WORKING
PAPER

Assessment of Pay-for-Performance Options for Medicare Physician Services: Final Report

MELONY E. SORBERO, CHERYL L.DAMBERG,
REBECCA SHAW, STEPHANIE S. TELEKI,
SUSAN L. LOVEJOY, ALISON H. DECRISTOFARO,
JACOB W. DEMBOSKY, CYNTHIA R. SCHUSTER

WR-391-ASPE

May 2006

Prepared for the Assistant Secretary for Planning and Evaluation, US Department of Health and Human Services

This product is part of the RAND Health working paper series. RAND working papers are intended to share researchers' latest findings and to solicit additional peer review. This paper has been peer reviewed but not edited. Unless otherwise indicated, working papers can be quoted and cited without permission of the author, provided the source is clearly referred to as a working paper. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.
RAND® is a registered trademark.

RAND HEALTH

SUMMARY

Pay-for-performance (P4P), the practice of paying health care providers differentially based on their quality performance, emerged in the late 1990s as a strategy for driving improvements in health care quality. The Centers for Medicare & Medicaid Services (CMS), the largest purchaser of health care services in the United States, is actively considering P4P for Medicare physician services, viewing this policy strategy as one way to increase physician responsibility for efficiently providing high quality, outcome focused care to beneficiaries of the Medicare program.

In September 2005, the Assistant Secretary for Planning and Evaluation (ASPE), within the U.S. Department of Health and Human Services, contracted with the RAND Corporation to help in its assessment of whether P4P can be effectively implemented in the Medicare physician service delivery and payment environment. RAND's tasks, within the scope of the project, were to

1. Review what is known about P4P and the empirical evidence about its effectiveness.
2. Describe the characteristics of current P4P programs
3. Assess whether features of these programs could help inform development of a P4P program for Medicare physician services.

This report summarizes what we, the RAND study team, did and what our findings were. Specifically, we describe

- Evidence from empirical studies about the effects of P4P.
- Lessons learned from currently operating P4P programs about how to design and implement these programs.
- Key P4P program design components and options.
- A framework for guiding the development of a P4P program.
- Challenges that CMS will face in designing and implementing a P4P program for Medicare physician services.
- Steps that CMS could take to prepare for building and supporting a national P4P program for physician services.

WHY IS MEDICARE INTERESTED IN PAY-FOR-PERFORMANCE?

Medicare's current interest in P4P is motivated by a range of concerns, especially continuing deficits in quality of care, rising health care costs, and the current Medicare fee schedule's inability to control costs.

The Quality Problem

A variety of studies document substantial deficiencies in the quality of care delivered in the United States (Asch et al., 2006; Institute of Medicine, 2001; Schuster et al., 1998; Wenger et al., 2003). A national examination of the quality of care delivered to adults found that they received only about 55 percent of recommended care on average and that adherence to clinically recommended care varied widely across medical conditions (McGlynn et al., 2003).

The Health Care Cost Problem

Health care costs continue to rise at a rapid pace; they are expected to account for nearly 19 percent of gross domestic product by 2014 (Heffler et al., 2005). In 2006, the Federal government will spend $600 billion for Medicare and Medicaid; by 2030, expenditures for these two programs are expected to consume 50 percent of the federal budget, jeopardizing funding for other, discretionary programs (McClellan, 2006). CMS Administrator Dr. Mark McClellan has stated publicly that if the United States is to continue funding these programs, it will need to redesign existing policies and practices.

The Current Medicare Payment Policy Problem

Approximately 484,000 physicians regularly bill for providing Medicare Part B services (MedPAC, 2006). Medicare's FFS payments for physician services follow a resource-based relative value fee schedule (RBRVS). The annual update to the fee schedule is determined by three factors: (1) the rate of change in the Medicare Economic Index (MEI), (2) a price index measuring changes in the costs of maintaining a physician practice, and (3) a sustainable growth rate (SGR) expenditure target. The annual update factor to the physician fee schedule is adjusted based on a comparison of cumulative past actual expenditures with the SGR.

Although the SGR was established as an expenditure control mechanism, the SGR target has been routinely exceeded because it is applied at the national level and treats all physicians the same regardless of their individual performance. Congress has protected physicians from negative updates resulting from expenditures exceeding the SGR targets

by the Medicare Modernization Act (MMA, 2003), which provided 1.5 percent updates for 2004 and 2005, and the Deficit Reduction Act (S. 1932), which provided 0 percent updates in 2006. Without modification, Medicare Part B expenditures will continue to exceed the SGR. Furthermore, the FFS payment system does not reward high quality care and often pays physicians more for treating complications that arise from poor quality of care.

Pay-for-Performance as a Means of Addressing These Problems

To close the gap between the care that is recommended and the care that patients receive, the Institute of Medicine (IOM) recommended reforms to the health system, one of which is the reform of current payment policies to create stronger incentives for providing high quality, efficient health care services (IOM, 2001). In response, a number of system reform experiments have been carried out in both the public and the private sector that offer financial and sometimes non-financial incentives to providers with the explicit goal of stimulating improvements in health care quality, provider accountability, and efficiency (Rosenthal et al., 2004; Epstein et al., 2004).

In 2005, the Medicare Payment Advisory Commission (MedPAC), which advises the U.S. Congress on issues related to Medicare, recommended that P4P be implemented for hospitals, home health agencies, and physicians (MedPAC, 2005). Congress has also shown interest in P4P, as evidenced through the multiple bills it has put forth. For example, the Medicare Value Purchasing Act (S.1356) proposed that a portion of Medicare reimbursement for physicians, hospitals, health plans, end-stage renal disease providers, and home health agencies be tied initially to the reporting of performance measures (either 2006 or 2007, depending on provider type) and then to actual performance (ranging from 2007 to 2009).

The Deficit Reduction Act of 2005(passed on February 8, 2006) does not include provisions for physician level P4P. It does, however, require MedPAC to submit a report by March 1, 2007, on mechanisms that could be used to replace the SGR system. Furthermore, the Deficit Reduction Act calls for hospital P4P to be implemented in fiscal year 2009, thereby setting the stage for future legislative activity to embed P4P in the reimbursement formula for Medicare physician services.

Important groundwork is being laid through a variety of CMS demonstrations. One of these, the Medicare Physician Group Practice (PGP) demonstration, is providing financial incentives to 10 physician group practices based on their quality and cost-efficiency performance. In addition, the Physician Voluntary Reporting Program (PVRP),

started in January 2006, will provide internal comparative performance feedback to providers but will not involve public reporting. CMS has started signaling its anticipated policy direction in public forums—first, by engaging in voluntary reporting of performance by physicians; then by moving to financially incentivize reporting; and then by implementing P4P (McClellan, 2005a; McClellan 2006). Finally, CMS is collaborating with the Ambulatory Care Quality Alliance in conducting a series of pilot projects around the country to test the feasibility of aggregating data across multiple payers and then scoring physicians and/or physician practices on a range of performance measures.

WHAT IS THE EMPIRICAL EVIDENCE FOR THE EFFECTIVENESS OF PAY-FOR-PERFORMANCE?

Neither the peer-reviewed literature on P4P programs nor ongoing evaluations of such programs provide a reliable basis for anticipating the effects of P4P in Medicare. Our examination of the peer-reviewed literature on P4P yielded 15 published studies whose goal was to determine the effect of directing financial incentives for health care quality at physicians, physician groups, and/or physician practice sites. All of these studies evaluated experiments that occurred in the late 1990s or in the early 2000s.

These studies do not, separately or in total, provide a clear picture of how P4P affects performance. The following is a breakdown of our findings:

- **The seven most rigorously designed studies (i.e., those using randomized controlled trials) provide an ambiguous message**: four show **mixed results** (Fairbrother et al., 1999, Fairbrother et al., 2001; Kouides et al., 1998; Roski et al., 2003), and three report **no effect** (Grady et al., 1997; Hillman et al., 1998; Hillman et al., 1999).
- **The two quasi-experimental studies report mixed findings** (Rosenthal et al. 2005; Levin-Scherz et al., 2006).
- **The least rigorously designed studies tend to report positive results** for at least one aspect of the programs examined (Francis et al., 2006; Greene et al., 2004; Amundson et al., 2003; Armour et al., 2004; Fairbrother et al., 1997; Morrow et al., 1995).

Drawing conclusions from the published literature about how P4P affects health care quality is problematic for a number of reasons:

- The interventions evaluated were small, and most were of very short duration, thus limiting the likelihood that an impact would be observed.

- The interventions typically occurred in one location with selected characteristics (e.g., targeting Medicaid providers), thus limiting the ability to generalize from the studies' findings.
- Many of the studies lacked control groups, thus making it difficult to distinguish the effects of P4P from the effects of other factors in the environment (e.g., medical group quality improvement interventions, public reporting of performance scores).
- The studies provide no information about the various design features that may have played a role in an intervention's success or failure, such as level of engagement and communication with providers and what share of a physician's practice the intervention represented (i.e., the dose effect).

In addition to the studies' methodological limitations, most of the programs evaluated in these studies do not resemble the P4P programs operating today in terms of size (i.e., number of measures or number of providers), duration, and magnitude of rewards. Thus, it is impossible to generalize from the findings in the published literature in order to estimate the effects of the newer generation of P4P programs. Some of these newer programs are being evaluated, but the results are just starting to emerge, and much of the new literature speaks only to lessons learned about the implementation process.

Furthermore, these new P4P programs are real-world experiments and, as such, suffer from some of the same methodological problems (i.e., lack of control groups, lack of random assignment to and not to incentives) as the studies evaluated in the peer-reviewed literature. These shortcomings will, of course, limit what the evaluations can reveal about how P4P affects performance. They do not, however, mean that these programs can offer no useful lessons for CMS.

WHAT CAN BE LEARNED FROM REAL-WORLD PAY-FOR-PERFORMANCE PROGRAMS?

As of December 2005, approximately 157 P4P programs were operating across the United States. These programs were sponsored by 130 organizations—including individual health plans, coalitions of health plans, employer groups, Medicare, and Medicaid—and they covered over 50 million health plan enrollees (Med-Vantage, 2005).

Because there is no published literature describing lessons learned in these programs, RAND held discussions with 20 private-sector P4P programs that target individual physicians or groups of physicians, as well as with six of the 10 medical groups participating in the CMS Physician Group Practice (PGP) P4P demonstration. The

PGP P4P demonstration is a three-year program to implement a P4P program for group practices with at least 200 doctors who care for FFS Medicare beneficiaries. The program's goal is to improve care for beneficiaries with chronic medical conditions by rewarding physician groups that manage patients across the continuum of care in a cost-effective, high quality manner. These discussions provided insights that could be useful to CMS as it embarks on a P4P program for Medicare physician services.

Some common themes emerged from our discussions with participants in the private-sector P4P programs:

- **P4P is not a panacea.** It is not, by itself, a solution for poor quality and rising costs. P4P needs to be implemented as part of a multi-dimensional set of strategies designed to change physician behavior so as to achieve quality and cost goals.
- **Physician involvement and engagement are critical to successful program implementation**. Sponsors of these programs found communicating with physicians to be a challenge, particularly in markets lacking sizable group practices or strong local physician leadership or organization. Traditional methods of communication, such as newsletters and mailings, were insufficient for raising awareness about the program and engaging physicians in quality improvement activities.
- **Health care remains local, and a one-size-fits-all approach may not work**. Discussions with P4P program participants revealed no consensus about the best way to design a P4P program (this lack of agreement is reflected in the design variations across existing programs). Variations occurred as a function of differences in the goals of individual P4P sponsors, the type of insurance product, and how physicians were organized within a geographic market.
- **It is essential to pilot test the implementation of measures and other implementation processes (e.g., audit, feedback reporting) at each step of the program**. Participants in programs that had not conducted pilot tests indicated that the omission was a serious mistake and strongly advised that all aspects of program design and implementation be tested. Two items repeatedly mentioned as being necessary were a willingness to be flexible and change, and the recognition that program development will involve some trial and error.
- **The accuracy and reliability of data underlying the measures must be ensured, and there must be a fair and equitable process for appeals.** These

two items were mentioned repeatedly as being essential for addressing providers' concerns about their performance scores and data accuracy.

- **Ongoing evaluation is needed.** Monitoring is needed to track programmatic effects and the process of implementation and to provide information that can be used to adjust program design and implementation.
- **Programs should start small, and success should be demonstrated.** These are seen as ways to build trust among program stakeholders.
- **Substantial infrastructure is required to support program operations.** Core operational functions needed are data warehousing; data aggregation; programming and analysis; data auditing; appeals management and data correction; performance feedback, such as report cards; communication with, engagement of, and support of physicians; measures maintainance; and payout computation and distribution. To support these operations, additional infrastructure investments in the form of both people and information technology will be needed, and sufficient resources must be allocated to support program operations.
- **Alignment of programs with the measures being used and physician requirements is vital.** This type of coordination is essential for reducing both confusion and the burden placed on providers who contract with multiple payers. Without it, providers may have to cope with inconsistent program requirements and measure specifications. Alignment among P4P sponsors within a market also strengthens the behavior-change signal to providers, increasing the likelihood that providers will in fact change their behavior.
- **Physicians need support for successful program participation**. Examples of the types of support being provided by programs are patient registries, technical support, and education.
- **To motivate physicians to change behavior, performance information must be actionable**. Providing rates is not sufficient; physicians must be able to act upon the information provided. For example, since most physicians in P4P programs continue to operate in an environment of paper records, they could be provided with specific lists of patients who need recommended care rather than being expected to start accessing population-based data and using it to track the provision of services and/or identify specific clinical areas in which there are less costly treatments that yield the same clinical outcomes as costly treatments for most patients (e.g., the use of ACE inhibitors instead of ARBs to treat

hypertension). Such lists would provide information in a manner that facilitates behavior change.

Early lessons that have emerged from the first year of the CMS PGP demonstration include the following:

- **Participation in the PGP demonstration was a key driver of performance improvement in the physician group organizations**. Four of the six physician groups noted that participating in the demonstration enabled them to implement changes (particularly to information systems) that had been discussed internally for years but never put into place. Once the demonstration was in place, changes began to happen or happened much more quickly than before.
- **Capital investments are required to support measurement and quality improvement work.** The physician groups told us they believe that the infrastructure investments needed to support P4P management and measurement will be "enormous." An influx of capital of this size suggests that vendors likely to fill infrastructure needs will have to be closely scrutinized.
- **Participation in P4P can prompt improved sharing of ideas to promote better care for the population.** Several physician groups mentioned their surprise that providers participating in the demonstration have embraced the concepts of population management underlying their case management strategies. The demonstration has improved the physicians' sharing of ideas for promoting better care within the organizations.
- **The support provided to PGPs was a critical feature of program design.** All physician groups mentioned at least one instance in which they contacted either CMS or its support services contractor, Research Triangle Institute (RTI), to comment on how a measure should be specified, to appeal the inclusion of beneficiaries in the group's target population under the attribution algorithm, to question the inclusion of patients in the denominator of measures, or to request help managing data.

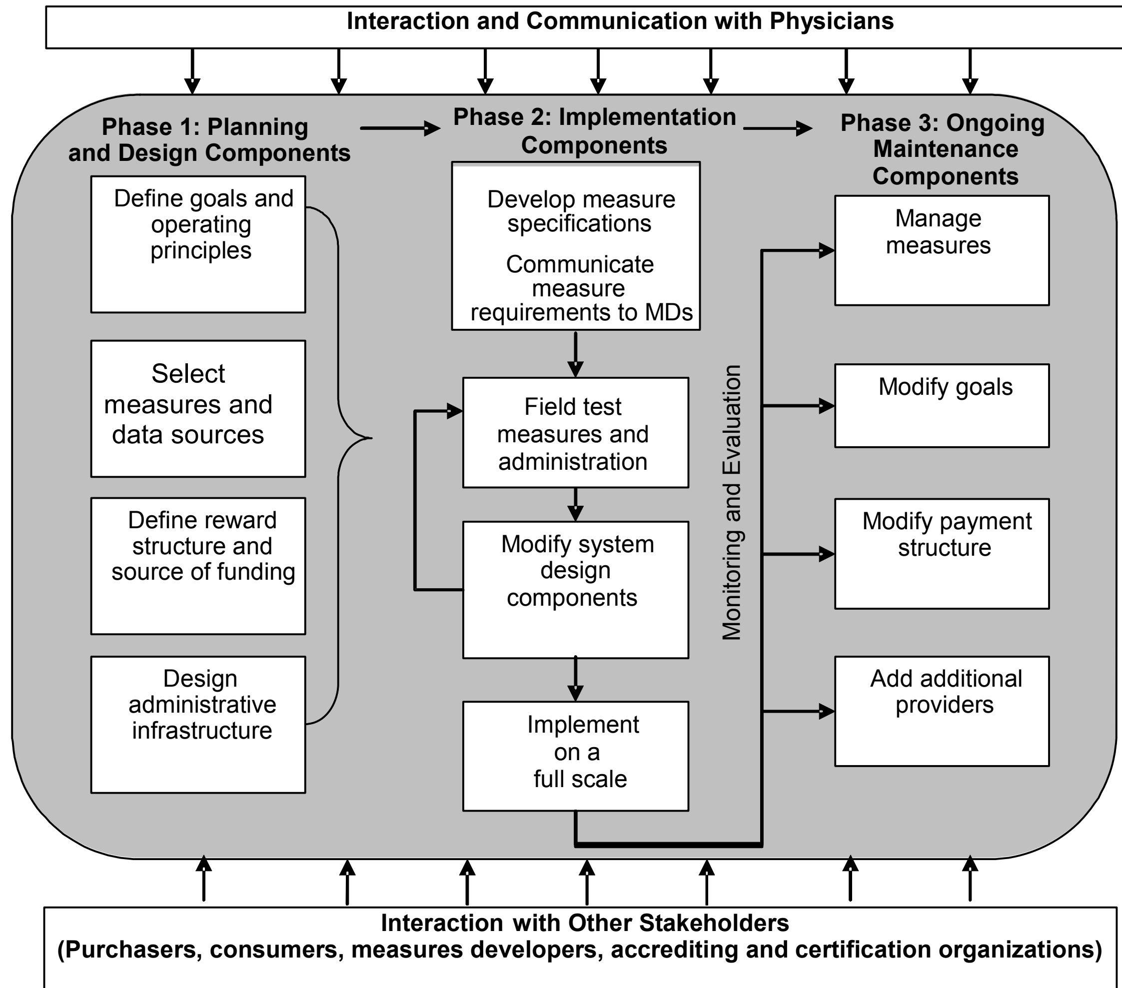
The organizations we spoke with, both in the private sector and in the CMS demonstration program, are firmly committed to P4P and believe that their programs, often in conjunction with other quality improvement activities, are resulting in care that is both of a better quality and more efficient. However, few of these programs are being evaluated in a rigorous manner, and only a handful have attempted to compute the return on investment (ROI) from their efforts.

WHAT DESIGN ISSUES AND OPTIONS NEED TO BE CONSIDERED?

Currently, there is no single strategy for designing and implementing a P4P program, so a great deal of experimentation and refinement is occurring as programs learn lessons along the implementation path. While all programs have key design components—such as attribution rules, payout structures, and measures selection—very little is known about the best form for these components or the relative importance of different components for achieving the program's goals. In some cases, newer programs are adopting the design components of more-mature programs, but there is substantial variation across P4P programs in terms of their approach to designing their programs. Programs are generally customized to address specific characteristics of the local health care market (e.g., organization of physicians, existence of physician leaders in the community), and little attention is paid to what theory suggests might be the best options of various program components to adopt. At this stage, absent empirical evidence to support one design approach over another, the variation in P4P experiments will allow opportunities for testing various design strategies.

Development of a P4P program is a complex undertaking involving many moving, interrelated parts. In addition, P4P programs are not static in terms of design. As a guide for those planning P4P programs, we used the steps identified in our discussions with P4P sponsors to construct a framework of core steps associated with developing and operating a P4P program. This framework, shown in Figure 1, displays the array of decisions that have to be made by any P4P program developer; it also highlights the interactions among various steps in the process.

Figure 1
A Framework to Guide Development of a Pay-for-Performance Program



Choosing among the various options typically reflects considerations of whether the approach helps to achieve programmatic objectives and what consequences may occur as a result. For example, if an explicit goal is to accelerate implementation of information technology (IT), the program developer may elect to include measures on the provider's IT capabilities. However, as was underscored in our discussions with P4P program developers, P4P program development is largely experimental in many respects, and the impact of various design components has not been studied and is not well understood.

Our review of the literature and discussions with a broad cross-section of existing P4P programs in the private sector revealed a host of options for the design components

that need to be addressed when developing a P4P program. The design issues that are of specific interest to ASPE and that we thus assessed are

- How should the initial performance areas that are to be subject to P4P be identified?
- What role should physicians and other stakeholders play in developing a P4P program for Medicare?
- What are appropriate measures for a P4P program?
- What unit of accountability should CMS measure and reward?
- Given the geographic variation in practice of care, should CMS pursue a national or a regional approach to implementation?
- How should patients be matched to individual physicians or group practices to ensure accuracy of measurement?
- How should rewards be structured?
- What should CMS be considering with regard to program infrastructure, including measure selection and specification, pilot testing, data collection and management, support to physicians, reporting and feedback, and monitoring?

We present information helpful in understanding the consequences or challenges associated with choosing particular design options. However, the effects of choosing one option over another are in many cases not known.

IS IT POSSIBLE TO IMPLEMENT A PAY-FOR-PERFORMANCE PROGRAM FOR MEDICARE PHYSICIAN SERVICES?

A P4P program for Medicare physician services can be implemented. However, in designing and implementing a P4P program, CMS will face significant challenges that include

- **The absence of an existing organizational infrastructure within CMS with which to manage the myriad components associated with running a P4P program.** This is particularly the case given a program of the size and scope necessary to measure and reward all or most physicians in the Medicare Part B program. To support a P4P program's operations, many systems will have to be designed, built, tested, and maintained, an endeavor that will require dedicated and sustained resources.
- **The absence of a P4P program comparable in size and scope to a P4P program for Medicare physician services**. There is no P4P program of comparable size from which to draw lessons.

- **The absence of infrastructure (personnel and information systems) at the individual doctor-level to support a P4P program's requirements**. For example, the majority of physician offices have neither electronic health records nor sufficient staff to perform the chart abstractions that might be required to provide information needed to construct the performance measure.
- **The difficulty of communicating with and engaging individual physicians in the program to achieve the desired behavior changes.** Organized medical groups have the staff and structure to facilitate communication between a P4P program sponsor and front-line physicians. At the level of the individual physician, however, there is no "local physician leadership" or point person who can help to facilitate communication with physicians about the program, engagement of physicians in the program, and assistance with behavior change.
- **The shortness of the timetable for ramping up a national operation.** Given Congress's mounting pressure for action, CMS is unlikely to have time to pilot a P4P program in multiple sites. There likely will be pressure to roll out a national program in a short period of time.
- **Physician resistance to transparency (public reporting) of performance data.** Some people have asserted that public transparency and accountability are valuable additions to P4P programs because they drive behavior change among physicians. However, physicians have expressed concerns about public reporting of performance results, especially about problems with data inaccuracies and failure to account for differences in patient populations served.

TAKING THE FIRST STEPS TO IMPLEMENT A P4P PROGRAM FOR MEDICARE PHYSICIAN SERVICES

There are several steps that CMS could undertake immediately and in the near term, as well as in the longer term, to prepare itself for designing and implementing a P4P program for Medicare physician services. The actions presented here, if taken, would provide information to guide program planning, would help generate awareness and engagement among physicians, and would begin to build the program infrastructure needed to support a P4P program.

Near-Term Steps (6 to 18 months)

Model critical design components using existing data.

CMS could start laying the groundwork for structuring a P4P program by modeling various program design components using existing Medicare claims data. Some of the critical design issues to be addressed in modeling the components are (1) the implications of different attribution rules, (2) the number of measures that can be scored today using claims data, (3) the number of physicians whose performance can be reliably scored using measures based on administrative data, and (4) the increase in the number of physicians that CMS could score if scores were based on composite measures versus individual indicators of performance.

Monitor the experiences of the Physician Voluntary Reporting Program and consider how to address emerging lessons in the design of P4P for Medicare physician services.

Implementation of the PVRP, a program started in January 2006 that will provide internal comparative performance feedback to providers on a starter set of 16 measures, offers CMS a potential foundation on which to build a P4P program. The lessons being learned in the PVRP will provide CMS with valuable information; in particular, the monitoring of physician participation and growth in participation over time will provide indications about the readiness of physicians nationally to provide information on the selected measures. Interviews with physicians could give CMS valuable insights about why physicians did or did not agree to participate. Participating providers could describe the challenges they experience with the data collection and reporting process, as well as their reactions to performance feedback reports. Non-participating providers could help to identify barriers to participation and actions needed to address them. Information gained from physician interviews could be useful in determining how to modify the program going forward as a stepping stone to full P4P. The interviews also would allow CMS to build communication channels with physicians before a P4P program is implemented and would constitute an important step in soliciting physician input on program design.

Mid-Term Steps (18 to 36 months)

Create incentives for participation in the PVRP as a way to help physicians move toward understanding performance measurement, to build systems to support measurement, and to work toward performance transparency.

Low participation in PVRP may suggest the need to provide inducements for participation, such as pay-for-reporting. Participation in PVRP is important in that it offers physicians the opportunity to gain experience with submitting data and receiving performance feedback, well in advance of P4P. PVRP participation also allows physicians time to see performance scores in a confidential manner, giving them the opportunity to improve systems for data capture and to identify and correct quality problems in advance of public reporting. This is an important step for CMS to take on the path to public transparency.

Expand the PVRP measurement set and administrative collection of measures.

CMS could also continue to expand the PVRP 16-measure set so that it is consistent with P4P program design decisions about which measures to reward to drive improvements. In addition, to support the administrative reporting of data to produce performance measures, particular attention should be paid to modifying the HCFA 1500, the form physicians use to submit claims to Medicare, to capture administratively the data elements needed to support performance measurement (e.g., working with the AMA to develop Current Procedural Terminology [CPT] supplemental codes).

Plan for program evaluation and collect baseline data.

It is also very important for CMS to build into its P4P design the continuous evaluation of program implementation and effects. Ongoing evaluation will give CMS critical information that it can use to adjust the program. Assessment of program effects will require that CMS collect baseline information about performance. If CMS expects to compute the return on investment, it will need to track program costs.

Longer-Term step (36 months and beyond)

Scale up incrementally and continue to build infrastructure capacity.

As the PVRP matures, CMS could scale up the program incrementally by adding measures and physician specialties and continuing to build infrastructure to accommodate

the program’s increasing size. By building gradually on successes, CMS will help to build trust within the provider community and will gain experience along the way.