesign and transformation of the health care system to ensure high-quality care for all. In such a system, all participants, providers, purchasers, and beneficiaries can potentially benefit."**

—Institute of Medicine, Rewarding Provider Performance: Aligning Incentives in Medicare, National Academies Press, August 2006.

### I. Stakeholder Engagement

Purchasers, providers, consumers and health plans are all important stakeholders in P4P initiatives. Providers are central, as the care they provide is measured for quality by both payers and patients. Their role in the health care system makes provider engagement in any process reform invaluable to the longevity of changes made.

Because of their financial interest, purchasers have historically had a significant interest in assessing the value of health care services. Although their interest was initially focused on appropriate pricing, purchasers have had a significant role in assessing whether the services they were purchasing were of high quality and were consequently of sufficient value to justify the expenditure.

In recent years, there has also been a push to motivate consumers to become more actively involved in their health care. Getting consumers involved is vital to the success of P4P initiatives because a significant portion of health care takes place outside the provider's office. In addition, consumers can play a role in P4P by actively seeking high-quality providers and making informed health care decisions.

Research has not found the best way to present information to consumers to get them fully engaged in their health care, yet evidence has shown that information must be presented in an easy-to-understand format.[3] In their work on consumer activation, Judith Hibbard, Dr.P.H. and colleagues note that although the need for informed and activated consumers is widely accepted, a central theory to guide activation is "underdeveloped."[4] Essential to the ability to engage consumers in their health care and P4P programs is an understanding of what will activate them, including how best to leverage the physician-patient relationship. Much more research is needed in this area.

#### Lessons From Rewarding Results – Stakeholder Engagement

A consistent lesson from the Rewarding Results demonstrations is that stakeholder engagement must be factored into every step of the design and implementation of a P4P initiative in order to maximize impact. In addition, methods for measuring stakeholder engagement (e.g., surveys, focus groups, expert panels) should be identified as part of the first step in P4P efforts. Opportunities to accelerate quality performance are forfeited when stakeholders are not engaged in the initiative. The absence of effective strategies to engage key stakeholders can result in

### Rewarding Results Experiences: Stakeholder Engagement

Involving recognized local champions and other stakeholders in the evaluation and selection of measures enhanced provider acceptance and engagement in RR sites. For example, both IHA and MHQP developed large and representative stakeholder councils to provide advice about increasing engagement. Public hearings were held and provider input was received before quality measures were selected. RIPA developed several locally approved measures based on community standards.

program failure for reasons such as lack of participation, misunderstanding, distrust over actual costs and who benefits, and lost opportunities for stakeholder-initiated redesign of clinical systems.

**Provider Engagement**
Based on extensive literature review and research, the Rewarding Results National Evaluation Team developed a conceptual framework identifying dimensions of provider attitudes that influence the effectiveness of financial incentives for quality. The specific dimensions are provider awareness of the quality incentive, size and structural salience, clinical relevance of quality targets, accountability for targets as measured by provider control and required cooperation, and unintended consequences.[5]

A recent article by Mark Meterko, Ph.D. and colleagues noted that providers must be motivated to initiate changes to their practice style, and this motivation is likely affected by their understanding of the incentives and measures being used. Providers must also feel that they have control over outcomes and quality targets, such as the ability to cooperate with other physicians or specialists involved with a patient.[6] A study examining performance-based quality improvements for diabetes care noted that engaging providers from the beginning in the development of P4P systems is critical.[7] When developing measures, for example, eliciting the expertise of providers helps to eliminate skepticism among the provider community.

## Translation to the Medicare Environment – Stakeholder Engagement

**Providers**
Medicare has the unique ability to make significant strides toward improving the health care system through P4P because it contracts with nearly all providers and has the largest market share in the health insurance industry. Medicare's more than 42 million beneficiaries account for the bulk of payments to providers. However, market share alone may not be sufficient to motivate change among its providers. Significant efforts may be needed to engage providers, which will be vital to the development and implementation of a performance-based system. As shown in the Rewarding Results sites, engagement of providers is necessary to overcome barriers to program success.

There are several specific actions that Medicare can take to promote engagement among providers:

- Design programs that can be tailored to local provider preferences and behavior.
- Involve providers in the development and selection of performance measures.
- Facilitate discussion and communication between different provider types and encourage the sharing of performance information to inform referrals and improve professional responsibility.
- Solicit feedback on performance reporting, such as what format of performance reports is most helpful in determining what needs to be improved.
- Personalize communication with providers. (While this will be difficult because of the number of providers working with Medicare and the potentially high costs, it will likely prove substantially important to the success of the program).

**Consumers**
Medicare beneficiaries may prove more challenging to engage than consumers in Rewarding Results due to their advanced age and poorer health status on average. However, early results from Medicare Advantage and the implementation of Medicare Part D have shown that it is possible to engage consumers to make informed choices. There are a number of steps Medicare can take to encourage consumer engagement in P4P:

- Implement mechanisms that help providers engage patients in their care and provide reminders based on medical records of when services or treatments are needed. (This is especially helpful in the Medicare population due to the prevalence of chronic illnesses, which require continuous care.)
- Inform patients about any changes being implemented (e.g., local information sessions).
- Provide patients with easy-to-understand performance information on their providers or providers in their area so that they can make decisions regarding their own care if they choose.
- Ask for patient feedback at several stages during implementation so that if the changes being made administratively or clinically are negatively affecting the patients, they can be identified and remedied as quickly as possible.
- Involve community organizations, as was done with Medicare Part D, to provide personalized help to patients who want to become engaged.

**Health Plans**
While Medicare holds the largest market share among health plans and, therefore, drives the systems and policies that are used, it is important for all payers to work toward the same goals. Unlike Medicare, health plans rarely contract with all providers in a region, and most providers contract with more than one health plan, making it difficult for one health plan to make a significant impact on quality in a region. Stakeholder interviews have shown that providers are driven by compensation method. By working together as an industry, health plans should be able to drive providers to provide high-quality and medically appropriate care, just as they were able to control health care costs through managed care in the 1990s.[8]

While CMS can begin to motivate change with private health plans through changes to the Medicare Advantage regulations, involving private plans in the development of changes to the Medicare system might encourage more rapid expansion of performance measurement across the entire health care system. In general, the best way to engage private health plans is to involve them in all discussions, but there are more specific actions that Medicare can take to foster industry relationships:

- Utilize best practices that have been learned through private health plans' many P4P initiatives.
- Ensure that requirements under Medicare Advantage are consistent with those in Medicare fee-for-service so that providers are exposed to similar requirements from both programs.
- Involve health plan representatives in discussions with providers regarding performance measures so that plans use standardized performance measures, when possible.
- Work with plans to limit the administrative burdens associated with reporting performance measures across the entire health care system.

## II. Infrastructure

Infrastructure refers to the physical environment in which care occurs and the systems that are in place to provide care. In P4P, infrastructure includes a variety of data and health information technology (HIT) systems that are necessary to implement the payment system. It also includes the communications activities that provide feedback to the providers and signal when improvements are necessary.

## Lessons From Rewarding Results: Infrastructure

Experiences from Rewarding Results have shown that infrastructure changes are often necessary to meet the quality improvement goals of P4P. Changing the infrastructure to obtain P4P rewards is an opportunity for more effective and efficient redesign of care delivery and can result in more

accurate reporting on quality measures almost immediately. In the demonstrations, some participants questioned whether performance had really improved or whether the same level of quality had been maintained but was now being captured and reported.

While better measurement may not reflect actual changes in provider behavior, infrastructure changes are likely to increase overall quality outcomes. For example, at one BCBSM site, changes in infrastructure included changes in IT systems. The process of making IT system changes had the positive effect of opening communication channels between finance and nursing departments. Changes in infrastructure at MHQP and IHA included steps to achieve multiple payer data aggregation, uniform measurement sets, common reporting methods, and calculation of aggregate investments and payments for performance.

### Spotlight on Rochester Individual Practice Association/ Excellus Health Plan

RIPA/Excellus was able to achieve system change using timely and accurate provider feedback. They collaboratively developed an infrastructure so that physician-level reporting could be distributed three times per year. Patient registries were provided to identify missing quality services. In addition, provider relations personnel were hired to increase contact and communication with providers when reports needed clarification.

### Rewarding Results Experiences: Infrastructure and Return on Investment

A common complaint in the demonstration projects has been that the incentives are not large enough to offset the cost of required infrastructure changes. Some administrative costs are reduced when a stable, common set of evidence-based quality measures are identified by multiple payers. It is recommended that the cost of infrastructure change be estimated, measured and included in the calculation of return on investment to help identify opportunities for greater efficiency.

## Translation to the Medicare Environment: Infrastructure

Many of the infrastructure changes that Rewarding Results sites have undertaken are also relevant in the Medicare environment. For example, P4P requires purchasers and providers to improve their data aggregation and coding methods, measurement and reporting infrastructure. Several RR sites also found that health IT adoption was an important component of their infrastructure changes.

### Data and Coding

Medicare faces particularly challenging data and coding issues. To begin with, data run-out time for Medicare is approximately six to nine months. This means that it can take nine months for CMS to receive and adjudicate claims, then merge, scrub and aggregate the data in a uniform format. This long run-out period impedes Medicare's ability to measure and reward performance in a real-time fashion. Furthermore, Medicare data come to CMS from multiple sources (e.g., providers, claims payment contractors, Quality Improvement Organizations (QIOs) and Medicare Advantage plans) and any changes to coding and data formats that a P4P initiative may require can be extremely complicated and affect multiple steps along the data aggregation path. Theoretically, CMS is well positioned to facilitate the aggregation of Medicare data with other payers' data sources in a P4P initiative. However, the intensity and time required to aggregate Medicare-only data suggests that efforts to merge other payers' data may take years of work and significant investments.

The Medicare Payment Advisory Commission (MedPAC) has emphasized that data collection and analysis for P4P should not be unduly burdensome and should be based, where possible, on data sources from which Medicare already collects.[9] Most Medicare data are administrative, generally meaning that they come from provider claims or discharge abstracts. Administrative data are a good source of process information but yield fewer outcome measures than medical records. Provider acceptance of administrative data validity is generally less than for medical record data. Providers are often concerned that claims data are not adequate for measuring clinical performance.[10]

Currently, Medicare has very little access to clinical records and faces many challenges in presenting data at the patient level. While Medicare, as the largest national payer, could leverage its market share to promote increased investment in clinical data collection, the program currently faces significant challenges in acquiring clinical data in a uniform format. In addition, mechanisms to collect clinical data, such as chart abstraction, can be very labor intensive and expensive. If CMS adjusts for differences in patient populations, data collection and analysis costs and burdens will be even greater. This may be necessary since risk adjustment is generally less effective when administrative data are used because detailed clinical information is unavailable. In the short term, Medicare may want to consider investing in smaller-scale clinical data pilots that build off work that is already taking place in certain communities.

### Measures Selection

Developing a robust measurement set is crucial to P4P success. Rewarding Results sites have used a wide array of performance measures in their P4P initiatives, including many process measures, as well as some structure, outcome and patient experience measures. Three RR sites—BCBSM, IHA and RIPA/Excellus—included efficiency measures. Some measures used in RR sites were nationally developed (e.g., HEDIS or Joint Commission measures), while many others were developed locally (i.e., "homegrown" measures). As CMS continues to develop and refine its Medicare P4P programs, several questions related to measures selection will be dominant:

How should performance measures be selected? It is important to start by narrowing down the most relevant types of measures to concentrate on, then defining the criteria. Measures need to be valid, evidence-based and clinically relevant. Many providers may already be deluged with multiple performance measures from different payers. Therefore, coordination and alignment of measures with other key national organizations can decrease data collection and reporting burdens. There are, however, relatively few nationally accepted measures at present.

What types of measures should be included? Measure types include structure, process and outcome. Based on a 2005 Medvantage survey,

### Spotlight on Blue Cross of California

BCC provided physician scorecards (Figure 1) that show percentile rankings based on individual measures, as well as composite measures that rank physicians compared with others. It is important to note, however, that in the composite rankings, physicians were required to have a minimum of 35 cases within the composite.

91 percent of P4P programs target clinical quality measures, 50 percent target cost efficiency, 42 percent target HIT, and 37 percent include patient satisfaction measures.[11]

Efficiency measurement is an emerging area of focus. While cost efficiency has not been a focus heretofore, many purchasers are recognizing that efficiency and quality must be measured together. However, there is a lack of agreement on how to measure provider efficiency. The IOM defines efficiency as avoiding waste. Other definitions include relative level of resource consumption and associated costs in the production of health care services, and the relationship between resource inputs and care outputs. The science behind efficiency measurement is currently weaker than most clinical quality measures. To the extent CMS considers including efficiency measures in its P4P demonstrations, it has an opportunity to significantly influence how efficiency measurement is defined.

How many measures should be included? If the principal focus of the P4P initiative is on quality improvement, a small set of measures should be identified. If the focus is on rewarding quality and/or efficiency, a larger set of measures is more appropriate in order to get the "big picture." Because most P4P programs aim to support quality improvement as well as reward strong performance, Medicare may want to consider a middle ground that includes a manageable set of measures that are most representative of provider performance.

**HIT Adoption**

A 2005 survey of P4P initiatives showed that 42 percent of P4P programs incorporated specific HIT requirements.[12] These initiatives include electronic health records (EHRs) and electronic patient registries. For several Rewarding Results sites, HIT adoption has been a significant part of their experience, and the intersection of HIT and P4P has manifested itself in several ways

Many experts think that providers' ability to report data for P4P and other quality initiatives hinges on widespread adoption of HIT, and P4P has the potential to transform HIT into a tool for quality improvement rather than simply the translation of medical records into electronic format. CMS is in a unique position to significantly influence widespread HIT investment because of its dominant market share. For example, Medicare could incent electronic medical record adoption by paying for outcomes, helping incorporate lab data into electronic databases, or promoting the use of patient registries for quality improvement as a step toward more widespread adoption of electronic health records. However, some argue that without significant additional investment, only the best-resourced providers can invest in and redesign workflow processes to incorporate these technologies. Many health care providers still do not use HIT, and those who do have HIT systems may have systems that are incompatible with others and have difficulty transferring and translating data.

**Communications**

Many P4P sponsors have found that important infrastructural changes need to be made to facilitate provider feedback and communications. In particular, while it can be challenging to attract providers' attention, giving frequent, clear and actionable feedback to providers has been found to be extremely important to improving performance. Tools to raise provider awareness of P4P, as well as infrastructure to facilitate regular communications between providers (e.g., patient registries), must be considered.

Medicare faces certain unique challenges in providing actionable feedback to providers. In particular, there are implicit challenges in providing patient-level feedback to providers and promoting patient-level data sharing among providers in a fee-for-service environment. Each provider will have different systems of communication. Furthermore, targeting providers within organizations such as integrated delivery systems or other groups may add additional challenges. Consideration should be given for using QIOs, claims processing contractors or local medical societies, which already have established vehicles for communication.

In addition to considering communications infrastructure for providers, Medicare P4P programs must also consider beneficiary engagement. For example, Excellus has extensive experience with providing reports to patients. However, given Medicare's many rules to protect beneficiary-level data, there may be significant challenges with CMS providing beneficiaries access to their individual data.

### III. Incentive Methodology

Incentives can be financial or nonfinancial. Nonfinancial incentives are generally reputational in nature, but they may drive referrals to or from providers. Financial incentives can include a reallocation of existing payments or additional new payments.

Payment changes, whether attached to performance measures or not, have been shown to have significant effects on provider performance. Questions still remain as to how payment incentives will improve health care quality and how they will interact with the underlying payment methodology. It is also important to note that payment incentives do not always reach their intended target. Physician groups have been shown to change the structure of the payments to individual physicians, which could reverse or change the effects of the intended incentives or disincentives.[14]

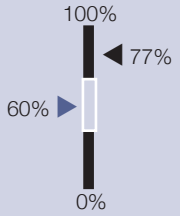
**Rewarding Results Experiences: Lessons in Provider Communication**

Rewarding Results demonstrations found that reliance upon traditional mail and regional meetings as channels of provider communication were not effective. Electronic communication alone also proved unsuccessful in one setting. Efforts to balance and personalize all communication channels were required. One demonstration successfully engaged a communication consultant, which resulted in renaming and redesigning the program.

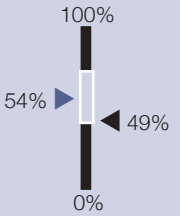
**Rewarding Results Experiences: Strategies for HIT Adoption**

Some RR sites have selected structure measures and developed incentives focusing on HIT adoption. For example, BCC measures timely electronic claims submission, IHA measures HIT investments, and BTE awards additional bonuses for innovative use of HIT. Some have invested in HIT as a strategy for improving quality and performance. RIPA/Excellus developed patient registries and physician notification systems as part of its quality improvement focus. BCC has used HIT tools (e.g., Web-based performance scorecards) as an incentive for high-performing physicians.

**Figure 1: Blue Cross of California Physician Scorecard[13]**

| Preventive Care and Screening | | | | | Below Peer Mean |
|---|---|---|---|---|---|
| Clinical Indicator | Observed Care | Eligible Cases | Your Rate ▶ (Confidence Interval) | Peer Mean ◀ | Composite Summary |
| Breast Cancer Screening | 22 | 30 | 73% | 87% |  |
| Cervical Cancer Screening | 19 | 30 | 63% | 81% | |
| Colorectal Cancer Screening | 31 | 60 | 52% | 63% | |
| **Composite Total** (Confidence Interval) | **72** | **120** | **60% (51% – 69%)** | **77%** | |



| Care Management | | | | | At Peer Mean |
|---|---|---|---|---|---|
| Clinical Indicator | Observed Care | Eligible Cases | Your Rate ▶ (Confidence Interval) | Peer Mean ◀ | Composite Summary |
| ACE Inhibitor Use in CHF | 4 | 9 | 44% | 43% |  |
| Compliance with Lipid Lowering Drugs | 11 | 16 | 69% | 47% | |
| Diabetes: Diabetic retinal exam | 18 | 62 | 29% | 32% | |
| Diabetes: Glycosylated hemoglobin measured | 52 | 62 | 84% | 74% | |
| Long-term control drugs for asthma | 5 | 6 | 83% | 77% | |
| Treatment of major depression – effective acute phase treatment | 2 | 9 | 22% | 44% | |
| Treatment of major depression – effective continuation phase treatment | 0 | 5 | 0% | 32% | |
| **Composite Total** | **92** | **169** | **54%** | **49%** | |



While payment changes have been shown to change the usage patterns of providers, they must be significant in order to have an impact. An evaluation of a performance-based reward system in upstate New York noted that reward programs have traditionally not been designed to produce real results in quality of care, stating that most do not use financial incentives that are significant to physicians.[15] A recent literature review of P4P initiatives found that, in general, "small-scale interventions do not appear to capture the attention of physicians."[16] Meredith B. Rosenthal and Richard G. Frank also noted that because of the multipayer system, any incentive by a single payer "must be a relatively large percentage of total reimbursement by that payer to justify any quality improvement effort with substantial fixed costs."[17] The good news for Medicare is that prior research has indicated that physicians change their behavior based on the dominant payer.[18]

The timing of incentives must be considered in a performance-based system, as it can affect behavior as much as the size of the incentives and efficacy of the measurements. Behavioral psychology has shown that in order to change behaviors, rewards must be provided in a timely and regular fashion.[19] A literature review done by Laura A. Peterson, et al. suggests that intermittent payments have more of an effect than "end-of-year" type incentives because of time delays between the actions being performed, their measurement and the reward.[20] Because of the variability of the health care system, it is nearly impossible to determine an exact number of incentive payments per year to best instigate change among providers.

Payers have several options when considering payout criteria. In the early stages of an initiative, payments for participation—especially reporting—are common. Key to this stage of P4P is the provision of timely and accurate information. Once the data are considered valid and reliable, then the accumulation of points for meeting pre-specified criteria must be developed. There are several options for payout criteria, including meeting certain thresholds, improvement or rankings.

Each method for determining payout criteria has advantages for a provider at a specific level of performance. Threshold performance distributions tend to favor existing high performers, since they are near the top to begin with. On the other hand, percent improvement distribution criteria tend to favor low performers, since they have more room to improve. Finally, distributions based on ranking by individual relative score tend to result in greater competition among providers.

One of the major critiques of performance-based payment systems lies in determining what part of the health care system affected a particular outcome.[21] For example, how do you account for a physician

who provides additional preventive or monitoring services and therefore keeps a patient out of the hospital? While the physician is utilizing additional services, they likely cost less than the avoided hospitalization. In some cases, the physician may consult over the phone or by e-mail and, therefore, is not traditionally compensated for the time being spent or for helping to avoid a costly hospitalization. Recently, work has been done exploring the use of various methods of physician profiling and their ability to effectively attribute care to an individual provider level.[22] In June 2006, MedPAC reported that episode groupers could link quality indicators to responsible physicians the majority of the time, and that in most cases a single physician was responsible.[23] These groupers, however, determine attribution between physicians and do not account for differences in costs and savings for hospitals, skilled nursing facilities, hospice or any other provider.

### Lessons From Rewarding Results – Incentive Methodology

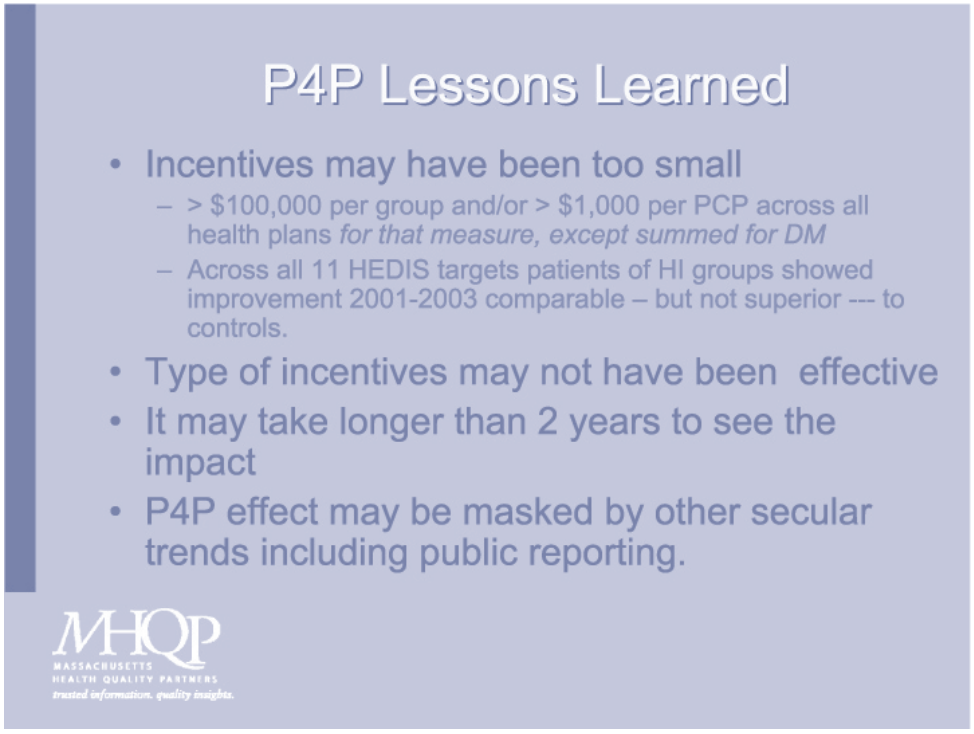
Nonfinancial incentives appeal to both providers' competitive nature and professional ethos to excel. At this early stage, it is unclear whether this method affects a provider's market share. However, payers in the MHQP and IHA markets are designing incentives for consumers to select providers with higher-quality scores. Consumer incentives have included reduced co-pay requirements and lower deductibles.

Rewarding Results methods to provide financial incentives include placing a percentage of financial withhold at risk, a direct bonus or a combination of both approaches. At RIPA, 10 percent of providers' HMO fees were set aside as withhold to be at risk. Between 50 percent and 150 percent of this withhold amount could be returned to the provider depending upon their ranked score. At BCBSM, hospitals were offered a direct 4 percent bonus in DRG payments for meeting quality measures. BTE offered physicians a $100 bonus per member per year for obtaining NCQA recognition for excellence in diabetic care.

### Spotlight on Integrated Healthcare Association

Initial efforts at IHA have included peer recognition for the best performers. In addition, there was the confidential release of blinded information that displayed a provider's relative rank in selected quality performance measures. Once providers were presented with opportunities to understand and verify the scoring process, relative rankings were made public.

**Figure 2: MHQP Lessons Learned**




For an incentive methodology to be effective, providers must understand and be in control of the services that are included in the quality measurements. Payers must know that payments are justified. In general, the available pool for payments is usually determined by the strategic self-interest of the payers without reliable cost-to-implement information from providers.

Reallocating or redistributing existing fee payments based on new quality criteria weakens the power of payers' financial incentives. Additional payments based on new quality criteria strengthen the power of financial incentives. The distinction between financial incentives from new money and old money is not lost on providers. When total incentive payments do not fully compensate for the cost of changes required in provider behavior, it is not likely that provider engagement will be sustained. In addition, as the methodology for measurement, calculation and reporting incentives becomes more complex, the understanding and acceptance of incentives is weakened.

The more the payout is directly related to the quality of service and the shorter the timing between the two, the better. One demonstration shifted from an annual payout to a rolling fee schedule enhancement and increased the potential increases to 14 percent for PCPs and 12 percent for specialists. Another demonstration changed its annual physician feedback to three times per year to increase provider's opportunities to improve their quality of care.

MHQP provided several important lessons from its project in a PowerPoint presentation at the Rewarding Results: Lessons Learned meeting on December 15, 2006, reproduced in Figure 2.[24] While specific to the MHQP project, these points provide significant lessons that can be translated to CMS efforts.

Other projects found that their initial incentive methodology may not have generated the outcomes expected. Although, as noted in the MHQP slide, there are a number of things that could mitigate incentive efforts. BTE, for example, learned that offering incentives at 5 percent to 10 percent of physician revenue (roughly $10,000 to $15,000 per physician) provided meaningful encouragement to engage in performance improvement efforts.[25] On the other hand, the RIPA project suggested that it would have redesigned its payout structure from a single, large, annual incentive to a number of smaller, more frequent payments.[26]

In every demonstration, providers with existing quality infrastructure seem most likely to succeed, regardless of payout criteria. Payers have the opportunity to collaborate on a standard set of measures to increase the total patient population measured and rewarded, as well as reduce the administrative burden of providers. Meanwhile, providers have the opportunity to aggregate incentive payments from multiple payers to compensate for care design changes and infrastructure required to reach higher-quality scores.

### Translation to the Medicare Environment – Incentive Methodology

**Size of Incentives**

Medicare could have a significant impact on provider behavior because, in most cases, it is the dominant payer, and other payers follow its payment lead. Medicare can take the following specific actions to help ensure that quality changes occur:

- Use large-scale financial incentives based on the fixed costs of services to elicit a change in usage behavior by providers.
- Work with other payers to align financial incentives collectively so that the incentives do not work against each other.
- Scale financial incentives based on the "additional" work being done or time spent by the provider.
- Utilize both financial incentives and withholds on top of base payments.

**Timing of Incentives**

As several of the Rewarding Results sites have reported, frequent payments and feedback encourage greater change among providers. But providing payments and feedback often, without being an overwhelming administrative burden to either Medicare or the providers, will prove a challenge. To execute a frequent payment and reporting system, an infrastructure must exist to support rapid data analysis and accounting mechanisms. There are several actions that Medicare can take to ensure that the timing of financial incentives is appropriate for a P4P system:

- Improve data sharing within CMS and its contractors so that available data can be used by multiple offices at the same time, allowing performance data to be evaluated more quickly and incentive payments to be paid out closer to the time of service.
- Establish incentive payments to occur several times a year.
- Set explicit goals for timelines of incentive payments as the initiative is put into place, with timelines becoming closer to the time of service as time goes on.

**Attribution**

The attribution of care and savings is extremely important for Medicare because the payment structures and financing for Part A and Part B are distinctive. The Physician Group Practice (PGP) demonstration has broken new ground in compensating large physician groups for realized savings in hospital and other Part A services. In addition, Section 5007 of the Deficit Reduction Act of 2005 (DRA) called for Medicare to establish a demonstration project that would share financial gains between hospitals and physicians that work together to improve the quality and efficiency of care.[27] These demonstrations should serve as a model for attributing savings between different provider groups.

Although significantly more research needs to be done in this area, Medicare can begin to distribute incentives accordingly by:

- Aligning incentives for providers to effectively provide preventive and maintenance care to avoid hospitalizations across all providers.
- Encouraging partnerships among providers, enabling smooth tracking of patients and attribution of savings.
- Promoting accountability by linking providers to the patient populations they serve when measuring quality.
- Establishing and evaluating acknowledged quality measures for transitions between settings.

### IV. Continuous Quality Improvement

One of the most important goals of P4P is to motivate providers to improve quality. In order to achieve reliable improvements in quality outcomes, the processes of care need to be carefully examined and reengineered to make it easier for physicians to do the right thing. Identifying the processes of care that need to be reengineered is an intensive activity that cannot be dictated from a payer or a central source but must be undertaken locally. In many cases, local expert panels or medical specialty societies have been used successfully to set standards, but providers within the organization must then examine current performance compared to the standard and identify a pathway for improvement. Within organizations, blinded peer comparison has been a helpful tool.

The process of taking quality measures, identifying care process systems that impact those measures and systematically examining those processes for potential improvement is a time-intensive, laborious activity, but it is essential to identify points of weakness in the system. Data are crucial for this activity and need to be timely so that they can truly impact performance.

While systematic quality improvement is the ultimate goal, continuous quality improvement and care process redesign require a deep focus on a small number of measures rather than a large data set. Each measure needs to be taken apart: examination of the patient population, what services were or were not provided, and by whom. Once sustained improvements have been obtained, new measurement sets can be updated and new goals can be set. This may require more customization of measures to particular providers.

**Patient Registries**

Patient registries are a way of organizing patients by relevant clinical information, such as diagnosis. Registries can be simple lists or more interactive components of an electronic health record. Registries are an important tool in care process redesign because they allow for like cases to be grouped together to facilitate similar care being provided to each patient. In addition, care can be analyzed at the patient or practice level rather than at the visit or service level. Patient registries can also be used to bring patients into the office for recommended follow-up care or preventive services to increase the likelihood of improved quality.

**Electronic Health Records**

Electronic health records (EHRs) are an important quality improvement tool because they facilitate the collection and analysis of data. In addition, they can be developed to incorporate reminders and pathways that make it easier for providers to give the appropriate care to patients. For example, EHRs can be designed with reminders that pop up when routine or preventive care is due and can warn physicians before prescribing contraindicated medications (i.e., medications that will interact with existing medications). While EHRs are a useful tool and will significantly reduce data collection and analysis costs, it is possible to implement quality improvement without EHRs.

Other types of technology that can be utilized to improve care include electronic prescribing tools that link physician offices with pharmacies and tools that link physician offices with laboratories for faster reporting of results.

**Technical Assistance**

Technical assistance is an important tool for care process redesign, especially for smaller practices that may not have sufficient staff or sophistication to analyze data and map measures to processes and outcomes. Some technical assistance methods that have proven useful include facilitating discussion among providers and case study analysis of high performers to identify any successful processes that are transferable to other settings.

### Rewarding Results Experiences: Evolution of Measures

In all the demonstrations, national accrediting organizations like NCQA and the Joint Commission, medical specialty societies, and local expert panels reviewed evidence-based medicine to set standards for continuous quality improvement. Annual provider scores were reviewed and adjusted to higher standards. For IHA and BCBSM, this led to retiring standards when the majority of quality improvement benefits were obtained. Two measures that were retired were pediatric immunizations and antibiotic administration time for pneumonia patients.

### Spotlight on Local Initiative Rewarding Results (LIRR)

Through provider interviews, LIRR realized that P4P had helped change the way well-baby care was provided. In many instances, providers reported increased outreach to schedule visits; others reported greater follow-up to their usual reminders. In a small number of instances, providers utilized sick visits to also perform well-baby care for those who had not scheduled it previously or were overdue.

#### Lessons From Rewarding Results – Continuous Quality Improvement

An important lesson learned from Rewarding Results is that care process redesign and continuous quality improvement methods are compatible with P4P initiatives because these initiatives are often data driven and designed to help physicians improve their work. For example, BCBSM did not pay out any incentives to hospitals unless performance met or exceeded the highest mean since the inception of the program. Continuous improvement was both expected and required.

When sufficient improvement is made, it may be appropriate to retire measures to focus on other aspects of care. However, this must be done carefully to avoid too many changes that may result in unsustainable quality improvement and wasted time and costs for providers to reach quality goals.

Changes to the measurement set to reflect progress and facilitate improvement must be made with care. IHA's Technical Committee proposes changes to its uniform measurement set annually to reflect progress and continuous quality improvement initiatives. Public review and comment periods are offered to all stakeholders, followed by rigorous testing of measurement specifications and distribution of sample results. During a normal two-year process, new standards for P4P are formalized. This consultative process ensures that changes are agreed upon by all stakeholders and facilitates their implementation.

Data challenges related to billing codes and data collection and aggregation are frequently a barrier to continuous quality improvement. While new data collection via chart review was feasible in the BTE demonstration, all the other demonstrations relied upon administrative claims data, so inaccuracies in data must be factored into expectations. For example, at health plans participating with IHA it was anticipated that data collection inaccuracies would make 100 percent compliance with quality improvement goals unrealistic. One health plan paid 100 percent incentives for clinical and service quality if the group scored at or above the 75th percentile. Until data accuracy improves, P4P initiatives favor payer-provider negotiation on selected quality targets and reasonable expected ceiling for improvement rates. It is anticipated that expanding HIT, including the EHR, will improve data accuracy and make 100 percent compliance feasible.

Accurate and timely performance feedback tools should precede annual performance measurement. At MHQP, provider groups received condition-specific patient registries and timely patient-level actionable information. At RIPA, a medical director was charged with providing training opportunities and technical assistance to identify patient-level interventions. Provider-specific, blinded, peer comparison of adherence to quality measures proved an excellent motivator for quality improvement.

BCBSM suggested that incentives should focus on chronic care management and prevention and that rewards should be given to physicians who participate in collaborative approaches to identify new process and outcome measures.[28] IHA demonstrated that limiting data collection to electronic efforts limits the capabilities, but it recognized that forcing the collection of electronic data through EHRs would improve data collection and exchange. Noting that administrative data is not sufficient for meaningful measurement, IHA recommended utilizing patient registries as a source of quality improvement efforts.

While data are the key to quality improvement, various tools have been used by other organizations to organize the data in a way that facilitates understanding of existing limitations and potential for change. Some demonstration sites, such as LIRR, have seen initial success with P4P initiatives but believe that further improvement may require technical assistance and quality improvement supports to improve impact.

While formal facilitation of quality improvement activities—where knowledgeable staff provides benchmarks, analytic support and guidance to initial activities—is particularly beneficial, providers sitting down together informally to discuss baseline measures and identify potential paths to improvement can also be helpful. According to the RIPA project: "There was an observed tension between primary care and specialists…RIPA put everyone in the same room to make them discuss things together."

A number of sites suggested that engaging consumers in the process would be an important addition to any program and one way of doing this is to foster increased relationships with patients outside of the office. This could include a range of improvements, such as encouraging the use of personal health records that can be updated by any provider or implementing a nurse outreach program to contact patients in need of chronic care by phone or e-mail between office visits.

### Spotlight on Bridges to Excellence

BTE structured its rewards so that offices will reinvest the money they get from the rewards. If providers can be encouraged to invest the money into structural improvements to sustain quality improvement, the demonstration has the potential to have impact far beyond the demonstration period.

### Spotlight on Blue Cross Blue Shield of Michigan

In the BCBSM demonstration, several successfully engaged hospitals reported that redesigned care systems enabled improved scores. For example, facilitating the discharge planning process made it easier for physicians to comply with quality measures and achieve higher-quality scores.

### Translation to the Medicare Environment – Continuous Quality Improvement

**Care Process Redesign**

Making timely data available to inform quality improvement activities has the potential to be challenging under Medicare, given the volume of data to be collected and analyzed and depending on the type of data to be used. To the extent that claims are used for quality measurement, providers may need to develop the capability to measure and track their own performance or special arrangements may need to be made with a contractor to provide early feedback. If other data are used, such as clinical data from chart reviews or other sources, arrangements may need to be made for providers to retain access to their own data upon collection and submission to CMS.

Medicare can foster continuous quality improvement and process redesign in a number of ways:

- It can allow and encourage some customization of measures by provider groups or organizations so that local providers can identify quality improvement targets that are most relevant to them. This might require a standardized set of measures and an optional list of measures that local providers could choose from.
- It can encourage solo practitioners and small groups at a local level to organize into collaboratives. QIOs might provide technical assistance and an infrastructure for these local groups to analyze data, map data to processes and identify potential process changes.
- It can document successful local efforts through case studies and disseminate them widely. These case studies provide concrete examples for other providers contemplating change and have proven useful in some demonstrations.

**Tools**

There are a number of tools that CMS could provide to facilitate quality improvement:

- CMS could encourage the use of registries within large provider groups and regionally in rural areas or among solo and small group practices. Software and training could be provided free to practices, or incentives could be built into the system to reward providers who use registries. Since most registries are disease specific, care must be taken to incorporate the recognition that patients—especially Medicare beneficiaries—may have multiple chronic diseases, otherwise, these registries can promote fragmentation.
- While EHRs and other HIT are not a necessary component of quality improvement, they make data collection and analysis easier. CMS could incent the acquisition of EHRs and other HIT by making HIT acquisition and use one of the performance metrics.
- The QIOs or other contractors could provide technical assistance, especially for small practices.

## V. Transparency

Transparency involves providing detailed, reliable, comparable cost and quality information to all stakeholders. Some of the information required for transparency includes explicit quality standards, comparable payout standards and the amount of rewards for high-quality health care.

P4P and public reporting are often viewed as complementary initiatives. Public reporting is sometimes used as a phase-in strategy for P4P, giving purchasers and providers time to focus on data collection and measurement before there is a financial stake. While many public reporting or "honor roll" programs are used as incentives in P4P, few studies have looked at the effects of report cards relative to P4P, and the value of public reporting remains undemonstrated. One study of facilities enrolled in the Premier demonstration did find that Premier facilities experienced more quality improvements than those hospitals participating only in a public reporting program.[29]

Advocates for public reporting argue that it injects competition, helps providers benchmark their performance, encourages purchasers to reward quality and efficiency, and helps patients make informed choices.[30] There is general agreement that public reporting must be timely, relevant and intelligible. Critics of public reporting commonly cite that cost and charge information are complicated to present, results may not be comparable, data accuracy and validity are difficult to assess, sample sizes may be insufficient, and risk-adjustment may not be applied appropriately. Public reporting may also have some unintended and negative consequences on health care.[31] For example, providers may avoid sicker patients in an attempt to improve their quality rankings, or they may be encouraged to achieve "target rates" for certain interventions even when it may be inappropriate for certain patients.

### Lessons From Rewarding Results - Transparency

The Rewarding Results experience has shown that transparency encourages openness; enhances trust and accountability; and can be facilitated by payer and provider collaborations on selecting measures, aggregating data, and designing incentives. Public reporting initiatives have been found to motivate provider participation in RR sites. For example, IHA uses a public scorecard comparing physician group performance as non-financial incentive, and MHQP created a new public reporting initiative to compare physician performance alongside its P4P strategy.

At IHA, transparency is facilitated by payer and provider collaborations on uniform measurement, data aggregation, common reporting and payment by six health plans. IHA reports the total payments for the IHA uniform measurement set, as well as the total payments for non-IHA measures for commercial HMO and POS members. In addition, methods for determining payments are made publicly available, including absolute thresholds and relative percentile ranking standards. The eligibility criteria are made public, including specific clinical, patient experience and information technology measures.

As national standards evolve from these early demonstrations and EHRs become more common, there will be increasing opportunity to provide transparent overall cost and quality information. MHQP and IHA learned that initial measurement, payment and reporting is statistically more reliable at the highest organizational level (physician group vs. individual physician) and with the largest population (multiple payers vs. individual plan). Meanwhile, comparable group, regional or national averages can serve to overcome statistical issues when physician profiling is based upon limited patient encounters

### Translation to the Medicare Environment - Transparency

As a government program, Medicare must operate in complete transparency. In addition, there is a tremendous push towards more transparency across the health care system. The Bush administration's Value-Driven Healthcare agenda promotes public reporting and transparency. In February 2007, CMS announced plans to make physician performance data available to

**Spotlight on Massachusetts Health Quality Partners**

MHQP's explicit, multiyear, public reporting plan designed to progressively report from the highest organizational level toward the individual provider motivates greater provider participation.

Medicare beneficiaries through the Better Quality Information to Improve Care for Medicare Beneficiaries (BQI) Project. Regional collaboratives will combine Medicare claims data with other insurers' data to produce comprehensive information on health care providers. The goals are twofold: to help physicians assess and improve their performance and to help beneficiaries make more informed decisions.

Medicare has a long history of reporting initiatives. In 1984, HCFA (now CMS) began publicly reporting hospital mortality rates for Medicare; however, these reports were controversial and short-lived. More recently, Medicare developed a number of public reporting initiatives for different provider settings. In June 2006, CMS began posting hospital-specific charge and payment information for 30 common procedures. In addition, Medicare payment information is now available at the county level for physician services, ambulatory surgery centers and hospital outpatient departments.

As Medicare considers public reporting investments for the future, it will be important to think about how public reports like Hospital Compare can provide a feedback loop for quality improvement at the provider level. CMS may want to consider a hybrid reporting arrangement in which public reports are available at the group level and private reports are available for individual providers for quality improvement purposes.

## VI. Measuring and Evaluating Impact

Although there is much interest in the potential for P4P to improve quality, and CMS is committed to implementing this approach, it is not a foregone conclusion that P4P will yield quality improvement.[32] Measuring and evaluating impact is critically important to ensure that scarce resources are used appropriately and the intervention is having the desired impact.

Measuring and evaluating impact from P4P should ideally occur at two levels: at the level of the provider that is being assessed to ensure that quality is being improved and at the programmatic level to ensure that program resources are being used appropriately and effectively. While this discussion is mostly focused on evaluating impact at the provider level, many of the aspects will also be relevant to programmatic evaluation.

In general, evaluation of impact can be complicated by a number of factors:

- Timelines can impact the potential for success. Some measures could result in improvement more quickly, while others may take some time to become apparent. Selecting the right measures and having appropriate expectations about when effects are likely is important.
- Impact can vary by population or disease group. Some provider groups may have a more difficult time showing improvement because of their patient population. However, risk adjusting the population by disease status or other factors may be somewhat controversial. Should providers that treat a more challenging population be held to a lower standard?
- Data quality may vary by provider. Steps to ensure complete reporting and audits to ensure accuracy are important to be certain that observed changes are due to changes in quality and not changes in reporting or data accuracy.

Determining attribution is difficult. Most patients are seen by a number of providers, and holding individual providers responsible for care may pose problems. In addition, the current fragmented environment in which different providers are reimbursed in silos offers little incentive or ability to recognize improvements in quality implemented by one provider group that impacts care in another setting (e.g., improvements in outpatient care may reduce hospitalizations). Despite these challenges, attribution is critical to defining denominators from which to measure performance for either quality or efficiency.

### Lessons From Rewarding Results – Measuring and Evaluating Impact

Rewarding Results and other programs demonstrate the potential for improvement, but there are numerous design and implementation features that could facilitate success or result in no impact. In spite of the potential for confounding effects, a systematic evaluation of P4P initiatives is required to inform senior decision makers about the role and value of incentives to improve quality. However, because of the complexity, experts in evaluation and attribution need to be involved as early as feasible in P4P programs.

In the absence of a randomized control trial, well-designed natural experiments and before-and-after observations provide opportunities to evaluate the impact of P4P in general and the impact of potential design and implementation options. RR evaluators have found that maintaining the continuity of the same providers, measured with the same evidence-based quality standards, and using the same reward calculation throughout the project make for a cleaner evaluation.

Demonstrations used multiple approaches to address the statistical concerns for a provider with a small number of patient encounters. At RIPA, the plan was to substitute the medical specialty or group average quality score. At BCC, individual measures are combined into composite scores and ranked as above, at or below peer mean scores. These composite scores include clinical quality, generic prescribing and administrative compliance, and efficiency categories. Efficiency measures often include radiology and imaging utilization and hospital length of stay.

There were some early lessons about calculating an all-stakeholder return on investment (ROI). Efforts at BCBSM, RIPA and BTE have resulted in P4P being used as a method to evaluate the investment and return to the payers and providers from specific initiatives. Early results indicate that incentives may not cover the cost of required interventions for each of the stakeholders and benefits may not flow equitably. One approach to deal with the inequitable returns on investment is to evaluate the ROI across all stakeholders, in aggregate, rather than by a plan or provider, and consider apportioning returns to each stakeholder to cover the required costs for sustaining quality enhancing interventions.

### Translation to the Medicare Environment – Measuring and Evaluating Impact

Evaluation of impact in the Medicare environment poses its own challenges over and above the challenges experienced in RR. However, the experiences of RR can provide some insights into possible options for Medicare.

#### Comparison Groups

Measuring the impact of P4P is difficult because often there is no real comparison group. Evaluation of impact usually involves before and after comparisons, but temporal trends, differences in providers and/or patients, changes to data collection, and other external factors can influence outcomes and result in an over or under assessment of the true impact.

RR sites have used creative ways of identifying comparison groups, including comparing performance among different regions and comparing the different levels and types of incentives for improved performance of different models. As P4P is implemented in demonstration sites, care should be taken to identify appropriate comparison groups. While regional variation in intensity and costs may make comparison of different regions difficult, because of Medicare's large market share it may be possible to randomize within a region to create comparison groups.

**Patient Identification and Targeting**
Quality measures used in P4P may not be applicable to all patients, and identifying the right patients may be difficult, especially when claims are used. This is because claims data may not include sufficient clinical detail to identify the appropriate subgroup for which measures can be applied. Care needs to be taken so that patient identification is specified in detail and records are audited to ensure that comparable patients are enrolled, measured and reported on accurately. Data collection for HEDIS provides a good model for P4P activities.

**Small Sample Sizes**
Accurate measurement requires sufficient sample size so that results can be attributed to differences in quality and not random fluctuations in population risk or data collection problems. Although Medicare's large market share makes small sample sizes less of a problem than for RR sites, sufficient sample size may be a significant problem in small or solo practices and for measures that are disease specific. One solution is the aggregation of data to larger units, such as regions. Elliot Fisher and colleagues have recommended that data for P4P be collected on a regional basis—based on hospital catchment areas—to alleviate the problems of small sample sizes.[33] There may be other levels of aggregation that are possible.

**Attribution**
On average, Medicare beneficiaries see two primary care physicians and five specialists in four different practices a year. Attributing outcomes to a particular provider may be difficult when the patient may have seen numerous providers.[34] In the past, CMS has used algorithms to determine attribution, and MedPAC has been examining the utility of different groupers for attribution.[35] Recently published research findings substantiate earlier indications that no more than half of beneficiaries' visits would be accurately assigned under current models.[36] Further research is needed in this area to determine how to best attribute beneficiaries' care so that financial incentives can accurately motivate physicians based on the care they provide.

However attribution is determined, most methods make determinations retrospectively and may be difficult to feed back into a quality improvement process at the provider level and create opportunities for providers to standardize care and improve quality for all populations (e.g., diabetics) and payers (e.g., commercial and Medicare). Also, attribution generally reduces sample sizes since the unit of analysis becomes the individual physician rather than the group, potentially exacerbating the small sample size problem.

**Return on Investment**
There are two levels for considering ROI:

- At the provider level, does the potential reward provide sufficient return on investment for infrastructure and other changes that are needed to achieve the reward?
- At the payer level, does the benefit achieved—in reduced expenditures or improved quality—exceed the investment in additional incentive payments?

For providers, the determination depends on the type of behavior that is being incented. For certain types of behaviors—especially preventive screenings—the level of investment to achieve a bonus may be relatively low. Other types of improvements may require relatively large investments, such as in infrastructure when HIT is involved.

**Evolution in Strategies**
As Medicare moves forward with P4P initiatives, recognizing emerging trends and challenges and incorporating new knowledge will be extremely important in designing and implementing these initiatives.

**Summary of Recommendations**
Previous P4P demonstrations, like Rewarding Results, provide important lessons for future Medicare P4P efforts. This section provides a summary of recommendations that CMS may want to consider as it moves forward with designing, implementing, evaluating and refining its Medicare P4P programs. These recommendations are discussed in more detail in the sections above.

**Stakeholder Engagement**
There are several specific actions that Medicare can take to promote provider engagement in P4P:

- Design programs that can be tailored to local provider preferences and behavior.
- Involve providers in the development and selection of performance measures.
- Facilitate discussion and communication between different provider types and encourage the sharing of performance information to inform referrals and improve professional responsibility.
- Solicit feedback on performance reporting, such as what format of performance reports is most helpful in determining what needs to be improved.
- Personalize communication with providers.

There are a number of steps Medicare can take to encourage consumer engagement:

- Implement mechanisms that help providers engage patients in their care, such as providing reminders based on medical records of when services or treatments are needed.
- Inform patients about any changes being implemented (e.g., local information sessions).
- Provide patients with easy-to-understand performance information on their providers or providers in their area, so that they can make decisions regarding their own care if they choose.
- Ask for patient feedback at several stages during implementation.
- Involve community organizations in providing personalized assistance to patients who want to become engaged.

Specific actions that Medicare can take to foster health plan engagement include:

- Utilize best practices that have been learned through private health plans' many P4P initiatives.
- Ensure that requirements under Medicare Advantage are consistent with those in Medicare fee-for-service.
- Involve health plan representatives in discussions with providers regarding performance measures so that plans use standardized performance measures, when possible.
- Work with plans to limit the administrative burdens associated with reporting performance measures across the entire health care system.

**Infrastructure**
Medicare can take several steps to improve data management:

- Facilitate the aggregation of Medicare data with other payers' data sources in a P4P initiative.
- Promote increased investment in clinical data collection, such as investing in smaller-scale clinical data pilots that build off work that is already taking place in some communities.

In order to improve the measures selection process, Medicare could:

- Coordinate and align measures with other key national organizations to decrease data collection and reporting burdens on providers.
- Monitor and influence how efficiency measurement is defined.
- Use a manageable set of measures that can both support quality improvement and reward strong performance.

To influence widespread HIT investment, Medicare could:

- Incent electronic medical record adoption by paying for improvements in care coordination and quality performance.
- Assist with incorporating lab data into electronic databases.
- Promote the use of patient registries on the path toward more wide-scale adoption of electronic medical records.

To improve communications with providers and beneficiaries, Medicare should:

- Consider using QIOs, claims processing contractors or local medical societies to provide actionable feedback to providers.
- Consider ways to improve communications with beneficiaries and increase their engagement in quality improvement.

**Incentive Methodology**

Medicare can take specific actions to help ensure that incentives lead to quality improvements:

- Use large-scale financial incentives based on the fixed costs of services to elicit a change in usage behavior by providers.
- Work with other payers to align financial incentives collectively so that the incentives do not work against each other.
- Scale financial incentives based on the "additional" work being done or time spent by the provider.
- Utilize both financial incentives and withholds on top of base payments.

There are several actions that Medicare can take to ensure that the timing of financial incentives is appropriate:

- Improve data sharing within CMS and its contractors so that available data can be used by multiple offices at the same time.
- Establish incentive payments to occur several times a year.
- Set explicit goals for timelines of incentive payments as the initiative is put into place, with timelines becoming closer to the time of service as time goes on.

Although significantly more research needs to be done on how outcomes are attributed, Medicare can begin to distribute incentives accordingly:

- Align incentives for providers to effectively provide preventive and maintenance care to avoid hospitalizations across all providers.
- Encourage partnerships among providers, enabling smooth tracking and attribution of savings.
- Establish and evaluate acknowledged quality measures for transitions between settings.

**Continuous Quality Improvement**

Medicare can foster continuous quality improvement and process redesign in a number of ways:

- Allow and encourage some customization of measures by provider groups or organizations so that local providers can identify quality improvement targets that are most relevant to them.
- Encourage solo practitioners and small groups to organize into collaboratives at a local level, with QIOs possibly providing technical assistance and an infrastructure for these local groups to analyze data, map data to processes and identify potential process changes.
- Document successful local efforts through case studies and disseminate them widely.

There are a number of tools that CMS could provide to facilitate quality improvement:

- Encourage the use of registries within large provider groups and regionally in rural areas or among solo and small group practices, with software and training being provided free to practices or incentives being built into the system to reward providers who use registries.
- Incent the acquisition of EHRs and other HIT by making HIT acquisition and use one of the performance metrics.
- Have QIOs or other contractors provide technical assistance, especially for small practices.

### Transparency

As Medicare considers public reporting investments for the future:

- Consider ways to support the use of public reports like Hospital Compare as a means of providing a feedback loop for quality improvement at the provider level.
- Consider using a hybrid reporting arrangement in which public reports are available at the group level and private reports are available for individual providers for quality improvement purposes.

### Measuring and Evaluating Impact

In order to measure and evaluate the impact of P4P demonstrations, CMS can:

- Carefully identify appropriate comparison groups (While regional variation in intensity and costs may make comparison of different regions difficult, it may be possible to randomize within a region to create comparison groups).
- Aggregate data to larger units in cases where small sample sizes may be a significant problem (e.g., in small or solo practices and for measures that are disease specific).
- Support research into how to best attribute beneficiaries' care so that financial incentives can accurately motivate physicians based on the care they provide.

## Next Generation Pay-for-Performance

As P4P initiatives mature and become more common in different care settings, the next generation of P4P will likely include new types of measures, shared savings and gain-sharing arrangements, a focus on underserved populations, and consumer incentives. In order to sustain quality improvement, P4P programs need to evolve and be refreshed as goals change and targets are achieved.

### New Areas of Focus

In addition to paying for performance on quality and efficiency measures, there will likely be new incentive arrangements that reward data collection and reporting functions or pay more for preventive care. In addition, initiatives to incent care coordination and chronic care management of populations that are underserved or more complex may become more common. Incentives could include larger, incremental payments for more complex care, or capital grants, technical assistance or special trainings to support care management.[37]

### New Measures

Evolving P4P programs will adopt new measures, update old measures as clinical evidence changes, and potentially retire measures that are no longer useful. P4P sponsors may find that new measures will be needed to support changing priorities and increase emphasis on certain areas. In

February 2006, the American Medical Association announced to Congress that it would soon develop more than 100 new performance measures.[38]

Efficiency will likely be a new area of measurement focus for the future. While there is currently a great deal of energy being devoted to measuring and incenting efficiency (approximately 50 percent of P4P programs currently target cost efficiency), there is little consensus on how it should be measured.

As best practices in efficiency measurement are identified, an increasing number of P4P sponsors will likely adopt efficiency measurement as a complement to quality measurement. Medicare may need to look at potential overuse reduction opportunities (e.g., in Medicare Part A) as a mode for funding a P4P bonus pool and/or ensuring budget neutrality. Medicare could solicit the help of the QIOs and partner with societies to set standards for measuring overuse and underuse, and evaluate potential unintended consequences of efficiency measurement.

### Consumer Incentives

As a complement to provider incentives, P4P sponsors may increasingly offer consumer incentives, particularly for chronically ill patients and those for whom adherence is a major challenge. There is evidence that financial incentives directed at consumers can enhance consumer compliance and increase positive preventive behaviors, including immunization rates.[39]

### Shared Savings and Gain Sharing

Similar to the Medicare Physician Group Practice, Health Care Quality demonstrations and RIPA, P4P programs will likely increasingly emphasize shared savings. Shared savings refers to arrangements where payers provide certain performance incentives to providers. The performance incentive payments are derived from savings that accrue to the payer but are achieved from providers' care management and clinical practice initiatives.

Gain sharing arrangements are also likely to increase. Gain sharing is when hospitals share savings with physicians for physician activities that are aimed at controlling hospital costs, such as appropriate use of imaging and testing services, good prescribing practices, reductions in medical errors, and use of outpatient instead of inpatient services.[40] Gain sharing is considered a potentially important strategy for aligning hospital and physician incentives to improve care quality.

### Care Coordination

Care coordination is intended to improve the patient care process and outcomes for those with chronic illness and co-morbidities. Coordination of care can be a complex process requiring significant infrastructure. In the future, care coordination strategies will likely evolve to better align incentives between purchasers, health plans, providers and beneficiaries. P4P has the potential to motivate providers and insurers to take a more active role in chronic care management and improve care coordination across settings and over time.[41]

### Strategies for Reducing Fragmentation

P4P offers some promise for reducing fragmentation in care. For example, MedPAC has recommended that Medicare consider adding P4P measures such as quality of transfers across settings and patient functioning a year after hospitalization.

### Co-morbidities Management

Many Medicare beneficiaries suffer from co-morbidities. It is important that Medicare avoid measuring and reporting on metrics that are not clinically relevant to the population. To date, however, there has been little focus on developing co-morbidities measures.

### Point of Care Transformation

P4P is an important mechanism to achieve transformation at the point of care. The appropriate use of data is a key element in successful efforts to transform care. Data requirements under P4P can be a valuable mechanism for providers to set priorities, make changes, measure improvement and ensure sustainability. However, providers need to move beyond "practicing to the measures." This can be achieved by balancing small, focused initiatives and broader organizational initiatives.

Outcome measures need to be defined around clear, concise and measurable goals. This may require sharing data across settings, since actions in one setting may affect outcomes in another. This is particularly difficult in our current fragmented fee-for-service system, where providers may not see the ultimate outcomes resulting from their actions. This is at odds with the inclination to incorporate measures that providers have control of, since providers tend to have control of more narrow processes of care. More research is needed to create linkages and make providers accountable for their actions. Perhaps incentives need to be developed that drive improvement to larger systems of care rather than specific clinical processes.

### Networks and Physician Groups

It is important to keep in mind that some physicians work in group settings, and incentives that are implemented at the payer level may not always be seen at the individual provider level. However, physician groups can provide an infrastructure that may be helpful in driving change. For example, physician leadership within a group may drive quality improvement, and many examples of change involve a local change champion. Data within the group can be analyzed and relative rankings can be used as an educational tool to encourage change. Creative use of clinical, non-physician staff or other systems of care may also enhance quality.

### Provider Tools

In addition to using registries, EHRs and electronic prescribing systems to improve quality, providers can also use PDFs and other tools at the bedside to minimize clinical errors. Furthermore, the proliferation of guidelines will provide benchmarks for appropriate care that providers can use to guide their clinical decision making.

## Acknowledgements

*The authors would like to thank Bert White and his colleagues from the Rewarding Results National Evaluation Team at Boston University and John Pilotte from the Centers for Medicare and Medicaid Services (CMS) for their thoughtful input into this report.*

## Endnotes

1 Committee on Quality of Health Care in America, Institute of Medicine. Crossing the Quality Chasm: A New Health System for the 21st Century. 2001.
2 Young GJ, White B, Burgess JF, Jr., et al. "Conceptual Issues in the Design and Implementation of Pay-for-Quality Programs." *American Journal of Medical Quality*, 20(3): 144-150, 2005.
3 Hibbard JH et al. "Strategies for Reporting Health Plan Performance Information to Consumers: Evidence From Controlled Studies." *Health Services Research*, 32(2), April 2002.
4 Hibbard JH et al. "Development of the Patient Activation Measure (PAM): Conceptualizing and Measuring Activation in Patients and Consumers." *Health Services Research*, 39(4), August 2004.
5 Young et al.
6 Meterko M et al. "Provider Attitudes Toward Pay-for-Performance Programs: Development and Validation of a Measurement Instrument." *Health Services Research*, 41(5), October 2006.
7 Beaulieu ND and Horrigan DR. "Putting Smart Money to Work for Quality Improvement." *Health Services Research*, 40(5), Part I, October 2005.
8 Zierler BK et al. "Effect of Compensation Method on the Behavior of Primary Care Physicians in Managed Care Organizations: Evidence From Interviews With Physicians and Medical Leaders in Washington State." *American Journal of Managed Care*, 4(2), February 1998.
9 MedPAC Recommendations. *Health Affairs,* March/April 2006.
10 Mehrotra A, Bodenheimer T, Dudley RA. "Employers' Efforts to Measure and Improve Hospital Quality:

Determinants of Success." *Health Affairs* (Millwood), 22(2): 60-71, 2003.
11 Baker and Carter. 2005 Medvantage survey.
12 Baker and Carter. 2005 Medvantage survey.
13 Curtis J. "PULSE: Blue Cross of California's PPO Physician Pay for Performance Program." Presented at Rewarding Results: Lessons Learned and Implications for Medicare Pay for Performance, held by the Centers for Medicare and Medicaid Services on December 15, 2006 in Baltimore, MD.
14 Reschovsky JD et al. "Effects of Compensation Methods and Physician Group Structure on Physicians' Perceived Incentives to Alter Services to Patients." *Health Services Research*, 41(4), Part I, August 2006.
15 Beaulieu et al.
16 Rosenthal MB and Frank RG. "What Is the Empirical Basis for Paying for Quality in Health Care?" *Medical Care Research and Review*, 63(2), April 2006.
17 Ibid.
18 Ibid.
19 Hitt DD et al. "Effects of Delayed Rewards and Task Interest on Intrinsic Motivation." *Basic and Applied Social Psychology*, 13(4), 1992.
20 Peterson LA et al. "Does Pay for Performance Improve the Quality of Health Care?" *Annals of Internal Medicine*, 145(4), August 2006.
21 Evans DB et al. "Measuring Quality: From the System to the Provider." *International Journal for Quality in Health Care*, 13(6), 2001.
22 Thomas JW and Ward K. "Economic Profiling of Physician Specialists: Use of Outlier Treatment and Episode Attribution Rules." *Inquiry Journal*, 43, Fall 2006.
23 MedPAC. Report to Congress: Increasing the Value of Medicare. Chapter 1, June 2006. Also see: www.medpac.gov/publications/congressional_reports/Jun06_EntireReport.pdf.
24 Singer JA. "Rewarding Results: Lessons Learned and Implications for Medicare P4P, The Massachusetts Experience." Presented at Rewarding Results: Lessons Learned and Implications for Medicare Pay for Performance, held by the Centers for Medicare and Medicaid Services on December 15, 2006 in Baltimore, MD.
25 DiLorenzo J. "Bridges to Excellence." Presented at Rewarding Results: Lessons Learned and Implications for Medicare Pay for Performance, held by the Centers for Medicare and Medicaid Services on December 15, 2006 in Baltimore, MD.
26 Curtin K. "Rewarding Results: Reporting Lessons Learned to CMS – Findings: Excellus and RIPA." Presented at Rewarding Results: Lessons Learned and Implications for Medicare Pay for Performance, held by the Centers for Medicare and Medicaid Services on December 15, 2006 in Baltimore, MD.
27 Centers for Medicare and Medicaid Services. "DRA 5007 Medicare Hospital Gainsharing Demonstration Solicitation." September 2006. See also: www.cms.hhs.gov/DemoProjectsEvalRpts/downloads/DRA5007_Solicitation.pdf.
28 Wheeler J. "Blue Cross Blue Shield of Michigan: Participating Hospital Agreement Incentive Program." Presented at Rewarding Results: Lessons Learned and Implications for Medicare Pay for Performance, held by the Centers for Medicare and Medicaid Services on December 15, 2006 in Baltimore, MD.
29 NEJM study from February 1, 2007.
30 Colmers J. "Public Reporting and Transparency." The Commonwealth Fund, January 2007.
31 Werner R and Asch D. "The Unintended Consequences of Publicly Reporting Quality Information." *JAMA*, March 9, 2005.
32 Meredith MR and Dudley RA, "Pay for Performance: Will the Latest Payment Trend Improve Care?" *JAMA* 297(7): 740-744, February 21, 2007.
33 Fisher ES, Staiger DO, Bynum JP, Gottlieb DJ. "Creating Accountable Care Organizations: The Extended Hospital Medical Staff." Health Affairs, 26(1), January-February 2007.
34 Pham HH et al. "Care Patterns in Medicare and Their Implications for Pay for Performance." New England Journal of Medicine, 356(11), March 15, 2007.
35 Ibid.
36 Ibid.
37 Rosenthal MR and Dudley RA, "Pay-for-Performance: Will the Latest Payment Trend Improve Care?" JAMA 297(7):740-744, February 21, 2007.
38 Pear R. "AMA to Develop Measures of Quality Medical Care." New York Times, February 21, 2006.
39 Giuffrida and Torgerson. National Primary Care Research and Development Centre, University of York. "Should We Pay the Patient? Review of Financial Incentives to Enhance Patient Compliance." BMJ, September 1997; Jepson et al. NHS Centre for Reviews and Dissemination, University of York. "The Determinants of Screening Uptake and Interventions for Increasing Uptake: A Systematic Review." Health Technology Assessment NHS R&D HTA Program, 2000.
40 Wilensky G et al. "Gain Sharing: A Good Concept Getting a Bad Name?" Health Affairs, Web exclusive, December 5, 2006.
41 Coleman EA. Journal of the American Geriatrics Society, 51, 2003; Berenson and Horvath, Health Affairs, 2003.