

WORKING P A P E R

An Environmental Scan of Pay for Performance in the Hospital Setting: Final Report

CHERYL L. DAMBERG, MELONY SORBERO,
ATEEV MEHROTRA, STEPHANIE TELEKI, SUSAN
LOVEJOY, AND LILY BRADLEY

WR-474-ASPE/CMS

November 2007

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

PREFACE

In recent years, pay-for-performance (P4P) programs have emerged as a strategy for driving improvements in the quality, safety, and efficiency of delivered health care. In 2005, with passage of the Deficit Reduction Act, Congress mandated that the Secretary of the Department of Health and Human Services (DHHS) develop a plan for value-based purchasing (VBP) for Medicare hospital services. VBP is one strategy for modifying the payment system to incentivize improvements in the quality of care delivered to beneficiaries in the Medicare program. The use of incentives—by paying differentially for performance—is a key component of building a value-driven health care system as called for by the DHHS Secretary’s Four Cornerstones Initiative.

To inform the development of the VBP plan for Medicare hospital services, the Assistant Secretary for Planning and Evaluation (ASPE), in collaboration with the Centers for Medicare & Medicaid Services, contracted with the RAND Corporation to conduct an environmental scan of the hospital P4P landscape. This report presents the results from the environmental scan of P4P and pay-for-reporting (P4R) programs; it also includes a review of the empirical evidence about the impact of these programs, a description of program design features, and a summary of lessons learned from currently operating P4P and P4R programs about the structure of these programs and implementation issues.

This work was sponsored by ASPE under Task Order No. HHSP233200600001T, Contract No. 100-03-0019, for which Susan Bogasky served as the Project Officer.

CONTENTS

PREFACE	iii
TABLES	vii
SUMMARY	ix
ACKNOWLEDGEMENTS	xxi
ABBREVIATIONS	xxiii
I. INTRODUCTION	1
Background	1
Development of the Value-Based Purchasing Plan	4
Content and Structure of This Report	9
II. A REVIEW OF THE EVIDENCE ON HOSPITAL PAY FOR PERFORMANCE	11
Summary of the Empirical Evidence on the Impact of Hospital Pay for Performance	11
Theoretical Literature and implications for p4p design	29
Limitations in using Economic Theories to Predict Behavioral response	35
Conclusions	38
III. SUMMARY OF DISCUSSIONS WITH PAY-FOR-PERFORMANCE PROGRAM SPONSORS	41
Methodological Approach	41
Findings From Discussions with Program Sponsors	43
Critical Lessons Learned	57
IV. SUMMARY OF DISCUSSIONS WITH HOSPITALS, HOSPITAL ASSOCIATIONS, AND DATA VENDORS	63
Methodology	63
V. SUMMARY OF FINDINGS FROM ENVIRONMENTAL SCAN	77
APPENDIX A: DESIGN ISSUES EXPLORED AS PART OF THE ENVIRONMENTAL SCAN	83
APPENDIX B: SUMMARY OF PAY-FOR-PERFORMANCE DESIGN PRINCIPLES	89
APPENDIX C: INPATIENT HOSPITAL MEASURES	121
APPENDIX D: LIST OF ORGANIZATIONS PARTICIPATING IN THE ENVIRONMENTAL SCAN	149
REFERENCES	151

TABLES

Table 1: Design Issues Explored with Program Sponsors and Hospitals	6
Table 2: Key Terms Used to Search the Literature for Hospital P4P Studies	12
Table 3: Summary of Design Features of P4P Programs Contained in Published Evaluation Studies	14
Table 4: Summary of Evaluation Studies Examining Hospital P4P Programs	15
Table B.1. P4P Principles and Recommendations from Stakeholders	89
Table B.2. Summary of P4P Design Principles and Recommendations.....	118

SUMMARY

Mounting cost pressures and substantial deficits in the quality of care within the U.S. health care system have led policy makers to consider various reform options. Pay for performance (P4P) has emerged as a leading reform strategy, in an effort to stimulate improvements in the quality, safety, and efficiency of delivered health care (IOM, 2006). In 2005, Congress passed the Deficit Reduction Act (DRA, Public Law 109-171, Section 5001(b)), which mandated that the Secretary of the Department of Health and Human Services (DHHS) develop a plan for value-based purchasing (VBP) for Medicare hospital services that would commence in Fiscal Year (FY) 2009. VBP, which is being applied by payers in both the public and private sectors, includes the use of both financial (e.g., P4P) and non-financial (e.g., transparency of performance scores) incentives to change the behavior of providers and the systems within which they work.

The use of incentives—by paying differentially for performance—and measuring and making quality information transparent are key components of building a value-driven health care system, as called for by the DHHS Secretary Leavitt’s Four Cornerstones Initiative. In support of this initiative, CMS has taken a number of steps toward using incentives and making quality information transparent, by funding pay-for-performance demonstrations in the hospital, physician, and home health settings, and by implementing pay for reporting (P4R) for hospitals, through the Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU) program, and for physicians through the Physician Quality Reporting Initiative (PQRI).

AN ENVIRONMENTAL SCAN OF HOSPITAL PAY FOR PERFORMANCE

The DRA required the Secretary of the DHHS to consider the following design elements when developing the VBP plan: (1) the process for developing, selecting, and modifying measures of quality and efficiency; (2) the reporting, collection, and validation of quality data; (3) the structure, size, and source of value-based payment adjustments; and (4) the disclosure of information on hospital performance. The CMS Hospital VBP

Workgroup was delegated the task of developing the VBP plan for Medicare hospital services.

To inform the development of the VBP plan the Assistant Secretary for Planning and Evaluation (ASPE) and CMS issued a contract to the RAND Corporation to conduct an environmental scan of the hospital P4P landscape. The environmental scan, conducted between August of 2006 and June of 2007, included:

1. **A review of the literature** to assess what is known about the impact of P4P and how various design features influence the effectiveness of these interventions. The review examined the hospital inpatient and outpatient P4P empirical literature as well as theoretical literature drawn from the economics and management disciplines regarding the use of incentives and behavioral responses;
2. **Discussions with key informants** to provide a picture of the current state-of-the-art in hospital pay for performance program design and to draw upon the experiences and lessons learned from existing P4P and P4R initiatives; and.
3. **A synthesis of the findings** from the environmental scan to inform the discussions and design considerations of the CMS VBP Workgroup.

To take advantage of the experimentation going on nationally with respect to P4P program design and implementation, discussions were held with 27 program sponsors, 28 hospitals, 7 hospital associations, 5 data support vendors, and a number of individuals with expertise in rural hospital issues. The discussions were necessary because this type of descriptive information and this level of detail about program design are not typically contained in peer-reviewed journal articles that summarize the results of P4P interventions. Additionally, many of the demonstration experiments are still in their infancy, and little has been formally documented about the related experiences. This report summarizes the findings from the environmental scan.

FINDINGS FROM THE LITERATURE REVIEW

The Empirical Literature on Hospital P4P

As of June 2007, few peer-reviewed studies existed on the use of financial incentives and their impact on quality, patient experience, safety, or the efficient use of resources. While more than 40 hospital-based P4P programs are operating in the U.S., little empirical evidence has emerged from these payment reform experiments to gauge the impact of hospital P4P in meeting programmatic goals or to understand how various design features affect such things as engagement in the program, the likelihood of creating unintended consequences (such as reductions in access to care for more difficult patients), or the distribution of payments to providers. Few P4P programs are undergoing formal evaluations to assess their impact, and challenges arise in conducting evaluations of real-world applications because the applications generally lack a comparison group that is required to assess the impact of the P4P intervention.

We reviewed the literature between January 1996 and June 2007 and found only nine published studies that address the impact of three separate hospital P4P programs in which formal evaluations have been occurring:

1. The Hawaii Medical Service Association (HMSA) P4P program
2. The Blue Cross Blue Shield (BCBS) of Michigan Hospital Incentive Program
3. The Premier Hospital Quality Incentive Demonstration (PHQID).

Of the eight studies examining changes in performance, each one reported improvements over time in at least some of the hospital performance measures or condition-specific composites included in the specific study; however, it is difficult to disentangle the P4P effect from the effect of other quality improvement efforts that were occurring simultaneously. The strongest evidence on the impact of hospital P4P to date has been shown through the Lindenauer (2007) study of the impact of PHQID relative to the Medicare RHQDAPU program. These studies, while showing a positive effect of P4P, reveal that the additional effects of P4P are somewhat modest relative to public reporting and other quality interventions that are occurring simultaneously.

Improvements in hospital performance have been observed in response to feedback reports (Williams et al., 2005) and public reporting, with a financial incentive for submitting data (Grossbart, 2006; Lindenauer et al., 2007). One study found improvements in a few performance areas associated with P4P as compared with what was seen for control hospitals participating in voluntary quality improvement activities (Glickman et al., 2007). It has been argued, however, that in order to accomplish sustained quality improvement, interventions should be multifaceted and focus on different levels of the health care system (Grol et al 2002; Grol and Grimshaw 2003). This suggests that to be most effective, P4P should be partnered with other activities such as public reporting and internal quality improvement activities, that also encourage quality improvement for the same clinical area.

There is less evidence of the effect of P4P on patient outcomes. One study (Berthiaume et al., 2006) found reduced complication rates for obstetrical and surgical patients in an uncontrolled study, though it was not reported whether those improvements were statistically significant. Glickman et al. (2007) did not find significant differences in inpatient mortality improvement for AMI between PHQID and control hospitals exposed to an AMI quality improvement intervention.. None of the studies evaluating PHQID separately analyzed the other patient outcome measures (for coronary bypass surgery and hip and knee replacement surgery) included in the program, so it is not clear whether improvements occurred in these measures.

Most of the published studies have significant methodological limitations. Six of the nine had no controls, which are critical for providing evidence of a link between P4P and performance improvements. This is particularly important given the documented temporal trend toward increasing performance on many hospital quality metrics. Another important issue to consider is whether the experience of these smaller-scale incentive programs, with the exception of the PHQID, could be generalized to reflect what the effects would be of wholesale national implementation of a hospital P4P program by Medicare.

Theoretical Literature and Implications for P4P Design

P4P is common in industries other than health care, and economists and management experts have studied and developed theories on how individuals respond to financial incentives. The economic and management theories that we reviewed suggest that the way in which P4P incentives are structured, or framed, could influence whether they achieve the desired behavioral response. Among the key highlights of this literature review:

- **Withholds May Have More of an Impact Than Bonuses** (Prospect Theory, Principle of Loss Aversion)—Individuals are more sensitive to incentives when they perceive they are losing as opposed to gaining something. The difference in the behavioral response for a choice framed as a loss rather than as a gain can be significant, almost twofold in magnitude (Kahneman and Tversky, 1979). P4P incentive payments can be structured as a withhold (a perceived loss in income) or as a bonus (a perceived gain). The theory of loss aversion suggests that if the goal is to drive hospitals to make changes that improve quality or efficiency, withholding dollars with the likelihood of later releasing them based on performance (i.e., framing the incentive as a possible loss) may lead to a greater behavioral response than framing the incentive as a “gain,” in the form of a bonus, even if the same amount of money is at risk.
- **A Series of Small Incentives Might Lead to More Quality Improvement Than Would One Large Incentive** (Principle of Diminishing Marginal Utility)—The perceived value of a sum of money becomes progressively lower when associated with an increasingly larger sum of money. People tend to judge such gains or losses as changes from their current state of well-being (or reference point), rather than their final states (Kahneman and Tversky, 1979). Thus, it may be more psychologically motivating to provide smaller, more-frequent incentive payments to providers than to provide a larger, lump-sum incentive payment.
- **Uncertainty May Reduce the Behavioral Response** (Principle of Risk Aversion)—Most people are risk averse; and when given a choice they will

choose an option with 100 percent certainty over an option involving an uncertain but likely more valuable outcome. This principle suggests that decreasing the risk or uncertainty in the likelihood of receiving a financial incentive is likely to lead to a greater behavioral response to the incentive. Relative thresholds based on provider rankings, found in many P4P program designs, create greater uncertainty for hospitals than do payment schemes that use absolute thresholds (i.e., a fixed target) for determining who receives an incentive payment. This is because the level of performance necessary to earn the incentive is unknown until after the performance period has ended.

- **Reducing the Time Lags Between Performance and Receipt of Incentive Can Help to Achieve Maximum Response** (Principle of Hyperbolic Discounting)—Individuals value having a sum of money now more than sometime in the future, even after accounting for inflation. Instead of discounting in a linear fashion, the individuals tend to discount at a steeper, hyperbolic curve. In the context of P4P program design, minimizing the lag time between the performance being incentivized and receipt of the incentive may strengthen the behavioral response. Substantial time lags between data collection and payouts may cause a hospital to see the incentive as occurring so far in the future that it is not worth pursuing.
- **A Series of Tiered Absolute Thresholds May Be Better Than One Absolute Threshold** (Goal Gradient)—An individual's motivation and effort when faced with a goal greatly depends on that individual's baseline performance. If baseline performance is far away from goal performance, the individual exerts little effort, because the goal is viewed as not immediately attainable. As baseline performance gets closer and closer to goal performance, the individual exerts more and more effort to succeed. However, as soon as the goal is achieved, the motivation to improve decreases significantly. Applied to P4P, this principle implies that there would be a greater behavioral response among hospitals if there were a series of quality performance thresholds to meet (e.g., increasing dollar amounts for achieving a 50 percent, a 60 percent, a 70

percent, an 80 percent, and a 90 percent performance threshold) rather than one (e.g., a 75 percent performance threshold). Another way to structure multiple thresholds is by paying for improvement, so that instead of thresholds there is a continuous scale across which performance and payments can be achieved.

- **Multidimensional Output or Multitasking**—Multitasking refers to situations in which the responsibilities of an individual encompass multiple activities or outputs that may require different types of skills to accomplish (Holmstrom and Milgrom, 1991). A hospital's output includes many different components, such as managing a patient's chronic illness, the timely and efficient diagnosis of a patient's new symptom, transitioning patients from the hospital to outpatient care, and providing emotional support to patients and their families.

Multitasking is relevant to P4P programs because the performance measures in these programs typically address only a narrow piece of a hospital's outputs or the processes that contribute to outputs. It is hypothesized that if a large incentive is applied to one type of output, other outputs will be neglected, and overall care might worsen (Holmstrom and Milgrom, 1991). Such concern is thought to explain why few private-sector corporations put large fractions of employee pay "at risk," making them dependent on measures of output for which only a small fraction of what contributes to output can be measured (Asch and Warner, 1996). A broader set of measures within a P4P program that includes process of care for a variety of clinical conditions, outcomes, patient experience, and efficiency could serve to mitigate this concern.

- **Intrinsic versus Extrinsic Motivation**—Intrinsic motivation is a person's inherent desire to do a task, while extrinsic motivation is the external incentive (such as P4P). Instead of supporting intrinsic motivation, extrinsic incentive "crowds out" intrinsic motivation, because when a task is tied to an extrinsic incentive, people infer that the task is difficult or unpleasant (Freedman, Cunningham, and Krismer, 1992). Increasing the size of the financial incentive is one way to address the crowding out of intrinsic motivation, though very

large incentives run the risk of having the hospital overly focus on measured areas of care to the detriment of unmeasured areas of care.

FINDINGS FROM THE KEY INFORMANT DISCUSSIONS

Design Lessons

Discussions with program sponsors, hospitals, and data vendors revealed the following lessons about P4P program design and operation:

- **Measures**—Hospitals are using an array of performance measures, though the focus at this stage is primarily on measures of clinical effectiveness, and within this category, most of the focus is on measures of underuse (i.e., process-of-care). Little is happening with respect to measuring efficiency, clinical outcomes, or patient safety. Sponsors noted there were limitations in the number and type of measures currently available for use in pay for performance and public reporting, and cited a need for additional measure development and testing. Hospitals expressed concerns about growing data collection and reporting burdens across the various P4P programs and reporting initiatives being developed by an array of sponsors, whose efforts are not fully aligned. Hospitals expressed a strong desire for measures to be aligned, for reporting efforts to be coordinated, and for use of evidence-based standardized measures to minimize physician pushback. While P4P program sponsors desire to expand the number and types of performance measures to ensure a more comprehensive picture of hospital quality, hospitals stated a desire for a more limited set of measures on which they could focus quality improvement efforts.
- **Payment structures**—Existing P4P programs primarily make reward payments on the basis of improving over time or relative performance. Hospitals universally agreed that payment structures should use absolute thresholds and reward all good performers, rather than providing incentives on a relative-performance basis (such as paying only to the top 10 or 20 percent of hospitals participating in a P4P program). This was seen as critical when the measures of

performance used have scores that “top out,” reflecting little meaningful difference in the performance across most hospitals. Program sponsors felt strongly that performance improvement as well as attainment of specific benchmarks should be included as a component of the payment structure, at least in the early years of a P4P program, in order to engage all hospitals. Hospitals also noted the difficulty of getting physicians to change their behavior absent aligned incentives on the physician side, and called for program sponsors to create parallel physician incentives focused on inpatient care for the same conditions used in hospital programs.

- **Data infrastructure**—Current validation efforts are weak, and program sponsors and hospitals acknowledged the need to strengthen validation as more money is put at risk in P4P programs. Hospitals also indicated a need for technical support to comply with P4P program requirements, and cited the important role played by QIOs and data vendors in helping them understand the program requirements, prepare data submissions, and develop tools and interventions to improve performance. Current information systems hamper the ability of P4P programs to substantially expand their measure sets because hospitals still rely on manual abstraction of hard copy medical records to produce the data required for P4P programs. Hospitals also expressed a desire that the P4P program data infrastructure be constructed in a way that enables regular, timely feedback to hospitals on their performance, for the purposes of making corrections and for quality improvement work.
- **Public reporting**—Hospitals indicate they do pay attention to how their institution looks publicly and that public reporting has forced their boards to more closely monitor quality and provide resources for quality improvement. Both program sponsors and hospitals cited a need for simplification of the performance information presented on consumer websites, such as the CMS Hospital Compare website, to facilitate consumer understanding and use of the information.
- **Engagement strategies**—Program sponsors noted the importance of engaging hospitals in the planning and execution of P4P programs to encourage a more

collaborative versus payer-driven approach to implementing this payment reform. Engagement strategies included involving providers in the measures selection process and program design more broadly, in ongoing planning as the program evolves over time, and structuring aligned incentives on the physician-side, as noted above.

- **Absence of Knowing What Works**—Because P4P is a newly emerging reform tool and little information is currently available about the impact of P4P or the influence of various design structures on P4P outcomes, P4P programs should incorporate evaluation and ongoing monitoring into their design as a means of building a knowledge base. Hospitals and P4P program sponsors recommended allowing experimentation, which would create models where learning could occur to inform future design structures. The discussants noted that the results of P4P may differ as a function of the program design features as well as the varying structure of local health care markets, and that much could be gained from examining the experience of these local experiments. Collecting and broadly disseminating this type of information will be critical to future efforts to construct P4P programs so that they can meet their programmatic objectives. Funding will be necessary to support program evaluation, and the evaluation work needs to be sustained over multiple years to fully assess impact and monitor for unintended consequences.

Program Implementation Challenges

The environmental scan also uncovered a number of program implementation challenges that warrant consideration during program design and implementation.

The small numbers problem: A sizeable number of hospitals have only a small number of events or cases to report for one or more measures. A small number of events to score will result in unstable estimates of performance as a basis for determining performance-based incentive payments. While this is a more acute problem for small and rural hospitals with a small number of patients per year, the problem also occurs in some medium- and large-size hospitals depending on their service mix, the details of measure

specifications, and the use of sampling during data collection. Using all-payer data, collecting and aggregating data over longer periods of time, using composite measures,¹ and identifying measures relevant to smaller providers are approaches that can help to mitigate the small numbers problem and allow for the construction of more stable estimates of performance.

The Burden of Data Collection: The data collection burden, which affects how many measures a P4P program can reasonably require a hospital to collect and report, creates challenges for efforts to comprehensively assess the performance of hospitals given the wide range of care and services provided within hospitals. The more comprehensive the measure set used, the greater the burden on hospitals in the near term, given that most of the data needed to construct performance measures is contained in paper medical records. In most cases, hospital information systems are not yet equipped to capture and easily retrieve the clinical information used to create performance measures, nor are they structured to enable routine monitoring of quality of care. Until health information systems are upgraded to capture this information, program sponsors may be constrained in the number and breadth of measures they can expect hospitals to collect and report. Once effective information systems are built and put into place, the number of measures included in a P4P program could be expanded.

Ensuring the Validity of Data used to Make Differential Payments: P4P programs are also challenged with an acute need to ensure the integrity of the data used to score hospitals and make differential payments, which requires resources for data validation. Allocating sufficient resources to validation work is critical for program credibility, and today only limited resources are being used for data validation within P4P programs. Most hospitals stated that the current level of validation is insufficient, and the incentives to game the system will increase as the amount of money at risk in P4P programs increases.

¹ There are a variety of ways to construct composite measures, not all of which would help mitigate the small numbers problem.

In summary, P4P programs have the potential to drive system improvements but their impact is likely influenced not only by their design but also by what other structures are in place to support P4P—such as enhanced information systems for quality monitoring and feedback, aligned payments across all providers, and transparency. The success of these programs in meeting improvement goals likely will be affected by their design, how they are implemented, and whether sufficient resources are allocated to provide the necessary day-to-day support for program operations and ongoing modification of the program.

Hospitals understand that P4P is likely to be part of their future and generally seem supportive of the concept. They face a number of challenges to their ability to successfully participate in these programs, including lack of physician engagement, inadequate information infrastructure that necessitates the manual collection of data from charts, and potentially conflicting signals from various organizations measuring hospital performance. These implementation challenges are important to consider carefully in the design of any hospital P4P program.

ACKNOWLEDGEMENTS

We gratefully acknowledge the sponsors of the pay-for-performance programs and the hospitals, hospital associations, and data vendors whose people willingly made the time to participate in individual discussions with us. They offered us valuable information and insights about their experiences in designing and implementing pay-for-reporting and pay-for performance programs.

We also extend our appreciation to the members of our Technical Expert Panel—Dr. Elliott Fisher of Dartmouth University, Dr. Jack Wheeler of the University of Michigan School of Public Health, Dr. Dale W. Bratzler of the Oklahoma Foundation for Medical Quality, and Dr. Howard Beckman of the Rochester Individual Practice Association—for their thoughtful review of the discussion guides to help ensure that pertinent topics and issues were addressed and their review of this report. In addition, we appreciate the assistance provided by Geoff Baker of Med-Vantage in helping us construct and narrow the list of candidate hospital pay-for-performance programs with which we held discussions. Finally, we thank Susan Bogasky, from the Assistant Secretary for Planning and Evaluation, who served as Project Officer for this contract. We also appreciate the guidance and feedback provided by Dr. Julie Howell, Project Coordinator Hospital VBP, CMS Special Program Office for Value-Based Purchasing, and Dr. Thomas Valuck, Director, CMS Special Program Office for Value-Based Purchasing.

ABBREVIATIONS

Abbreviation	Definition
AAFP	American Academy of Family Physicians
AAMC	Association of American Medical Colleges
ACC	American College of Cardiology
ACR	American College of Radiology
AHA	American Hospital Association
AHIP	Association of Health Insurance Plans
AHRQ	Agency for Healthcare Research and Quality
AMA	American Medical Association
AMGA	American Medical Group Association
AMI	Acute Myocardial Infarction
APU	annual payment update
ASO	administrative services only
ASPE	Assistant Secretary for Planning and Evaluation
AVR	Aortic Valve Replacement
BCBS	Blue Cross Blue Shield
BCBSA	Blue Cross/Blue Shield Association
CABG	Coronary Artery Bypass Graft
CAH	Critical Access Hospital
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CAP	Community Acquired Pneumonia
CART	Clinical Abstracting and Reporting Tool
CDC	Centers for Disease Control
CHA	Catholic Health Association
CHF	Congestive Heart Failure
CMS	Centers for Medicare & Medicaid Services
CPOE	computerized physician order entry
CQI	continuous quality improvement
CRUSADE	Can Rapid Risk Stratification of Unstable Angina Patients Suppress Adverse Outcomes with Early Implementation of the ACC/AHA Guidelines
DHHS	Department of Health and Human Services
DRA	2005 Deficit Reduction Act
DRG	diagnosis-related group
EHR	electronic health record
EMR	electronic medical record
FAH	Federation of American Hospitals
FFS	fee for service
FY	fiscal year

Abbreviation	Definition
AAFP	American Academy of Family Physicians
GWTG	Get with the Guidelines
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
HIT	health information technology
HMO	health maintenance organization
HMSA	Hawaii Medical Service Association
HQA	Hospital Quality Alliance
ICU	intensive care unit
IHA	Integrated Healthcare Association
IHI	Institute for Healthcare Improvement
IOM	Institute of Medicine
IT	information technology
<i>JAMA</i>	<i>Journal of the American Medical Association</i>
JC	Joint Commission
MedPAC	Medicare Payment Advisory Commission
MGMA	Medical Group Management Association
NACH	National Association of Children’s Hospitals
<i>NEJM</i>	<i>New England Journal of Medicine</i>
NQF	National Quality Forum
NRHA	National Rural Health Association
NSQIP	National Surgical Quality Improvement Program
NVHRI	National Voluntary Hospital Reporting Initiative
ORYX	A quality improvement initiative introduced by JC that utilizes performance measures
P4P	pay for performance
P4R	pay for reporting
PGP	Physician Group Practice
PHQID	CMS–Premier Hospital Quality Incentive Demonstration
POS	point of service
PPO	preferred provider organization
PPS	Prospective Payment System
PQRI	Physician Quality Reporting Initiative
QALY	quality adjusted life years
QIO	Quality Improvement Organization
RHQDAPU	Reporting Hospital Quality Data for Annual Payment Update
ROI	return on investment
SCIP	Surgical Care Improvement Project
STS	Society of Thoracic Surgeons
TEP	Technical Expert Panel

Abbreviation	Definition
AAFP	American Academy of Family Physicians
VBP	value-based purchasing
VHA	Voluntary Hospital Association
VTE	Venous Thromboembolism

I. INTRODUCTION

BACKGROUND

The Cost and Quality Problems

Substantial, well-documented deficiencies exist in the quality of care that is provided to patients in the United States (Institute of Medicine [IOM], 2001; Schuster, McGlynn, and Brook, 1998; Wenger et al., 2003). In a landmark study published in 2003, McGlynn et al. (2003) found that adult patients received only about 55 percent of recommended care and that adherence to clinically recommended care varied widely by medical condition. The follow-on analysis, conducted by Asch et al. (2006), found that the quality deficit was persistent across all sociodemographic subgroups and that although quality of care varied moderately across the sociodemographic subgroups, there was substantial underuse of recommended care regardless of income, race, or age. Other studies, such as those by Fisher et al. (2003a and b), have shown that among Medicare beneficiaries, there is substantial regional variation in the use of services and health spending. Also, regions where more services were provided did not show additional benefit to patients either through improved outcomes or improved satisfaction with care. These studies highlight that problems occur in both the underuse of recommended care services and the overuse of services.

Health care costs continue to rise at a steady pace and are anticipated to account for 18.7 percent of gross domestic product by 2014 (Heffler et al., 2005). In 2006, the federal government spent \$600 billion for Medicare and Medicaid for care delivered to its approximately 87 million beneficiaries; and it is anticipated that by 2030, expenditures for these two programs will consume 50 percent of the federal budget, a financial burden that will place funding for other discretionary programs at risk (McClellan, 2006). To improve quality and hold down growth in the costs of the Medicare and Medicaid programs, the Centers for Medicare & Medicaid Services (CMS) will need to explore alternatives to existing policies and practices.

The Disconnect Between Payments and Performance

Existing mechanisms for paying hospitals, both Medicare's per-hospitalization payments using diagnosis-related groups (DRGs) and the per diem payments used by commercial payers, do not differentiate payments to hospitals providing efficient, high quality care. Current payment policies in both the public and the private sector reward the quantity rather than the quality of care delivered and provide neither incentive nor support for improving quality of care. Historically, hospitals have gotten paid the same regardless of the quality of care they provided and, in some cases, may have even received additional payment for treatment of avoidable complications and for readmissions and complications that occurred as a result of providing poor quality care. Starting in 2008, CMS has announced that it will no longer pay Prospective Payment System (PPS) hospitals for the additional costs of certain preventable conditions acquired in the hospital (CMS, 2007a).

Calls for System Reform

The 2001 IOM report *Crossing the Quality Chasm* called upon policymakers in the public and private sectors to make reforms that would address problems of quality and inefficiencies. A key reform recommended by the IOM was to create financial incentives for quality and to make performance information transparent to ensure public accountability. More recently, the IOM made specific recommendations for implementing payment rewards for performance within Medicare in its 2006 report titled *Rewarding Provider Performance: Aligning Incentives in Medicare*. Additionally, the Medicare Payment Advisory Commission (MedPAC), which advises the U.S. Congress on issues related to the Medicare program, has recommended that Medicare adopt pay for performance (P4P) across various settings, including Medicare Advantage plans and dialysis providers and hospitals, home health agencies, and physicians (MedPAC, 2005).

Federal Actions to Reform the System

On August 22, 2006, President Bush issued an Executive Order, "Promoting Quality and Efficient Health Care," that requires the federal government to: (1) ensure that federal health care programs promote quality and efficient delivery of health care and

(2) make readily useable information available to beneficiaries, enrollees, and providers. These actions are designed to drive improvements in the value of federal health care programs.

To support this mandate, Department of Health and Human Services (DHHS) Secretary Michael Leavitt embraced “four cornerstones” for building a value-driven health care system:

- Connecting the health system through the use of health information technology (HIT)
- Measuring and making transparent quality information
- Measuring and making transparent price information
- Using incentives to promote high-quality and cost-effective care.

Building on these four cornerstones, CMS has taken steps toward using incentives and making quality information transparent in order to become a value-based purchaser of care. The steps taken include funding a number of demonstrations regarding use of financial incentives across hospital, physician, and home health settings, and implementing pay for reporting (P4R) for hospitals and physicians through the Reporting Hospital Quality Data for Annual Payment Update (RHQDAPU) program and the Physician Quality Reporting Initiative (PQRI). In particular, the RHQDAPU program, which was mandated under the Medicare Prescription Drug Improvement and Modernization Act of 2003 (MMA),² required hospitals to submit data on a defined set of performance measures to receive 0.4 percentage points of their annual payment update (APU). The performance data from RHQDAPU are made transparent to Medicare beneficiaries and the public through the CMS Hospital Compare website (<http://www.hospitalcompare.hhs.gov>). Section 5001(a) of the 2005 Deficit Reduction Act (DRA) expanded the set of RHQDAPU P4R performance measures and increased the differential payment for reporting from 0.4 to 2 percentage points.

² Public Law 108-173, December 8, 2003.

The 2005 DRA also authorized the DHHS Secretary, under Section 5001(b), to develop a plan for value-based purchasing (VBP) for Medicare hospital services commencing fiscal year (FY) 2009. Congress specified that the VBP plan consider the following design issues:

- The process for developing, selecting, and modifying measures of quality and efficiency
- The reporting, collection, and validation of quality data
- The structure, size, and source of value-based payment adjustments
- Disclosure of information on hospital performance.

Through implementation of VBP for Medicare hospital services, CMS would provide differential payments to hospitals based on their performance (i.e., P4P).

DEVELOPMENT OF THE VALUE-BASED PURCHASING PLAN

In response to the DRA mandate, CMS created an internal hospital VBP workgroup with responsibility for developing the VBP plan. To inform the development of the plan, the Assistant Secretary for Planning and Evaluation (ASPE), in collaboration with CMS, contracted with the RAND Corporation in July 2006 to conduct a literature review to synthesize the empirical evidence that exists on P4P in the hospital setting and an environmental scan of the existing P4P landscape.

To take advantage of the experimentation going on nationally with respect to P4P program design and implementation, RAND held discussions with P4P program sponsors, hospitals, hospital associations, data support vendors, and organizations experienced with small and rural hospitals to capture the array of experiences connected with the design and implementation of P4P and P4R programs. The discussions were necessary because this type of descriptive information and this level of detail about program design are not typically contained in peer-reviewed journal articles that summarize the results of P4P interventions. Additionally, many of the demonstration experiments are still in their infancy, and little has been formally documented about the related experiences.

RAND was tasked to:

- Identify and describe the concept of inpatient and outpatient hospital P4P
- Review the existing literature on inpatient and outpatient hospital P4P (theoretical and applied)
- Review existing inpatient and outpatient hospital P4P programs, examining their design features and evaluating the lessons being learned
- Summarize and synthesize the findings from the environmental scan, which would then be used to inform the discussions and design considerations of the CMS VBP workgroup tasked with developing the VBP plan for Congress.

Table 1 highlights core design issues that were examined as part of the environmental scan. Appendix A contains a complete listing of the design issues that were explored.

Table 1: Design Issues Explored with Program Sponsors and Hospitals

Overview
<ul style="list-style-type: none"> • The goals of existing P4P programs and demonstrations in the hospital setting
<ul style="list-style-type: none"> • Whether and how hospitals were included in the design and implementation of P4P and P4R programs
<ul style="list-style-type: none"> • The mechanisms used to monitor for unintended consequences, such as inappropriate clinical care or gaming of data to secure bonus dollars
<ul style="list-style-type: none"> • Lessons learned by organizations with P4P and P4R programs in practice or participating in demonstrations
Measures
<ul style="list-style-type: none"> • The measures of performance (clinical effectiveness, efficiency, patient experience, care coordination/transitions, etc.) that are currently being used for both inpatient and outpatient hospital care in practice and in demonstrations
<ul style="list-style-type: none"> • The measures selection criteria being used by P4P and P4R programs
<ul style="list-style-type: none"> • Methodological issues around P4P, including the level of aggregation of measures (i.e., composite scoring, weighting); the establishment of benchmarks, thresholds, and targets; risk adjustment; and opportunities for gaming
Data
<ul style="list-style-type: none"> • The data collection, data management, reporting infrastructure, and data outreach required to implement existing P4P programs
<ul style="list-style-type: none"> • Methods being used to validate data for use in P4P programs
Payment Mechanism
<ul style="list-style-type: none"> • The types of incentives, financial or non-financial, that currently exist or are under consideration, and what has been learned from various incentive structure designs
<ul style="list-style-type: none"> • Examining the basis for payment, such as paying on meeting a threshold, improvement, and/or high achievement
<ul style="list-style-type: none"> • The levels (fixed dollar, percentage of payments) and types (negative versus positive) of financial incentives being used
Public Reporting
<ul style="list-style-type: none"> • How information from public reporting systems is being used, and the impact of this information
<ul style="list-style-type: none"> • Strategies for simplifying public reports to facilitate use and understanding
Outpatient
<ul style="list-style-type: none"> • Whether outpatient hospital services should be incorporated into VBP in the future
<ul style="list-style-type: none"> • Extent to which current P4P programs include measures of hospital outpatient services

This chapter builds the foundation for subsequent chapters of this report by defining P4P and its dimensions and by providing the policy context underlying the rationale for P4P as a system reform strategy.

Defining Value-Based Purchasing

VBP is a strategy that strengthens the link between quality and provider payments by rewarding providers that deliver high-quality, cost-efficient care. VBP encompasses a number of activities that can be used individually or as a mutually supportive set to engender provider behavior change. One activity that falls under the VBP umbrella and has garnered much attention and interest in recent years is P4P. P4P explicitly links health care providers' pay to their performance on a set of specified measures such that better-performing providers receive higher payments than do lower-performing providers. The term *provider*, which we use throughout this report, encompasses a broad spectrum of health care providers: hospitals, individual physicians, physician practices, medical groups, and integrated delivery systems.

P4P programs seek to align measurement of and payments to providers with a program sponsor's goals, such as the delivery of high-quality, cost-efficient, patient-centered care. For example, if a program sponsor is seeking to improve patient outcomes, the program will include either measures of risk-adjusted mortality or complications rates or clinical measures, such as the provision of disease-specific services. If that program sponsor also seeks to improve the cost efficiency of care, the program may also include readmission rates or risk-adjusted length of stay. P4P programs are designed to financially reward those providers whose performance is consistent with the program sponsor's identified goals.

Three other mechanisms that use financial and non-financial incentives also seek to incentivize changes in provider and/or consumer behavior as means to improve quality and efficiency in health care delivery. These three mechanisms were excluded from our environmental scan of P4P in the hospital setting per se, although public reporting is often a component of P4P programs and is a core quality improvement strategy that CMS is currently implementing through the RHQDAPU program. The mechanisms are as follows:

- **Provider profiling (or report cards)** is an internal activity through which a health plan or other organization distributes comparative performance information to providers in either a blinded or an unblinded fashion. This

information may be used as the basis for structuring tiered or high-performance networks, for P4P programs, or for quality improvement.

- **Public reporting** makes provider performance information available to consumers and the public more broadly to help inform decisionmaking and to hold providers publicly accountable as a means to incentivize providers to improve.
- **Tiered provider networks** separate providers into categories on the basis of costs and/or quality performance and provide financial incentives to consumers (i.e., lower co-payments or deductibles) to use providers placed in the high-performing tier.

Principles for Pay-for-Performance Programs

Numerous organizations have developed design principles for P4P programs in the hopes of influencing how CMS and other P4P sponsors structure their P4P programs (see Appendix B). Among these organizations are MedPAC, the Joint Commission, employer coalitions, the American Medical Association (AMA) and other physician groups, the American Hospital Association (AHA), and the Association of American Medical Colleges (AAMC).

The principles cover a wide variety of program design and implementation issues, and at times the recommendations made by the different organizations directly oppose one another. Five major areas of disagreement about P4P design and implementation issues are:

- Should P4P programs, especially in Medicare, be budget neutral or based on “new money”?
- Should P4P programs include negative financial incentives for participating providers?
- Should P4P programs include efficiency measures?
- Should P4P programs initially include measures of patient outcomes?
- Should the measures included in the program be stable or be modified over time?

There was also variation in the topics explicitly included by organizations in their statements. For example, physician organizations frequently include these principles: voluntary participation, no link between rewards and the ranking of physicians relative to one another, reimbursement of physicians for the administrative burden of collecting and reporting data, and physician involvement in program design.

There are, however, areas of consensus. Nine or more organizations endorsed the following principles/recommendations:

- P4P programs should be based on accepted, evidence-based measures.
- Risk-adjustment methods should be used to prevent providers from avoiding caring for patients who are more difficult to treat (i.e., are sicker or non-compliant).
- Incentives should be aligned with the practice of high-quality, safe health care.
- Programs should include positive incentives for the adoption and utilization of IT.
- Rewards should be based on improvements in care and exceeding benchmarks.
- Data collection for P4P programs should not place an undue burden on providers, or providers should be reimbursed for the costs of collecting and reporting data.

CONTENT AND STRUCTURE OF THIS REPORT

The remainder of this report presents the findings of RAND's environmental scan of hospital P4P. Chapter 2 reviews the empirical literature on the impact of hospital P4P. It also draws from the economics and organizational management theoretical literature that has examined the effect of incentives on behavior to assess possible implications for P4P program design. Chapter 3 summarizes our discussions with hospital P4P program sponsors nationally, focusing on a description of the measures being used by these programs, the structure of the incentive payments, operational issues associated with implementation, and lessons learned. Chapter 4 summarizes our discussions with hospitals that have been exposed to P4P and P4R efforts (such as the CMS RHQDAPU program, the Premier P4P demonstration, or private-sector P4P programs), hospital associations, and data vendors that support hospitals in their data submissions to the array of performance-reporting efforts. Our emphasis in these discussions was on learning what

hospitals thought about the set of performance measures for which they were being held accountable, the structure of the incentive payments, issues related to data submissions and the quality and validity of data used to score their performance, the importance of public reporting, barriers they saw as hampering their ability to comply with the program requirements, and lessons they had learned. As part of these discussions, we also focused on understanding the unique issues of small, rural, and Critical Access Hospital (CAH) hospitals that would affect their ability to participate in P4P programs. Chapter 5 concludes by summarizing the key findings from the environmental scan.

II. A REVIEW OF THE EVIDENCE ON HOSPITAL PAY FOR PERFORMANCE

This chapter summarizes the empirical evidence on the effect of P4P in the hospital setting, based on application and theory. We begin with a review of published studies that assess the impact of P4P programs on health care quality, safety, and/or resource use, including studies that address P4P in either the hospital inpatient or the hospital outpatient setting. We then follow with a summary of relevant lessons for hospital P4P that can be drawn from the management and economic literature on how individuals in general respond to incentives, and we consider the implications for structuring incentives to achieve the desired behavioral response.

SUMMARY OF THE EMPIRICAL EVIDENCE ON THE IMPACT OF HOSPITAL PAY FOR PERFORMANCE

Methods

Our review of the empirical literature on the effects of P4P included all peer-reviewed published studies describing the impact of a hospital P4P program for either inpatient or outpatient hospital services. We defined outpatient hospital services as any medical or surgical services performed primarily in an outpatient/ambulatory care setting that are billed through a hospital. Examples of outpatient hospital services include chemotherapy, outpatient surgery, and diagnostic tests such as colonoscopy. The review included any randomized control studies, quasi-experimental trials, and pre-/post-intervention studies. We only retained articles that reported empirical findings related to the effect of paying for quality, patient experience, and safety or resource use, specifically excluding articles focused only on the impact of changes in hospital payment, such as the shift to the Prospective Payment System (PPS) and P4P as applied to physicians in the ambulatory setting. Only studies that were in English and published in the last 10.5 years were included.

We searched for articles published between January 1996 and June 2007 using five bibliographic databases (PubMed, EconLit, CINAHL, Psycinfo, and ABInform) that

could include articles related to P4P and financial incentives specific to the hospital environment. Table 2 displays the search strategy and terms used to identify relevant articles for hospital inpatient and hospital outpatient settings separately.

Table 2: Key Terms Used to Search the Literature for Hospital P4P Studies

Hospital Inpatient	Hospital Outpatient
<ul style="list-style-type: none"> • #1 pay for performance OR p4p OR “pay for quality” OR “pay for value” OR “value based purchasing” OR “financial incentives” OR “monetary incentives” • #2 (bonus* OR reward* OR (incentive reimbursement)) AND quality • #3 hospital OR hospitals • #4 (Results from search #1 or #2) AND (Results from Search #3) • #5 NOT (organ donation) 	<ul style="list-style-type: none"> • “pay for performance” OR p4p OR “pay for quality” OR “pay for value” OR “value based purchasing” OR “financial incentives” OR “monetary incentives” <p>This resulted in a database of 1,575 articles. Within this database, we retained any article that included the following keywords:</p> <ul style="list-style-type: none"> • #1 “Outpatient clinic(s)” OR “outpatient hospital” OR “outpatient” • #2 “Annual payment update” • #3 “Chemother*” (chemotherapy) • #4 “Radio*” (radiology) • #5 “Emergency” (emergency room) • #6 “Physical ther*” OR “occupational ther*” OR “speech” (physical therapy, occupational therapy, speech therapy) • #7 “Ambulance” • #8 “Durable” (durable medical equipment) • #9 “ambulatory surg*” OR “outpatient surg*” OR “surgery” (ambulatory surgery) • #10 “laboratory” • #11 “colono*” or “endosc*” (endoscopy) • #12 “pathol*” (pathology) • #13 “catheter*” (cardiac catheterization)

We combined the results of this search strategy for each setting (conducted initially in November 2006 and update with articles published through June 2007) from

the five different databases and then eliminated duplicate articles. Titles and abstracts for these articles were reviewed, and potentially eligible articles were identified. The full text of the set of potentially eligible articles was then read to determine whether the article was appropriate for inclusion. Reference lists of the included articles were checked to identify additional relevant studies. To ensure our scan was comprehensive, we also consulted experts in the field of P4P and retrieved references from recent reports on P4P and payment reform from the IOM, the Joint Commission, MedPAC, and the Agency for Healthcare Research and Quality (AHRQ).

From the initial search strategy, we identified 902 non-duplicated articles for the hospital inpatient setting and 162 non-duplicated articles for the hospital outpatient setting. After the abstracts were reviewed, eleven articles were targeted for further review for the inpatient setting and zero for the hospital outpatient setting. Of the eleven articles, eight met our criteria for inclusion. After consultation with P4P experts and a review of relevant reports, one more paper was thought to be sufficiently important to include. It is a white paper, not published in the peer-reviewed literature, describing the early results of the CMS–Premier Hospital Quality Incentive Demonstration (PHQID). Our summary therefore focuses on the findings from nine articles that describe P4P intervention in the inpatient setting.

The methodological quality of the articles was assessed by evaluating the overall study design in terms of its strength in determining a causal relationship or an association between the intervention and the outcome. For example, we determined whether the study design was a pre-post measurement without a control group, a pre-post study with a control group (a quasi-experimental study design), or a randomized control trial. If there was a control group, we also assessed its adequacy, such as whether hospitals in the control group were reasonably similar to hospitals exposed to the P4P intervention. If there was no control group, we assessed whether the study controlled for pre-intervention trends in performance. Lastly, we assessed the studies' use of appropriate statistical methods for estimating an intervention effect. These characteristics were used to determine the quality of the studies being reviewed, with randomized control trials providing the strongest evidence of a causal relationship between the implemented

program and changes in performance measures, and uncontrolled studies providing weaker evidence.

Findings from the Literature Review

As of June 2007, few peer-reviewed studies existed on the use of financial incentives to affect quality, patient experience, safety, or the efficient use of resources. While more than 40 hospital-based P4P programs are operating in the U.S., few of them are undergoing formal evaluations to assess their impact.

The nine articles in our review address the impact of three separate hospital P4P programs in which formal evaluations have been occurring:

1. The Hawaii Medical Service Association (HMSA) P4P Program
2. The Blue Cross Blue Shield (BCBS) of Michigan Hospital Incentive Program
3. The PHQID.

Table 3: Summary of Design Features of P4P Programs Contained in Published Evaluation Studies

Hospital P4P Program	Type of Measures					Type of Performance Target		Form of Financial Incentive		
	Outcome	Process	Structure	Patient Experience	Patient Safety	Absolute	Relative	Bonus	Withhold	Penalty
HMSA	X	X	X	X		X	X	X		
BCBS of Michigan		X	X		X	X	X	X		
PHQID	X	X			X		X	X		X

Table 3 presents a high-level summary of key design features of each of these three P4P programs. Table 4 provides descriptive data on the evaluation studies. More detailed findings from our evaluation are in the following subsections.

Table 4: Summary of Evaluation Studies Examining Hospital P4P Programs

P4P Program	Article	Type of Study	Change in Performance	Control Group
HMSA P4P Program	Berthiaume et al., 2004	Describes uptake of one component of program and how many dollars were dispensed	No	No
	Berthiaume et al., 2006	Describes trends in measures	Yes	No
BCBS of Michigan Hospital Incentive Program	Nahra et al., 2006	Cost-effectiveness analysis	Yes	No
	Sautter et al. 2007	Qualitative interviews with leadership of 10 participating hospitals	NA*	No
	Reiter, Nahra, and Wheeler, 2006	Survey of participating hospitals to track behavioral responses	No	No
PHQID	Premier White Paper	Describes improvements in quality measures	Yes	No
	Grossbart, 2006	Evaluates improvements in quality versus a “matched” control group	Yes	Yes
	Lindenauer et al., 2007	Evaluates improvements in quality versus a “matched” control group	Yes	Yes
	Glickman et al., 2007	Evaluate improvements in quality versus a control group	Yes	Yes

*Change in performance was used to select hospitals for the interviews and not the outcome examined by the research.

Hawaii Medical Service Association Pay-for-Performance Program

Two papers evaluated the impact of the HMSA P4P program, which started in 2001 and targeted all 17 hospitals in Hawaii. The program had four components:

1. Compliance with the AHA's "Get with the Guidelines—Coronary Artery Disease" program, which encourages hospitals to improve compliance with the latest scientific guidelines for management of coronary artery disease. Hospitals could earn points by signing up for an AHA workshop, being recognized as a "Get with the Guidelines" hospital, using a patient management tool for data collection, and reaching 85 percent performance on at least three out of five process measures related to Acute Myocardial Infarction (AMI) care.
2. The hospital's case-mix adjusted rate of clinical complications and length of stay.
3. Patient satisfaction and physician satisfaction with emergency department and hospital inpatient care.
4. The hospital's self-reported success in implementing internal quality improvement programs.

The complication and length-of-stay measures focused on patients admitted to the obstetric service or undergoing one of the 18 most common surgical procedures, which accounted for approximately 50 percent of the surgical case volume. The HMSA hospital P4P program has been evaluated, and the results of the evaluation are contained in two articles by Berthiaume and colleagues (2004 and 2006).

Berthiaume et al., 2004: This study looks at the rates of participation in the "Get with the Guidelines—Coronary Artery Disease" component of the HMSA P4P program. The authors report that of the 13 hospitals in Hawaii with more than 30 admissions for acute coronary artery disease, 10 earned some points associated with participation in "Get with the Guidelines." The average incentive amount to the 10 hospitals ranged from \$5,514 to \$114,574 in one year. The authors state that the fact that 85 percent (11/13) of the eligible hospitals participated in "Get with the Guidelines" is noteworthy because this level of program adoption "is much higher than would be predicted by models of diffusion of innovation in healthcare." The authors report that the incentive dollars helped

provide support within hospitals for salaries and travel costs and led to substantial changes to the systems of care.

This study suffers from several limitations that restrict our ability to assess the impact of the P4P program. It reports only how many hospitals participated in the program at a single point in time, 2003—not whether participation, number of points earned, or scores on the myocardial infarction process measures increased over the intervention period. Since there was no control group, it is unclear whether participation in the “Get with the Guidelines” care improvement effort was truly driven by the incentive program versus other factors. Hospitals around the country were being encouraged to enroll in the program, and many of the measures that the program used were also being used by the Joint Commission and CMS as part of their quality measurement and improvement efforts. This study does not provide evidence on the impact of the incentive program in changing clinical process or outcome measures and how the results might generalize more broadly.

Berthiaume et al., 2006: This second study by Berthiaume and colleagues reports changes in the following HMSA P4P program areas: length of stay, complication rates, patient satisfaction, and the hospital’s internal quality initiatives. It does not report changes in the clinical process of care measures for AMI. The study design used pre-post measurement with 2001 as the baseline year and 2004 as the final year of available data. The HMSA program awarded \$9 million in financial incentives across all parts of the program in 2004.

The authors report that complication rates for both obstetric and surgical patients declined approximately 2 percentage points between 2001 and 2004. Average length of stay also decreased for both types of patients; surgical patients experienced a decrease in length of stay of approximately 1.2 days, whereas length of stay for obstetric patients decreased by approximately 0.4 days. Patient satisfaction with inpatient care remained stable (78 percent in 2001 versus 79 percent in 2004); satisfaction with emergency room care increased from 71 percent in 2002 to 75 percent in 2004. Lastly, the scoring mechanism for internal quality initiatives was changed halfway through the program; but between 2003 and 2004, the scores increased from 4.25 to 6.5 points out of a total of 10

possible points. The authors do not state whether the observed differences between time periods were statistically significant. However, confidence intervals shown in figures contained in the article appear to indicate that only the change in surgical length of stay was statistically significant.

The authors state that it is unclear whether these upward shifts in performance were caused by the HMSA P4P program intervention or other factors occurring more broadly, such as greater national emphasis on improvements in AMI care or efforts to reduce utilization. As is typical for P4P programs being implemented nationally, the HMSA program did not have a control group to determine the effect of the HMSA intervention separate from other factors that may have caused the observed changes.

Blue Cross and Blue Shield of Michigan Hospital Incentive Program

Two published papers have examined the impact of the BCBS of Michigan Hospital Incentive Program. This program was initiated in 2000 and fully implemented in 2001 between BCBS of Michigan and the 86 hospitals statewide with which it contracts. Under the incentive program:

1. Hospitals could earn up to a 2 percent bonus of the hospital's heart-related DRG payments by exceeding the median performance of all participating hospitals on several process of care measures related to the care of patients with AMI and Congestive Heart Failure (CHF).
2. Hospitals could earn incentives through participation in patient safety initiatives and community health improvement projects.

As of this review, no results have been published describing changes in quality metrics in response to this program. The three evaluation studies that have been published examine the cost-effectiveness of the program (Nahra et al., 2006), results of qualitative interviews with leadership at 10 participating hospitals (Sautter et al., 2007) and the results of a survey of organizational changes that participating hospitals reported making in response to the P4P program (Reiter, Nahra, and Wheeler, 2006).

Nahra et al., 2006: This study estimated the cost-effectiveness of the Michigan BCBS Hospital Incentive Program from the sponsor of the health plan program's

perspective. In estimating the costs, the researchers included incentive amounts paid to hospitals by BCBS and the costs of administering the program. Benefits from the program were estimated by using increases in performance on the process measures to calculate the number of patients receiving improved heart care. These calculations were combined with published clinical trials data to estimate how many quality adjusted life years (QALYs) would be saved from the improved heart care over the 2000–2003 period. The researchers estimated that the clinical quality improvements observed would lead to savings of 733 to 1,701 QALYs. Based on this calculation and the cost of the program to the health plan, the cost per QALY was between \$12,967 and \$30,081, a range generally considered to be cost-effective (Ubel et al., 2003). This study illustrates that modest quality improvements can lead to substantial gains in QALYs saved. Additional unpublished information obtained from the program evaluator (private communication J Wheeler) indicated hospitals reported incremental costs for participation in the P4P program were on average \$36,915 for large teaching hospitals and \$28,525 for other hospitals. Even taking these into account, the program would be considered cost effective.

One limitation of this evaluation is the absence of a control group or trend data from the period prior to intervention to know whether the observed improvements in heart care are attributable to the BCBS Hospital Incentive Program or other secular trends in care for heart disease (such as the CMS RHQDAPU pay-for-reporting program, the Joint Commission quality improvement initiatives, or the CMS 7th Scope of Work quality improvement efforts).

Reiter, Nahra, and Wheeler, 2006: This study reports the results of a survey of the 86 hospitals participating in the BCBS of Michigan Hospital Incentive Program. The survey measured the effect of participating in the program on hospital behavior. The study outcomes were the number of hospitals self-reporting that the incentive program had triggered a structural change or a process change within the hospital. Structural changes included the formalization of a quality management staff position or a change in the person responsible for quality. Process changes included implementation of a computerized physician order entry (CPOE) system or creation of case-management

teams. Of the 86 hospitals participating in the program, 66 responded to the survey (70 percent response rate). Of the respondents, 32 (48 percent) reported that they had made a structural change and 39 (59 percent) reported they had made a process change in response to the P4P program. Overall, 75 percent of the responding hospitals reported making at least one type of change as a result of the BCBS Hospital Incentive Program. The most common structural change was involvement of leadership and greater board engagement in quality improvement. The most common process changes were instituting physician education, developing case-management teams, and increasing leverage with hospital physicians. The authors observed that since most of the process changes focused on physician behavior, a hospital's ability to improve quality might depend on its "willingness or ability to exert influence over physicians."

While this study found changes in the behavior of hospitals in response to the P4P program, it does not demonstrate that the changes made by hospitals resulted in clinical quality improvements. Additionally, the combination of the BCBS P4P program and other quality improvement interventions that were occurring simultaneously (e.g., CMS P4R, Joint Commission quality improvement) may have created a tipping point for the hospitals to make the reported behavioral changes. This study does not include a control group, which means there is no way to determine whether hospitals not exposed to the BCBS of Michigan Hospital Incentive Program were making similar changes.

Sautter et al., 2007: This qualitative study described the findings of semi-structured interviews with senior management and cardiologists at 10 Michigan hospitals participating in the P4P program. Fifty-four hospitals that participated in the P4P program and reported cardiac care performance to BCBSM 2002-2004 were placed into strata based on their changes in performance on one of the quality measures used in the incentive program, assessment of ventricular function among CHF. Hospitals from each strata were selected for interviews to obtain variation in hospital characteristics, such as size and teaching status. Among the 10 hospitals selected for interview, 7 had improved their performance, 2 were top performers at baseline and remained top performers, and 1 hospital showed declining performance. Only two of the 10 hospitals interviewed reported that the P4P incentives were a driver for quality improvement; eight of the 10

reported their facilities were undertaking these activities anyways or that the incentive was not large enough to be effective. The authors, however, are not sure these responses imply that without financial incentives performance would have improved to the same degree. They note, “incentive rewards clearly enabled some hospitals to make investments in quality.” In explaining the variation in quality improvement, the authors believe “underperforming hospitals with some infrastructures for quality improvement had the greatest success when presented with incentives.”

CMS–Premier Hospital Quality Incentive Demonstration

Four studies have analyzed the effects of the PHQID, a three-year CMS-sponsored demonstration project initiated in 2003. The PHQID program allowed for voluntary enrollment (i.e., hospital self-selection into the study) and only included hospitals using the Premier Perspectives data system—two factors that may hinder the ability to generalize the experience of the demonstration hospitals to non-demonstration hospitals to the extent that participants differ in important ways from non-participants. It should also be noted that at the start of the Quality Incentive Demonstration period, CMS had already begun implementing its RHQDAPU P4R program, whose set of measures overlapped substantially with that of the PHQID. The PHQID program includes 34 measures of which 22 overlap with RHQDAPU measures in the areas of AMI, pneumonia, CHF, and surgical infection prevention.

The PHQID demonstration includes 262 hospitals across 38 states. Hospitals were paid an annual bonus based on their composite performance scores in five clinical areas: AMI, Coronary Artery Bypass Graft (CABG) surgery, Community Acquired Pneumonia (CAP), CHF, and hip and knee replacement surgery. The bonus dollars represented new money. Hospitals that did not achieve a minimum level of performance in the third year of the program (defined by the lowest two deciles of performance in the first year of the program) were assessed a financial penalty.

Premier, Inc., 2006: Premier published its own report describing the PHQID and the observed quality improvements from the first year of the incentive program’s implementation. Premier reported that between the first and fourth quarters of the first

year of the program (October 2003 to September 2004), significant gains were made across the measures in the study, with an average 6.6 percentage point improvement across the five clinical areas. Within each of the five clinical composites, AMI performance increased from 87.4 percent to 90.8 percent, CABG surgery performance improved from 84.9 to 89.7 percent, CAP improved from 69.3 percent to 79.1 percent, CHF increased from 64.6 percent to 74.2 percent, and hip/knee replacement improved from 84.5 percent to 90.1 percent.

Although these results are positive, it is difficult to draw conclusions from this study about the effect of the PHQID program. An important challenge with this study is trying to assess whether non-participants were achieving similar gains in performance given the absence of a control group. As documented by Williams et al. (2005), there has been a strong trend across the country toward improvement in many of the same measures used as a basis for incentives in the PHQID. Disentangling the impact of the CMS-Premier demonstration from concurrent Joint Commission and CMS quality improvement efforts (i.e., RHQDAPU and the 7th Scope of Work) requires that there be a set of comparison hospitals with similar characteristics but no exposure to the PHQID. Selection bias is another issue to contend with in explaining the observed outcomes, since Premier hospitals that chose to participate in the PHQID had higher baseline quality scores than did Premier hospitals that chose not to. Thus, improvements in performance may be stem partly from characteristics of the hospitals that participated rather than from the incentive program itself.

Grossbart, 2006: This study examined the impact of the PHQID but focused solely on a subset of hospitals participating in the Premier system. The study followed the performance of hospitals in the Catholic Healthcare Partners system—four that chose to participate in the PHQID and six that chose not to participate and were used as controls. The analysis was limited to a subset of 17 of the 34 measures used in the PHQID initiative (for three clinical conditions, AMI, CAP, and CHF) that were collected by both intervention and control groups of hospitals as part of reporting for Joint Commission ORYX Core Measures program.

All 10 hospitals showed significant improvement across the measures. Those participating in the PHQID had a greater statistically significant increase in performance than did the non-participants. Across 17 measures, PHQID hospitals improved their scores by 9.3 percentage points, versus 6.7 percentage points for non-participating hospitals. Although the researchers matched hospitals on a number of key characteristics, one important limitation of this study is that they did not match them on baseline performance. The findings are confounded by the fact that the participating hospitals started at a higher level of quality than the non-participants did (80.4 percent versus 78.9 percent).

Much of the observed difference between the two sets of hospitals was driven by greater improvement in CHF care (19.2 percentage points for PHQID hospitals versus 10.9 percentage points for non-participants). Across the 17 measures examined, the two measures with substantial differences in improvement between PHQID and non-participating hospitals were (1) discharge instructions for patients with CHF (40.1 percentage points improvement for PHQID hospitals versus 14.6 for non-participants), and (2) pneumococcal vaccine delivery for patients admitted with pneumonia (31.6 percentage points improvement for PHQID hospitals versus 22.1 for non-participants). These two measures likely drive a substantial fraction of the overall observed differences in improvement between participating and non-participating hospitals.

The PHQID P4P intervention did not occur in isolation; it was conducted in an environment in which several national quality improvement efforts already in play were focusing on the same measures, particularly the HQA measures. These efforts included the CMS RHQDAPU program, the Joint Commission's quality improvement initiatives, and the CMS 7th Scope of Work. Across the subset of ten HQA measures, the study found that there was no difference in the amount of improvement: 5.4 percentage points for PHQID hospitals, and 5.1 percentage points for non-participating hospitals. This very modest difference, while not statistically different, raises questions about the added value of P4P incentives above and beyond other quality measurement and feedback efforts, particularly the RHQDAPU P4R intervention, which appears to have driven improvements in performance nationally (Lindenauer et al., 2007). Similar levels of

improvement were observed among all hospitals nationally, both those exposed to P4P and those exposed to public reporting, measurement, and feedback interventions.

The author described why only some Catholic Healthcare Partners hospitals chose to participate in PHQID. With the exception of those with the highest volume, hospitals saw the costs of participation, particularly for the extra staff required for the additional data collection, as being too high; and most hospital CEOs believed there was little to be gained by participation. Those that chose to participate thought the experience would provide them with a market advantage and a head start given the growing numbers of P4P programs in the market.

It is unknown from this study whether the ten Catholic Healthcare Partners hospitals making up the set are similar to or different from other hospitals nationally in ways that are important. To the extent that these hospitals differ in important ways from other hospitals, the results may not be more broadly generalizable. Another unknown is how Catholic Healthcare Partners hospitals and the system in which they operate may differ from other hospitals nationally, such as in the amount and type of systems and quality resource support that were provided. The six hospitals serving as the control group were selected because of “similar levels of service,” and the hospitals were shown to be similar in terms of availability of an open heart program and average number of beds, discharges, and case-mix index. A more rigorous method of selecting controls would have been to match each intervention hospital to a control on these characteristics as well as on baseline performance.

Lindenauer et al., 2007: This study provides the most comprehensive evaluation of the impact of the PHQID that has been published to date. The paper describes changes in performance on 10 measures that occurred over a two-year period, between the fourth quarter of 2003 and the third quarter of 2005. The study examined 207 PHQID hospitals and 406 control hospitals that were submitting performance data as part of the RHQDAPU program. Hospitals in this study were matched on bed size, teaching status, region (Northeast, Midwest, South, or West), location (urban or rural), and ownership status (for-profit or not-for-profit).

On an overall composite measure constructed from the 10 measures, PHQID hospitals experienced greater improvement than the control hospitals did (9.6 percentage point improvement versus 5.2 percentage points). This difference was seen consistently for each of the three clinical conditions (AMI, CAP, and CHF) for most individual measures and on an appropriate care measure.³ The greatest amount of improvement was seen among hospitals with the lowest baseline performance.

The authors did a number of sensitivity analyses to assess whether this differential response stemmed from a volunteer bias, meaning that Premier Perspectives hospitals that volunteered to select into the PHQID program were inherently different from Premier Perspectives hospitals that did not volunteer. The researchers found that after controlling for baseline performance and volume of patients, the difference in improvement decreased from 4.3 percentage points to 2.9 percentage points, but the improvement was still statistically significantly higher in PHQID hospitals. When all hospitals *eligible* to participate in the PHQID program were compared to all other hospitals nationally (so those exposed to RHQDAPU), the performance differential remained, but the gap was smaller (the difference in absolute performance point improvement was 2.1 points). Overall, this article provides the strongest evidence that the PHQID is improving performance beyond what is accomplished by public reporting of performance for some of the 10 measures, albeit modestly, once the hospitals' baseline performance and characteristics are controlled for. Because this study describes the impact of the P4P intervention on top of the measurement and public reporting intervention, we do not know how the impact of the P4P intervention would have differed absent public reporting.

Glickman et al., 2007: This study examined the impact of the PHQID on hospitals voluntarily participating in the national quality improvement initiative Can Rapid Risk Stratification of Unstable Angina Patients Suppress Adverse Outcomes with Early

³ An appropriate care measure is a composite measure that assesses what percentage of time a patient with a given clinical condition (e.g., AMI) received all of the recommended processes of care—in other words, how often a hospital provided “optimal” care for a patient with a given clinical condition.

Implementation of the American College of Cardiology/American Heart Association (ACC/AHA) Guidelines (CRUSADE). Hospitals participating in CRUSADE received performance feedback, including comparisons with other CRUSADE hospitals and national standards, as well as a variety of educational interventions. Trends in the cardiac care of patients with non-ST-segment elevation AMI from July 2003 to June 2006 were compared for 54 CRUSADE hospitals participating in PHQID and 446 CRUSADE hospitals not participating in PHQID (i.e., controls). In addition to the AMI measures included in PHQID, the comparison also used eight AMI process measures not included in the demonstration. The study sought to determine whether participation in the P4P intervention gave an additional boost to performance improvement above that from the CRUSADE intervention.

Both PHQID and control hospitals improved performance on PHQID measures and the other AMI measures over the period examined. There were not statistically significant differences between improvement in the PHQID and control groups on the composite measure for either PHQID (7.2 percentage points and 5.6 percentage points, respectively) or other AMI measures (13.6 percentage points and 8.1 percentage points, respectively). PHQID hospitals had significantly greater improvement on three individual measures—two that were included in PHQID (aspirin prescribed at discharge, $p = .04$; smoking cessation counseling for active or recent smokers, $p = .05$) and one that was not included in the demonstration (lipid-lowering agent prescribed at discharge, $p = .02$). There were no statistically significant differences in improvements in inpatient mortality between the two groups. In both groups, hospitals with lower levels of performance at the start of the observation period demonstrated greater improvements in performance than did higher-performing hospitals.

The authors concluded that P4P leads to only very small improvements in performance beyond what can be accomplished through engagement in quality improvement initiatives. Like the Lindenauer et al. (2007) article, the Glickman et al. article demonstrates the importance of using control hospitals and controlling for baseline performance in any analysis of the impact of hospital P4P. This study's limitations are its focus on only one of the clinical areas included in PHQID and its narrow focus on

patients with non-ST-segment elevation myocardial infarction. In addition, since the hospitals included in the study voluntarily participated in CRUSADE, it is not known whether hospitals would demonstrate the same level of performance improvement if participation were not voluntary.

Summary of the Evidence on Hospital P4P Programs

As of June 2007, there were only nine studies on the impact of hospital P4P programs, one of which was not peer reviewed. All of these studies evaluated programs that targeted the inpatient setting, and none examined P4P interventions in the hospital outpatient setting. Among the studies examining changes in performance, each one reported improvements over time in at least some of the hospital performance measures or condition-specific composites included in the specific study; however it is difficult to disentangle the P4P effect from the effect of other quality improvement efforts that were occurring simultaneously. Improvements in hospital performance have been observed in response to feedback reports (Williams et al., 2005) and public reporting with a financial incentive for submitting data (Grossbart, 2006; Lindenauer et al., 2007).

The two studies with control groups saw very modest improvements in performance associated with P4P compared with what was accomplished with public reporting (Grossbart, 2006; Lindenauer et al., 2007), but one of these studies saw improvements in a few performance areas associated with P4P compared with what was seen for control hospitals participating in voluntary quality improvement activities (Glickman et al., 2007). It has been argued, however, that in order to accomplish sustained quality improvement, interventions should be multifaceted and focus on different levels of the health care system (Grol et al 2002; Grol and Grimshaw 2003). This implies that to be most effective, P4P should be partnered with other activities such as public reporting and internal quality improvement activities that also encourage quality improvement for the same clinical area.

There is less evidence of the effect of P4P on patient outcomes. Berthiaume et al. (2006) found improvements in complication rates for obstetrical and surgical patients in an uncontrolled study but did not report whether those improvements were statistically

significant. In the study by Glickman et al. (2007), they did not find significant differences in inpatient mortality improvement for AMI between PHQID and control hospitals. None of the studies evaluating PHQID separately analyzed the other patient outcome measures (for coronary bypass surgery and hip and knee replacement surgery) included in the program, so it is not clear whether improvements occurred in these measures.

Most of the published studies have significant methodological limitations. Six of the nine had no controls, which are critical for providing evidence of a link between P4P and performance improvements. This is particularly important given the documented temporal trend toward increasing performance on many hospital quality metrics. It is challenging to disentangle the effects of the increasing use of financial incentives from the effects of greater use of quality improvement initiatives on the local and national level as well as the increasing use of public reporting when all activities are focused on the same clinical conditions. One of the studies that used a control group only included six control hospitals, and it is unclear whether the controls utilized were appropriate.

Beyond the specific limitations of the nine studies, another important issue is whether the experience of these geographically confined incentive programs that took place in the context of established relationships between the individual hospitals and the program sponsors would reflect the experience of wholesale national implementation of a hospital P4P program by Medicare. Medicare is the largest payer of inpatient care in the nation, accounting for 30.4 percent of third-party payments for hospital expenditures (CMS, 2007b). Given the importance of this revenue source for hospitals, it is possible that the level of engagement by hospitals in a national P4P program would be higher than that experienced in the programs in Michigan and Hawaii; though in both Hawaii and Michigan, the incentive program was administered by the dominant commercial payor in each of those states. Another issue to consider when interpreting the impact of these smaller P4P programs and demonstrations is that they all generally focus on a small set of process measures covering a handful of diagnoses. It is unknown what the impact on raising quality performance more broadly might be if Medicare were to adopt a more comprehensive set of measures.

THEORETICAL LITERATURE AND IMPLICATIONS FOR P4P DESIGN

The published literature on the use of financial incentives in health care is sparse and provides little information about how specific design features may influence behavioral responses. P4P is common in industries other than health care, and economists and management experts have studied and developed theories on how individuals respond to financial incentives. In the sections that follow, we describe theories that are drawn from the economics and management literature and consider the implications of applying the findings from tests of these theories to the design of a P4P program. Our review is not exhaustive; instead it focuses on selected theories to illustrate how theory might inform program design to achieve the desired behavior changes. It should be noted that the theories described have examined the behavioral responses of individuals, not institutions. It is thus uncertain whether application of these theories would elicit the same type of behavior responses from organizations, such as hospitals.

Prospect Theory and the Role of Framing in Decisionmaking

P4P incentives are designed to change the behavior of providers and the systems in which they operate in ways that will improve quality or efficiency. Various factors, such as the size of the incentive, are likely to influence a hospital and its physicians' behavioral responses to a P4P program. For example, a large incentive would likely lead to a larger behavioral response than would a small incentive. Another factor is how an incentive is structured, or "framed," which can determine the behavioral response to it. Prospect theory is an economic theory that attempts to explain how individuals respond to the framing of choices (Kahneman and Tversky, 1979). What follows is a description of several applications of prospect theory and an exploration of the potential implications for structuring a P4P program.

Withholds May Have More of an Impact Than Bonuses

One aspect of prospect theory is the principle of "loss aversion," which finds that individuals are more sensitive to incentives when they perceive they are losing as opposed to gaining something. This effect has also been described as "losses loom larger

than gains.” This behavioral effect has been demonstrated in a series of experiments in which both doctors and patients are asked to make a choice of treatment—either surgery or radiation—for a patient with lung cancer. Both doctors and patients made different choices depending on whether the choice was framed as a loss (the probability of dying after surgery) or as a gain (the probability of surviving after surgery) (McNeil et al., 1982). In another experiment, Meyerowitz and Chaiken (1987) showed that a pamphlet that framed the benefits of self-breast examinations as a loss (lost ability to detect cancer early) led to a greater increase in the percentage of women doing these examinations than did an identical pamphlet that framed the benefits as a gain (gained ability to detect cancer early). The difference in the behavioral response for a choice framed as a loss rather than as a gain can be significant, almost twofold in magnitude (Kahneman and Tversky, 1979).

The principle of loss aversion may have implications for structuring a P4P incentive payment. Incentive payments can be structured as a withholding (a perceived loss in income)—for example, a portion of the hospital’s full payment for a service could be held back until the end of the measurement period and then released only if the hospital met the performance target—and they can be structured as a bonus (a perceived gain). The theory of loss aversion suggests that if the goal is to drive hospitals to make changes that improve quality or efficiency, withholding dollars with the likelihood of later releasing them based on performance (i.e., framing the incentive as a possible loss) may lead to a greater behavioral response than framing the incentive as a “gain,” in the form of a bonus, even if the same amount of money is at risk.

While framing something as a loss rather than a gain may result in a larger behavioral response, experiments have shown that doing so generally causes a negative reaction and violates what the parties exposed to the incentive believe to be fair. This point was illustrated in a study in which subjects were asked to respond to two decision scenarios. The economic impact of the two scenarios was the same, but one was framed as a loss, the other as a gain. In the first scenario, subjects were told that there was no inflation in the community and that employees were being asked to take a 7 percent wage cut (a loss). In the second scenario, subjects were told that there was 12 percent inflation

and that employees were being given a 5 percent raise (a gain). The result in both of these decision scenarios was the same—employees would all experience a 7 percent reduction in net earnings—but the emotional response differed. A majority of subjects (62 percent) judged the first scenario to be unfair, whereas only 22 percent thought the second was unfair (Kahneman, Knetsch, and Thaler, 1986).

In terms of P4P program design, this research suggests that hospitals would be more likely to perceive a bonus in a positive light than they would a payment withholding, even if the net financial impact is the same. This conclusion is supported by a finding from a recent survey of 79 physician group leaders: When given a choice in the structure of a P4P program, 59 percent preferred a bonus, 24 percent preferred a withholding, and 17 percent felt they were the same (Mehrotra et al., 2007).

A Series of Small Incentives Might Lead to More Quality Improvement Than Would One Large Incentive

Why do people go across town to save \$10 on a clock radio but not to save \$10 on a large-screen TV? After all, the same amount of money can be saved in both cases.

The explanation for the difference in behavioral response in these two scenarios is called the principle of “diminishing marginal utility” (Lowenstein, 2001): the perceived value of a sum of money becomes progressively lower when associated with an increasingly larger sum of money. Thus, for example, an individual perceives the difference between \$0 and \$10 as being greater than the difference between \$100 and \$110, which is perceived as being greater than the difference between \$200 and \$210, and so on. This principle asserts that people tend to judge such gains or losses as changes from their current state of well-being (or reference point), rather than their final states (Kahneman and Tversky, 1979).

When we apply these findings to hospital P4P program design, it may be more psychologically motivating to provide smaller, more-frequent incentive payments than to provide a larger, lump-sum incentive payment. As an example, consider that a total of \$1,000 in incentives is to be provided to a hospital based on its performance. According to the principle of diminishing marginal utility, the hospital’s behavioral response is

likely to be greater if the \$1,000 is divided into a number of payments—say, ten payments of \$100 each—rather than paid as a lump sum. The reason for the greater motivation is that each \$100 is perceived as a new \$100 gain, capitalizing on the steepest portion of the utility function (the difference between \$0 and \$100), rather than simply as an addition to the previous gains (for example, from \$500 to \$600) (Thaler, 1985).

One way to structure this type of incentive in a P4P program would be to link the incentive payment to each applicable hospitalization. For example, the hospital could receive an extra payment of \$100, on top of its usual DRG payment, for every patient admitted for pneumonia that received the care designated by the quality measure(s). This approach could lead to a greater behavioral change by the hospital than if it were to receive a lump sum, equal in dollar value, at the end of the year.

Uncertainty May Reduce the Behavioral Response

When given a choice, most people are risk averse; they will choose an option with 100 percent certainty over an option involving an uncertain but likely more valuable outcome. This principle of risk aversion is illustrated in a study in which subjects were given a choice between a one-week vacation that was certain or a three-week vacation they had a 50 percent chance of winning. The vast majority of subjects chose the one-week vacation (Kahneman and Tversky, 1979). Even though the 50 percent chance of a three-week vacation might be considered a more rational choice, most people will choose the sure thing because they perceive it to be a better choice than the possibility of getting nothing at all.

With regard to P4P program design, the principle of risk aversion suggests that decreasing the risk or uncertainty in the likelihood of receiving a financial incentive is likely to lead to a greater behavioral response to the incentive. Some P4P payment structures use relative thresholds, such as paying those in the top quartile of performance, as the basis for determining who “wins.” This type of payout scheme creates greater uncertainty for hospitals than do payment schemes that use absolute thresholds (i.e., a fixed target) for determining who receives an incentive payment. The reason for the greater uncertainty with relative thresholds is that the level of performance necessary to

earn the incentive is unknown until after the fact, when hospitals can be sorted by rank order of performance. In contrast, absolute thresholds known in advance and thus provide greater certainty to the individual or institution trying to hit the target. Because of the uncertainty they create, relative thresholds may reduce the behavioral response to an incentive more than an approach using an absolute threshold will. Similarly, a shared saving program, such as is being used in the CMS Physician Group Practice (PGP) demonstration, might lead to a reduced behavioral response, in this instance because the providers in the PGP face uncertainty about whether there will be cost savings to fund incentive payments. In contrast, the most certain incentive would be an adjustment to the fee schedule. For example, for every admission for myocardial infarction, a hospital would receive an extra \$100, on top of its DRG payment, if the patient received all applicable processes of care. In such an incentive system, the hospital would know that if its physicians provide these processes, it would definitely obtain the additional payment.

Reducing the Time Lags Between Performance and Receipt of Incentive Can Help to Achieve Maximum Response

In economics, the principle of discounting is based on the fact that individuals value having a sum of money now more than sometime in the future, even after accounting for inflation. The concept of discounting and the use of a discount rate are well accepted in both accounting and economics. Studies have found, however, that individuals discount in a way different than would be expected by classic economic theory. In one study, the vast majority of individuals chose to receive \$10 immediately rather than \$21 in one year (Loewenstein and Prelec, 1992). But when asked to choose between \$10 in one year and \$21 in two years, fewer individuals selected the \$10. Instead of discounting in a linear fashion, the individuals in these experiments were discounting at a steeper hyperbolic curve, which led to the name of this phenomenon: hyperbolic discounting.

The application of hyperbolic discounting to P4P program design suggests that minimizing the lag time between the performance being incentivized and receipt of the incentive may strengthen the behavioral response. Money received right away is

perceived as different in value from money to be received in the future—even the near future. For example, a hospital is more likely to implement an electronic medical record (EMR) if they know the money associated with doing so will be received quickly (e.g., within the next month) rather than years after the implementation. One criticism of current performance measurement and reporting programs is that the substantial lag between the provision of care (i.e., performance) and the reporting of results renders the results not actionable (Davies, 2001). Similarly, in a P4P program, the time required to collect and validate data and make the payout determination might mean that the incentive payment comes long after actual delivery of care. Substantial time lags may cause a hospital to see the incentive as occurring so far in the future that it is not worth pursuing. Strategies that tie payment to the provision of individual services or more frequent payouts may help reduce the time lag.

A Series of Tiered Absolute Thresholds May Be Better Than One Absolute Threshold

An individual's motivation and effort when faced with a goal greatly depend on that individual's baseline performance. Economists and psychologists have described this phenomenon as a "goal gradient" (Heath, Larrick, and Wu, 1999). If baseline performance is far away from goal performance, the individual exerts little effort, because the goal is viewed as not immediately attainable. As baseline performance gets closer and closer to goal performance, the individual exerts more and more effort to succeed. However, as soon as the goal is achieved, the motivation to improve decreases significantly. This phenomenon was illustrated in a study of a coffee shop reward program in which the tenth coffee purchased was free. Participants in this experiment slowly decreased the time between purchases of a coffee as they got closer to the free coffee (Kivetz, Urminsky, and Zheng, 2006).

The notion of a goal gradient may have application in structuring a hospital P4P program. This principle implies that there would be a greater behavioral response among hospitals if there were a series of quality performance thresholds to meet (e.g., increasing dollar amounts for achieving a 50 percent, a 60 percent, a 70 percent, an 80 percent, and a 90 percent performance threshold) rather than one (e.g., a 75 percent performance

threshold). If, for example, there is just one 75 percent quality threshold (rather than a series of thresholds), a hospital whose baseline performance is at 45 percent is likely to see the goal as too difficult and not likely to be achieved, and is thus less likely to devote resources to quality improvement. If there is also a 50 percent quality threshold, however, the hospital's leadership may see reaching the threshold as feasible and thus be more likely to devote resources to improving quality. A series of quality thresholds might also lead to a different behavioral response among hospitals that are doing well. In a single-threshold system with a goal of 75 percent, a hospital that is at 80 percent would have little reason to devote more resources to improve its quality performance any further. In a graded performance threshold system, however, this hospital would have an incentive to reach the highest threshold, 90 percent, to achieve additional payment. To stimulate continual improvement, some P4P programs have elected to use relative performance targets so that the bar keeps moving upward. However, absent some gradients or some allowance for payment along the entire continuum of improvement, a single relative threshold creates a cliff effect—meaning all or nothing winners.

LIMITATIONS IN USING ECONOMIC THEORIES TO PREDICT BEHAVIORAL RESPONSE

Multidimensional Output

Multidimensional output, or multitasking, refers to situations in which the responsibilities of an individual encompass multiple activities or outputs that may require different types of skills to accomplish (Holmstrom and Milgrom, 1991). A hospital's output includes many different components, such as managing a patient's chronic illness, the timely and efficient diagnosis of a patient's new symptom, counseling and advice on how to prevent illness, and emotional support.

Multitasking is relevant to P4P programs because the performance measures in these programs typically address only a narrow piece of a hospital's outputs or the processes that contribute to outputs. For example, a program may measure the provision of aspirin for a patient with AMI but not other processes or outputs that are difficult to measure, such as diagnostic acumen for a patient hospitalized with unclear symptoms. It

is hypothesized that if a large incentive is applied to one type of output, other outputs will be neglected, and overall care might worsen (Holmstrom and Milgrom, 1991). This reasoning is used to explain why few private-sector corporations put large fractions of employee pay “at risk,” making them dependent on measures of output for which only a small fraction of what contributes to output can be measured (Asch and Warner, 1996). A large financial incentive based on a narrowly focused set of measures may lead to the unintended consequence of having a hospital “teach to the test,” devoting resources to those things being measured and neglecting other important outputs that are not being measured.

There are several potential ways to overcome or minimize the problem of multitasking. One is to create an incentive program that addresses a broad array of a hospital’s outputs by applying a comprehensive set of performance measures. This approach has been taken by the primary care physician P4P incentive program in the United Kingdom, which has over 146 quality indicators covering clinical care for ten chronic diseases, organization of care, and patient experience (Doran et al., 2006). The challenge with this approach is to avoid creating a program that may be overly complicated and costly—absent efficient measurement tools. Another approach that employers in other industries have used is low-powered incentives (Asch and Warner, 1996). With this approach, the majority of an employee’s income is fixed, and only a small fraction is tied to an incentive. The incentive emphasizes the importance of the measured area but is not large enough to induce undesirable behaviors, such as gaming of the data to win or avoiding caring for sicker patients.

Intrinsic Versus Extrinsic Motivation

Empirical meta-analyses of studies that examined incentive programs show that such programs have a mixed response; some studies show an impact, and many others show little or even a negative impact (Rothe, 1970; Deci, Koestner, and Ryan, 1999; Cameron, Banko, and Pierce, 2001). Researchers have tried to reconcile the mixed results by theorizing that they are caused by a conflict between intrinsic motivation, which is a person’s inherent desire to do a task, and extrinsic motivation, which is the external

incentive—such as might be provided in a P4P program. Researchers theorize that instead of supporting intrinsic motivation, extrinsic incentive “crowds out” intrinsic motivation (Deci, Koestner, and Ryan, 1999). This theory is used to explain why financial incentives for blood donation are ineffective: they inhibit the altruistic benefit of blood donation (Titmuss, 1970). The explanation for this crowding-out effect is that when a task is tied to an extrinsic incentive, people infer that the task is difficult or unpleasant (Freedman, Cunningham, and Krismer, 1992).

Empirical evidence of this effect was provided by a study in which students who were asked to collect money for a charity were put into two groups, one that was given an external incentive (a small amount of money), and one that was not. The group that was given the incentive collected less money than the other group did (Gneezy and Rustichini, 2000). A meta-analysis supported this study’s finding that performance-contingent rewards significantly undermine intrinsic motivation (Deci, Koestner, and Ryan, 1999), but the finding is not without critics (Cameron, Banko, and Pierce, 2001). Similar concerns have been raised about the effect of P4P in health care and how it may violate a physician’s sense of professionalism (Berwick, 1995). Application of this theory would imply that a small P4P incentive could actually lead to lower performance if it is tied to something hospitals are intrinsically motivated to improve, such as quality of care.

A potential way to address the crowding out of intrinsic motivation is simply to increase the size of the financial incentive. A very large external incentive will crowd out any inherent intrinsic motivation; but, in turn, it may create a greater behavioral response than would be obtained through intrinsic motivation alone. Gneezy and Rustichini, in “Pay Enough or Don’t Pay at All” (2000), illustrated this concept in a study of the average percentage of correct answers on an IQ test for four groups of college students that were given different incentives—one group received no incentive for each correct answer, one received a small incentive for each correct answer, one received a medium incentive for each correct answer, and one received a large incentive for each correct answer. The group given no financial incentive outperformed the group given the small financial incentive (56 percent versus 46 percent of questions correct, respectively), and

the groups given the medium and large financial incentives (68 percent of questions correct in each group) outperformed both of the other groups.

The idea of using a large financial incentive to overwhelm the potential loss of intrinsic motivation is at odds with the recommendation to use low-powered incentives to mitigate the incentive to overfocus on measured areas of care to the detriment of unmeasured areas of care.

CONCLUSIONS

Together, the economic and management theories that we reviewed suggest that the way in which P4P incentives are structured, or framed, may influence whether they achieve the desired behavioral response. Incentives that are framed as withholdings, paid out in small and frequent payments, and paid out close to the time that care is delivered might drive the greatest behavioral response among targeted hospitals. Furthermore, in comparison to relative thresholds or one absolute threshold, a stepped number of absolute thresholds may be more likely to induce hospitals to devote resources to quality improvement. The two potential unintended consequences discussed serve as a helpful counterpoint to the economic theories. They emphasize that P4P incentives could lead to the neglect of other important, but unmeasured outputs in a hospital and that P4P programs could even have a negative impact on quality. Therefore, any program should closely monitor for these unintended consequences.

There are several important limitations and caveats to this interpretation of these theories. First, as noted above, the theories were developed to describe the behavior of individuals, not institutions; and it is possible that institutions may behave differently. Researchers have, however, applied theories of individual behavior to organizations and there is some anecdotal evidence that organizations respond similarly (Bazerman, Baron, and Skonk, 2001). Another caveat is that there are often practical reasons for not choosing the options suggested by these economic theories. For example, it was noted above that a more frequent payout might lead to a greater behavioral response. Yet this result might be outweighed by the higher administrative costs to the program sponsor of more frequent processing of data and payouts. An absolute threshold with an associated

incentive with a fixed dollar amount might lead to a greater behavioral response than a relative threshold with an associated uncertain incentive. Yet such an approach leads to greater risk for the payer, which could face the prospect of paying out much more in incentives than was budgeted if providers outperform the predicted improvement. In the United Kingdom's primary care physician P4P program, provider performance greatly exceeded the 75 percent predicted when the scheme was negotiated, so the cost to taxpayers was considerably more than expected (Doran et al., 2006). This could be avoided by setting a fixed incentive budget.

III. SUMMARY OF DISCUSSIONS WITH PAY-FOR-PERFORMANCE PROGRAM SPONSORS

Given the scarcity of empirical data showing the effects of P4R and P4P programs on improving quality, safety, or efficiency and showing the effects of design elements that may influence provider behavior, RAND held discussions with a broad cross-section of P4P programs to gather information on the current state-of-the-art of P4P program design and operation. In this chapter, we describe key design features of hospital P4P programs that were being operated by both private- and public-sector sponsors across the United States as of October 2006. In addition to this cataloging of the designs, we asked about issues confronted in implementing and operating a hospital P4P program. The insights and perspectives gathered through these discussions reflect more than half of all hospital P4P programs in operation at the time the environmental scan was conducted.

METHODOLOGICAL APPROACH

We used several sources to identify candidate private- and public-hospital P4P programs to construct the universe of hospital P4P programs:

- The published literature on P4P programs (e.g., CMS PHQID).
- The Med-Vantage annual survey of P4P programs (2006) and a review of our candidate list by Med-Vantage staff who had conducted the annual survey.
- Information provided by research and policy staff within leading professional organizations, including the Association of Health Insurance Plans (AHIP), AHA, Blue Cross/Blue Shield Association (BCBSA), and Joint Commission.
- The Leapfrog Compendium of incentive and reward programs (Leapfrog Group, 2007).

- A Lexis/Nexis search of major U.S. newspapers, a broad Google-based Internet search, and a search of relevant trade journals.⁴
- The knowledge accumulated by RAND project staff who have been directly involved in evaluating a number of P4P demonstrations, and
- Input from the project's Technical Expert Panel (TEP), some of whose members currently operate or are involved with P4P programs.

From this scan, we identified 41 candidate organizations thought to sponsor hospital P4P programs. We then cataloged the 41 programs by a range of characteristics (e.g., type of sponsor, geographic region, type of insurance product) and selected a subset of hospital P4P program sponsors for discussions. During the selection process, we attempted to include a broad cross-section of programs that would encompass the range of variation in program design and operation. The goal of pursuing this strategy, as contrasted with a pure random sample, was to provide a rich base of information for consideration by ASPE and CMS.

The characteristics we sought to balance in our purposive approach to sampling were:

- The inclusion of a broad array of sponsor types, such as single organization sponsors, multi-stakeholder coalitions, private- versus public-sector sponsors.
- The inclusion of different types of insurance products, such as health maintenance organizations (HMOs), preferred provider organizations (PPOs), point of service (POS), administrative services only (ASO), Medicare, and Medicaid.
- The programs needed to cover various geographic areas of the country because of the variation in market characteristics that could affect design.

⁴ The journals searched were *Managed Care*, *Hospitals and Health Networks*, *Modern Healthcare*, *Managed Health Care Executives*, *Healthcare Intelligence Network*, *Medical Economics*, *Managed Care Weekly*, *Modern Physician*, *Business Insurance*, *California Healthline*, *Managed Care Online*, and *Managed Care Magazine*. The search terms used included pay for performance, pay for quality improvement, financial incentive, bonus, reward, hospital payment, performance improvement, and quality initiative.

From the 41 programs, we selected 31 organizations and requested their participation in the discussions. We held discussions with 27 of the 31 organizations between August and December 2006. Of the four organizations that did not participate, one had no hospital P4P program, one declined to participate, one never replied, and for one we were unable to establish correct contact information.

The numerical statistics presented in the following sections reflect 23 of the 27 organizations. The four organizations excluded from our tabulations were in the planning stages of designing a P4P program or were the national plan office that delegated operation of P4P programs to the local plan. We did, however, include information gathered from our conversations with these four organizations in our descriptive summaries.

FINDINGS FROM DISCUSSIONS WITH PROGRAM SPONSORS

General Descriptive Characteristics of Hospital P4P Programs

- **Length of Time in Operation.** Of the 23 P4P programs, a majority were relatively new. Seven had made their first incentive payment to hospitals in 2006 or were about to make a payout early in 2007, five had made their first payout in 2004 or 2005, and 11 had made their first incentive payments starting in 2003 or earlier. Only one program reported making its first payout prior to 2000. Planning efforts for the P4P program typically started two to three years in advance of making the first payout.
- **Program Sponsorship.** Most programs were sponsored by individual commercial health plans and did not involve partnerships with other organizations. Only six of the 23 program sponsors reported partnering with other organizations to develop and operate their programs.
- **Type of Insurance Products.** Eleven of the programs included all commercial product lines in their hospital P4P programs, while the others focused their incentives on a narrower set of products. Of the P4P programs with a narrower focus, six focused on PPO populations, five on HMO populations, five on

- **Program Goals.** Nearly all sponsors (21/22) reported that the primary goal of their P4P programs was to improve the quality of care delivered to their members.⁵ Other program goals mentioned included improving the efficiency with which care is delivered (6/22), improving patient safety (5/22), and rewarding and recognizing top-performing hospitals (4/22). A number of sponsors also noted that they were interested in strengthening hospital quality improvement department/activities, improving patient experience, and improving their relationships and ability to work collaboratively with hospitals.
- **Overall Program Structure.** Programs were typically voluntary (17/22). Hospital P4P sponsors reported that they often implemented P4P through contract negotiations (11/22), meaning that the program was rolled out on an individual hospital basis as individual contracts came up for renewal, and that the specific terms may have been customized to the individual hospital. Consequently, this process translated into a slower program rollout compared with programs that shifted to universal adoption of the P4P program in a single contract modification affecting all hospitals at the same point in time. Several sponsors noted that some hospitals have considerable leverage in these contract negotiations as a function of having significant market share or being “the only game in town.” This situation contrasts with the experience of physician-level P4P programs, in which the majority of physicians practice individually or in small practices, which means they have less bargaining strength to negotiate the terms of the P4P contract. Although the programs were voluntary, our discussions with hospitals revealed that most hospitals approached by P4P sponsors agreed to participate, so penetration was high. Sponsors reported that they usually did not include specialty and small and/or Critical Access Hospitals (CAHs) in their P4P programs,

⁵ Any denominator less than 23 indicates that one or more of the organizations did not respond to the question. Non-responses were typically caused by limited time or a respondent’s inability to answer the question.

primarily because of the challenges of not having enough patient events to score to produce stable performance estimates (i.e., the small-numbers problem). There was an exception; one program sponsor designed a P4P program specifically to enable participation by rural hospitals.

Measures

- **Measure Set Determination.** We identified two general approaches used by sponsors to determine the measure set for their hospital P4P programs. The first is a standardized, “one-size-fits-all” approach in which the measures applied to hospitals in the program do not vary. The second approach involves customization in one of two ways: (1) each hospital, in consultation with the program sponsor, selects from a structured, pre-determined menu of measures a subset on which to be measured (i.e., measures are from a pre-determined menu), or (2) each hospital works with the program sponsor to create a customized set of measures from the universe of measures that exist (i.e., measures are not from a pre-determined menu). Regardless of how the measure set was determined, many programs used all-payer data to construct the measures, primarily to ensure adequate amounts of data to score hospitals (i.e., to avoid the small-numbers problem).
- **Common Measure Types.**
 - **Clinical Quality.** Consistent with their key goal of improving clinical quality, all sponsors included clinical process and/or outcome measures as part of their hospital P4P programs (23/23). Process-of-care measures were much more commonly included (22/23) than outcomes were (3/23). The reasons cited for the focus on process measures included the availability of measures and performance scores collected and reported by national organizations such as the Joint Commission and CMS, and concerns about the adequacy of risk adjustment for outcome measures. There is substantial overlap between the measures included by the Joint Commission, CMS, and HQA (as shown in Appendix C, which lists existing hospital measures and their sources). The most frequently used process measure sets were:

- The Joint Commission's "core" measures (10/23)
- The CMS' P4R (RHQDAPU) ten starter-set measures (7/23)
- The HQA-approved measures (that have since been incorporated into the RHQDAPU program) (5/23), and
- The Surgical Care Improvement Project (SCIP) measures (3/23).

The most frequently tracked outcome measures were:

- Complications of care (e.g., Healthcare Cost and Utilization Project measures concerning pneumonia after major surgery) (3/23)
 - Mortality (3/23).
- **Patient Safety.** Another important area of measurement used by a large number of program sponsors (16/23) was patient safety. Among the most commonly used measures were:
- 3 Leapfrog Leaps
 - CPOE (12/23)
 - Use of Intensivists (9/23)
 - Evidence-based Referral based on Volume (6/23)
 - National Quality Forum (NQF) Safe Practices (4th Leapfrog Leap) (7/23)
 - Safe Medication Practices (6/23).
- **Efficiency or Resource Use.** Approximately half of the program sponsors included measures of efficiency or resource use in their P4P programs (11/23). A challenge cited in this area was identifying reliable and valid measures, given that their development has lagged that of clinical measures. Resource use measures most frequently included were:
- Readmission rates (5/23)
 - Average length of stay (4/23).

Other resource use measures used by sponsors included unit cost, avoidable days, and admissions per 1,000 members.

- **Patient Experience.** Measures of patient experience were used by many sponsors in their P4P programs (9/23). They often used “homegrown” metrics (6/23). Many said that, in moving forward, they anticipated using the emerging national standard, H-CAHPS, which was undergoing approval by the NQF and they expected would be required by CMS under the RHQDAPU program.
- **Structure.** Some sponsors were also focusing on the structural components of hospitals (9/23). Typically, these measures center on use of an electronic health record (EHR) or other IT implementation beyond the use of CPOE (5/23). A notable exception was one sponsor’s inclusion in its P4P program of whether hospitals used rapid response teams.
- **Quality Improvement.** Some sponsors (8/23) included metrics related to hospital quality improvement activities, which was consistent with their desire to improve the quality of care delivered to their members. More specifically, some are taking into account participation in the following quality improvement efforts:
 - Regional quality improvement initiatives (3/23)
 - National registries/databases (3/23)—for example, the registries managed by the ACC and the Society of Thoracic Surgeons
 - Internal quality improvement initiatives (2/23)
 - Institute for Healthcare Improvement’s (IHI’s) 100,000 Lives Campaign (2/23)
 - AHA’s “Get with the Guidelines” program (coronary artery disease, stroke) (2/23).
- **Administrative.** Only a small number of the sponsors with whom we spoke included administrative performance measures (5/23). When used,

these primarily focused on metrics having to do with claims submissions, such as:

- Number of claims re-submitted (2/23)
 - Electronic claims submitted (2/23).
- **Measurement Selection Criteria.** Sponsors consistently said that one of the most important criteria they use in selecting measures for their hospital P4P programs is consistency with other reporting activities (17/23), the objective being to help minimize hospital reporting burdens (15/23). They said that coordinating with other efforts, such as Joint Commission core measures and CMS RHQDAPU measures, makes it easier to launch and maintain their own programs. Doing so was considered essential for avoiding a cacophony of measures and to help set a collaborative, rather than combative, tone with hospitals. Although many of the sponsors valued the ability to use existing CMS and Joint Commission reported measures, they reported that the current set of measures was too narrow in scope and that there was a need to expand the set of measures to more comprehensively measure the performance of a hospital. Additionally, the sponsors indicated that performance has “topped out” on many of the measures (e.g., care for AMI), rendering them of less utility for quality improvement or for distinguishing differences between hospitals. Evidence-based measures (13/23) and/or endorsement by known organizations (such as NQF, Joint Commission, or HQA) (12/23) were also cited as key factors used in selecting measures. This not only assists with consistency across programs, but also reduces “pushback” from hospitals, especially in the case of measures that have been endorsed by HQA. Lastly, the practical points of ease of data collection (12/23) and data availability (12/23) were also important considerations in measurement selection.
 - **Risk Adjustment.** Many sponsors risk-adjust some of the measures in their program (15/23), generally outcomes of care, complications, and/or cost/efficiency measures. All sponsors noted that they use the risk adjustment methods recommended by the organization that developed the measure.

- **Composites.** Many sponsors used composite measures, which summarize performance across multiple individual measures, in contrast to reporting individual metrics (17/23). Composites are typically being used for payout (10/23) or in report cards to facilitate consumer understanding (8/23). Composites were frequently produced at the condition level, such as AMI or CHF. Composites can take a variety of forms, ranging from an average of performance on the individual measures weighted by the size of the denominators, to assessing whether the patient received all of the measured care for which they were eligible (referred to as the appropriate care composite). Because fewer hospitals provide the right care 100% of the time to patients with any given condition, the use of an appropriate care composite typically results in a performance score that is lower than scores for individual measures. Shifting the performance measure to achievement of all recommended care can reduce the extent to which hospital scores “top out,” which may have occurred for individual measures comprising the composite.
- **Piloting Measures.** Sponsors expressed mixed thoughts about the need to pilot the measures being used in their P4P programs prior to payout. Some felt strongly that a trial run is “necessary to be fair,” especially if using newly created or not commonly used measures. Others, primarily those adopting measures used by the Joint Commission or CMS, thought that hospitals have had enough time to get used to both measurement and P4P and that, consequently, it was time to “just get on with it.”

Data Collection and Validation

- **Data Collection.**
 - **Data Sources.** As with measurement selection, a key driver of data sources used was the goal of minimizing hospital burden. As such, there was heavy reliance on the use of data already collected by other entities (e.g., CMS, JCAHO) (14/23) or administrative data (either their own or from state reporting efforts). However, the clinical information used to populate measures for national measurement efforts is largely still being

gathered from medical records, as opposed to claims data or EHRs, so this still represents a significant burden to hospitals. Although EHRs in particular are often touted as a panacea for the burden of data collection, many organizations do not yet have EHRs. And even if they do, the data captured by EHR are in text versus data fields, which makes the tool difficult to use for measure construction. Even with an EHR, manual review is still required to extract relevant information. Other data sources used by program sponsors included (1) hospital self-reports, such as formal attestations (e.g., Leapfrog) or informal, in-depth conversations (e.g., with small programs) (16/23); (2) plan administrative/claims data (13/23); (3) patient experience survey data (10/23); and (4) national databases (3/23).⁶

- **“Small-numbers problem.”** Lack of an adequate number of cases was mainly an issue for hospitals that were small and/or CAHs, according to the sponsors with whom we talked. However, even for larger hospitals, a small number of events could occur; and if the data were based solely on a single payer’s data, the numbers would be insufficient for producing a stable score. Sponsors reported addressing the small-numbers problem primarily by using all payer data (versus only sponsor data) to score hospitals. Additionally, some sponsors allowed the data to drive which measures were tracked—by looking to see which measures had substantial patient volume. Another approach was to use participation in quality improvement activities or implementation of health information technology. Few sponsors reported using composite measures⁷ or multiple years of data, which borrow strength across the data to address the small-numbers problem.

⁶ Sponsors cited use of the AHA “Get with the Guidelines” database, the American College of Cardiology, and the Centers for Disease Control’s National Health Safety Network (NHSN).

⁷ As previously described, there are a variety of methods that can be used to construct composite measures and all of the methods would help mitigate the small numbers problem. For example, the appropriate care model does not create more denominator events to be scored.

- **Timeliness.** Timeliness of data was a concern, especially for quality improvement purposes. According to many sponsors, the typical lags of several months to half a year or more—for data collection, cleaning and processing, validation, and reporting—rendered the information useless to hospitals for improving performance in real time. These lags also affected the length of time between actual performance and when incentive payments were made, leading to a disconnect between these two events. Sponsors expressed a desire to obtain data as close to real time as possible in order to strengthen the impact of feedback to providers and other hospital staff.
- **Accuracy.** Sponsors expressed concern about the accuracy of coding administrative data, noting that hospitals potentially face the conflicting goals of coding to increase reimbursement versus coding to reflect care that was actually provided.
- **Data Validation.** Almost no sponsors were engaged in their own validation of the data used to score hospitals. Instead, they relied heavily on the audit functions of the organizations that originally collected the data (e.g., CMS, Joint Commission). When measures are generated from all-payer claims data, any validation that occurs typically consists of a review of the final performance scores by hospitals prior to payout and/or public reporting of results. Sponsors indicated that it was too labor intensive and expensive to validate data. While sponsors recognized that CMS and the Joint Commission may not have foolproof validation methods in place, many reasoned that “if it is good enough for the government or Joint Commission, it’s good enough for us.”

Payment Structure

- **Payout Method.** The sponsors with whom we spoke tended to use one of two performance-based rewards. About half (10/22) pay a lump-sum bonus, usually annually. The other half (9/22) pay the reward on a continuous basis (e.g., an ongoing “bump up” to per diem or DRG payments) and use past performance to determine the future year’s payment increase. The payment method selected was

usually determined by operational ease of implementation for the sponsor. A key consideration was budget planning related to how the payment was structured. For some, continuous, smaller payments spread out during the year were easier to plan for financially, rather than a one-time, larger bonus. For others, the situation was the reverse.

- **Reward Determination.** Most sponsors determined rewards based on improvements over time/meeting quality improvement targets (12/22) or relative performance (e.g., percentile ranking) (10/22). To a lesser extent, some used absolute thresholds (7/22), such as national percentile rankings from the prior year. Many of the sponsors to whom we spoke (8/22) used multiple forms of reward determinations in a single program. For example, for a given measure or set of measures, there might be a minimum threshold that a hospital must meet to even be considered for a reward. Then, for the hospital to receive the reward, it might have to demonstrate some pre-determined level of improvement. Some sponsors grouped hospitals by type when determining the reward in order to ensure “apples to apples” comparisons; for example, sponsors might compare and determine rewards for CAHs separately from other types of hospitals. Regardless of the way in which sponsors determined the reward, however, the majority measured performance using all-payer data but based the reward amount on the their own service volume in the particular plan products included in the P4P program (e.g., HMO, PPO, “all commercial”).
- **Weighting.** Most of the sponsors we spoke to (15/22) use differential weighting of their P4P metrics to determine a hospital’s performance score. Typically they use a differential point system grouped by domain. For example, a reward program may be based on 100 total points with 40 allocated to clinical measures, 30 to quality improvement activities, 20 to patient experience, and ten to structural measures. Given that many sponsors negotiate P4P with hospitals one by one, weighting is often tailored to individual hospitals as contracts come up for renewal.

- **Reward Funding Source(s).** Sponsors are funding their reward payments primarily through reallocation of existing resources (13/22). A few (7/22) are using premium increases and negotiated increases in hospital contracts as a way to fund the P4P program. Several program sponsors noted that compared to individual physicians, hospitals have greater bargaining strength, which makes it difficult for sponsors to take money off the table. Withholds were used only by five of the programs, largely because sponsors wanted to set a collaborative tone rather than a “take away” tone. Only two sponsors (2/22) mentioned savings from cost reductions as a funding mechanism; the others expressed uncertainty about where or even whether there would be cost savings from performance improvements to fund the program. One sponsor used “tiers” of participation, with higher levels requiring more measures but offering a larger “upside” in terms of the incentive payment.
- **Other/Non-Financial Incentives.** In addition to financial rewards, many sponsors include other, non-financial incentives as part of their incentive programs. Public reporting is a key non-financial motivator used (12/22), with results frequently posted on publicly available websites. Some sponsors (11/22) also use peer comparisons to motivate hospitals. Such comparisons tend to be included in reports shared with all hospitals participating in a given program. To set a collaborative, rather than punitive, tone, most sponsors present hospitals with blinded comparisons to peers; however, a few stated that they present unblinded data. Some sponsors also present performance scores grouped by hospital type (e.g., rural, academic medical center) and/or hospital size in an effort to make comparisons across similar types of institutions. Only a few sponsors use public recognition (5/22) (e.g., naming high performers on a public website) or tiering (2/22) (e.g., charging higher co-payments to consumers who go to lower-performing hospitals).

Public Reporting

- **General Comments.** Sponsors had mixed thoughts on public reporting. Some saw public reporting as a critical part of the incentive program, saying that it

captures the attention of all levels of hospital staff, as well as consumers. Others saw public reporting as creating a negative tone that is at cross-purposes with collaborative, quality improvement efforts between hospitals and program sponsors. Regardless of whether they were reporting specific data from their own programs or not, many sponsors provided a website link to the CMS Hospital Compare public report card that shows performance results for approximately 3,534 hospitals participating in the RHQDAPU program.

- **Reporters.** Sponsors that reported publicly (12/22) usually posted performance scores on websites intended for health plan members (i.e., usually password protected). Data were often presented in a simple format (such as stars displaying different levels of performance) rather than as specific numeric values, and summary scores were commonly used. Most sponsors reported doing minimal to no testing of report presentation with consumers and did not know whether consumers understood or found useful the information as presented.
- **Non-Reporters.** Sponsors not reporting data publicly tended to give two practical reasons for this. First, customized programs that are rolled out contract by contract do not permit comparisons, since not all hospitals have performance results or the same set of performance results. Second, some programs do not include all hospitals in a given area, again making comparisons difficult. Additionally, several sponsors underscored their desire to use their programs to work collaboratively with hospitals and thought that hospitals often viewed public reporting as a punitive strategy.

Hospital Assistance and Engagement

- **Engagement.** The majority of sponsors consulted with hospitals about overall program design (15/22), typically through in-person or telephone meetings during which they discussed ways to structure the P4P program. These sponsors strongly felt that such engagement was and continues to be critical to the success of their programs. In addition, they emphasized the importance of continuing to work collaboratively with hospitals as the program evolves. As part of their ongoing

interactions with hospitals and efforts to help them engage on quality improvement, many sponsors (21/23) provided performance reports to participating hospitals that often contained detailed information on individual metrics rather than just summary measures.

- **Assistance and Support.** Most sponsors (13/20) offered assistance to hospitals, usually in the form of (1) education about the program (e.g., goals, background information on metrics) (10/20) and/or (2) technical assistance (e.g., instructions on how to submit data electronically, clarifications on measure specifications) (7/20). Some sponsors also noted that they make themselves available for one-on-one, on-site consultations with program participants on an as-needed basis. Other techniques used to support hospital participation in P4P programs included sharing of best practices among participating hospitals (3/20) and the use of breakthrough collaboratives (2/20).

Program Evolution

- **Measures.** Looking forward, many sponsors (11/20) plan to expand and/or modify the measure sets they are currently using. They anticipated including more measures in one or more of the following areas:
 - **Expanded clinical processes:** Some sponsors noted that current performance is “topping out” on the measures that are part of existing measure sets. Consequently, they plan to expand the metrics they track to include areas that have received less attention to date, such as measures of surgical infection prevention and other new areas being added to RHQDAPU.
 - **Clinical outcomes:** Sponsors indicated that they want to shift the focus of their programs to include health outcomes, as opposed to solely using process measures, currently the primary focus of most programs.
 - **Patient experience:** Given CMS’ requirements to collect the Hospital CAHPS (HCAHPS) data starting in 2007 (with public reporting in 2008) as part of its RHQDAPU program, many sponsors foresee moving to this survey in the near future (www.hcahponline.org).

- **Resource use/efficiency:** There was significant interest in this area but a lack of sound metrics, according to sponsors. As reliable and valid measures are developed, sponsors plan to make this area a larger part of their programs.

Sponsors emphasized the need to ensure that programs are both “broad and deep” in terms of metrics. They noted, however, that achieving this goal is a challenge because they seek not to overburden hospitals with extensive data collection and submission requirements.

- **Other Modifications.** In addition to the changes to measures noted above, sponsors anticipated increasing selected aspects of their programs, such as
 - **The number of hospitals participating in the program:** Sponsors anticipate including more hospitals in their P4P programs as contracts come up for renewal. They noted that non-participating hospitals were beginning to feel pressure to sign up for the programs.
 - **The amount tied to performance:** Sponsors plan to increase the magnitude of the financial incentive that is tied to performance as they update their contracts with hospitals. In at least one case, a sponsor plans to begin tying payments to both inpatient and outpatient hospital services and was in the planning stages of developing outpatient hospital performance measures.
 - **The level of consumer engagement:** Increasingly, employers demand that the health systems with which they contract encourage consumer VBP through full disclosure of hospital performance. In response, some sponsors intend to incorporate “tiering” or other similar mechanisms into their programs as a way to encourage consumers to seek care from high-performing institutions.

Program Evaluation

Most sponsors to whom we spoke were not conducting formal evaluations of their hospital P4P programs (5/22). However, some noted anecdotal evidence of positive

program impact. For example, some said hospitals have improved their quality improvement infrastructure (e.g., dedicated quality improvement staff, regular quality improvement meetings) in response to P4P. Other sponsors reported seeing improved performance scores for participating hospitals. There was significant interest in tracking ROI, but there was also a lack of knowledge about how to do this and general difficulty estimating the costs associated with program development, implementation, and ongoing administration. For the most part, sponsors were not monitoring for potential unintended consequences of their hospital P4P programs, such as reduced attention and decreased quality of care in unmeasured areas. Sponsors did, however, recognize the need to do this, especially as P4P programs become more widespread and the amount of money tied to the financial incentive increases.

CRITICAL LESSONS LEARNED

We asked hospital P4P program sponsors to discuss the key lessons they have learned and the challenges they have faced in designing, implementing, and maintaining their hospital P4P programs. Their insights and recommendations based on their experiences are presented here for six key areas: overall design, measures, data collection, payment structure, hospital engagement, and public reporting.

Overall Design

Program sponsors said that **coordinating and aligning** their P4P programs with other P4P programs and hospital reporting requirements constituted one of the most important considerations in designing a successful program. They noted that hospitals are often overwhelmed with requests for disparate information from a variety of organizations, and that streamlining these requests is key to making program participation feasible. An article by Pham et al. (2006) noted that on average, hospitals face 3.3 reporting requirements from various entities which are typically not fully aligned and which create additional reporting burdens.

Sponsors underscored the importance of striving for a **simple program design** and avoiding a “black box” that is difficult to understand and explain. They also noted that simplicity helps to win over skeptics.

Although a number of sponsors had programs tailored to individual hospitals, they noted the administrative **advantages of a standardized program design and implementation**. They felt, however, that separate programs may be necessary for small, rural, and CAH hospitals to accommodate their distinct challenges related to performance scoring, such as small case volume, less-educated patient populations, different mixes of services and patients, and different pools of providers.

Regional experimentation would allow various models of program design to be tested. For national programs, such as those that might be sponsored by a large insurer or CMS, sponsors felt a regional approach would allow for experimentation, which they saw as important for two reasons. First, several noted that health care is local and there are variations in infrastructure and patterns of care across regions; so, clinical areas that may be problems in one area may not be an issue in another area. As such, quality improvement may be best carried out through local initiatives that take into account local practices and organizational structures. Second, the best way to design a P4P program is not yet known (or there may be more than one best way, depending on the characteristics of the market).

Finally, sponsors said it was important for them **to keep abreast of CMS' future actions to facilitate advance planning and allow them to align their own programs** with those of CMS.

Measures

Program sponsors said that based on their experience, the **use of evidence-based measures** that are standardized and have achieved a consensus base (i.e., are NQF and HQA endorsed) reduces hospital pushback. Sponsors noted that they would like to **expand measurement** beyond areas in which hospitals are already doing well to avoid the “teaching-to-the-test” phenomenon and to enable a more comprehensive assessment of performance. Areas suggested for additional measurement include:

- Outcomes
- Resource use/efficiency
- Transitions in care

- Medication management
- Patient experience related to safety
- Outpatient hospital services.

The shortage of evidence-based measures in some of these areas will slow efforts to expand measures.

Sponsors reported that they were **relying on CMS to take the lead nationally in both developing and maintaining measures**. Sponsors believe CMS is the most suitable entity to develop reliable and valid measures. They feel CMS' national presence and leverage will greatly facilitate adoption, leading to more programs using the same measures and thus decreasing the burden placed on hospitals to respond to the growing number of data requests and other new program requirements.

Data Collection

Sponsors reported that **minimizing the data collection burden was critical for hospital acceptance of P4P programs**. Suggested strategies for minimizing hospital burden included (1) alignment of measures and data collection across programs and (2) selection of a reasonable number of measures to include as part of the P4P program. Sponsors were unable to specify the precise number of measures that would be considered reasonable to include in a P4P program but stressed that there must be some limits. One suggestion was to retire measures as hospitals reach high-performance levels. However, this tactic raised concern that the areas no longer tracked would be ignored going forward. A suggestion for addressing this concern is to continue to track all measures but transition the high-performance metrics to threshold metrics after a specified amount of time. As such, a hospital would have to meet a certain level of performance on some metrics to be eligible for the financial incentive, but payouts would only be made based on performance on the current set of measures.

Payment Structure

The majority of P4P program sponsors advocated making the program as positive as possible. In this spirit, they suggested **focusing on collaboration and rewards** and

avoiding financial withholds, which are viewed as punitive. This sentiment is consistent with the principle of framing noted in our review of economic theories in Chapter 2. Program sponsors found a more positive, collaborative approach yields the best results in terms of quality improvement. Sponsors also recommended rewarding improvement in combination with top performance to keep all hospitals engaged. Many sponsors believe that it is important to “spread the wealth” by rewarding top performers and also incentivizing the lowest performers to improve. Some sponsors also suggested supporting or rewarding participation in regional continuous quality improvement (CQI) efforts to improve systems of care. One sponsor noted that quality improvement efforts may best be served by focusing on systems of care, rather than relying on the current “one off” model of tracking performance on individual measures. They recommended expanding the focus of hospital P4P programs to include rewards for participating in quality improvement efforts at the system level.

Hospital Engagement

Sponsors unanimously agreed that **interaction with hospitals** is critical to P4P program success. They stated it was important to engage and work collaboratively with hospitals “early and often” in all aspects of the program design and operation. Sponsors noted that this builds a sense of ownership and partnership among hospitals involved, which, in turn, helps increase acceptance of and support for the P4P program. Program sponsors also feel it is important to provide quality improvement guidance and support to hospitals as part of an ongoing feedback loop. Many sponsors viewed their role not only as the operational manager of the P4P program, but also as an important quality improvement resource for hospitals. They underscored that if performance improvement is truly a goal of the P4P program, mechanisms must be built in to provide assistance to hospitals that are trying to improve.

Public Reporting

Not all sponsors agreed that public reporting should be a part of P4P programs. While some viewed it as an important component that compliments the financial incentive, others saw it as contentious and detrimental to creating a collaborative

relationship with hospitals. Sponsors suggested that if public reporting were part of the program, **performance should be reported on a wide range of measures—such as clinical, patient experience, and resource use—in order to communicate a complete picture of health care to consumers.** Sponsors said that consumers do not make health care decisions in a vacuum and need additional information. As noted previously, many program sponsors provided links on their websites to the Hospital Compare website. Some sponsors suggested that the **Hospital Compare website should be simplified for ease of use by consumers.** Specific recommendations included (1) the use composite or summary measures within a service area or at the condition level, with information on individual measures available through “drilldown” capabilities to those wanting more-specific information and (2) increased consumer testing of the website to ensure that the information is understandable and useful.

IV. SUMMARY OF DISCUSSIONS WITH HOSPITALS, HOSPITAL ASSOCIATIONS, AND DATA VENDORS

RAND held discussions with a broad cross-section of hospitals, hospital associations, and hospital data vendors to learn about the experiences hospitals and their support vendors have had with the Medicare RHQDAPU P4R program, various private-sector P4P programs, and/or the CMS PHQID. Within the hospitals, we spoke to the Chief Executive Officer (CEO) or President; within the hospital associations, we spoke to the CEO and/or the lead policy and research staff dedicated to performance measurement and reporting. This activity was part of the larger environmental scan that RAND conducted to describe the current P4P and P4R landscape, in terms of how programs are designed and what lessons are being learned, in order to help inform the development of a VBP program for Medicare hospital services.

METHODOLOGY

RAND drew a purposive sample of hospitals from the universe of hospitals included in the RHQDAPU program and PHQID to obtain a range of perspectives. RAND selected hospitals from the national pool of hospitals that provide services to Medicare patients, reflecting an array of characteristics:

- Large and small
- Urban and rural
- Eligible to participate in the PHQID program but had declined
- Invited to participate in the CMS RHQDAPU program but had declined to submit data
- Submitted data and failed the data validation processes for RHQDAPU
- CAHs (which are not required to submit data under any current P4P or P4R initiatives) voluntarily submitting data under RHQDAPU.

We also spoke to a small number of hospitals exposed to a statewide private-sector P4P program, again selecting hospitals that were both large and small in terms of

number of beds. In addition, we held discussions with the major hospital associations and a small number of vendors that support the hospitals in their data submissions to comply with P4P and P4R reporting requirements.

Between October of 2006 and March of 2007, RAND held discussions with:

- Twenty-eight hospitals in five categories:
 - Twelve PHQID hospitals, seven of which volunteered to participate in the P4P demonstration and five that elected not to participate.
 - Five hospitals exposed to a private-sector P4P program.
 - Seven small and CAH hospitals that had submitted RHQDAPU data and were listed on Hospital Compare website.⁸
 - Three hospitals that failed data submission for RHQDAPU.
 - One PPS hospital that elected not to participate in the voluntary RHQDAPU program but was eligible to submit data.
- Seven major hospital associations:
 - The AHA, Federation of American Hospitals (FAH), AAMC, Voluntary Hospital Association (VHA), National Association of Children's Hospitals (NACH), National Rural Health Association (NRHA), and Catholic Health Association (CHA).
- Five hospital data vendors that support hospitals in submitting data for the RHQDAPU program.

To understand the unique characteristics and issues facing rural and CAHs hospitals that would affect their ability to fully participate in a VBP program, we held telephone discussions with seven hospitals (four rural, three CAHs), two government agencies with expertise in rural health issues, three state hospital associations located in states with a large number of rural providers and CAHs, one research center with

⁸ CAHs serve as a "proxy" for the likely experience of small hospitals. CAHs are not required to submit data under RHQDAPU, although some voluntarily do so. CAHs are not Subsection D hospitals and are excluded from the proposed Medicare VBP program, as outlined in the Deficit Reduction Act of 2005.

expertise in rural health issues, and three consultancies with extensive experience working with rural providers and CAHs. For the rural hospital assessment, the organizations with which we spoke were identified through two sources: (1) hospitals reporting on the Hospital Compare website and (2) experts in the rural health field who were interviewed and asked to identify key organizations and individuals with rural health expertise in the hospital setting.

Hospital Experiences with the Medicare RHQDAPU P4R Program

In our discussions with hospitals about the Medicare RHQDAPU program, which as of 2007 held 2 percent of a hospital's APU at risk for reporting, there was widespread sentiment that they would publicly report on these measures absent the RHQDAPU effort. The historical evidence suggests the contrary, however. Prior to tying reporting of performance measures to the APU, only a small number of hospitals (400 out of approximately 3,800 PPO hospitals) voluntarily reported performance data under the National Voluntary Hospital Reporting Initiative (NVHRI).

Helping the Hospitals Prepare for P4P. Most hospitals were fairly positive about their experience to date with the RHQDAPU program. Hospitals accepted the measures and agreed that the measures addressed important areas; they also felt that hospitals should be held accountable for these indicators of care. There was a unanimous belief among hospitals that P4P was inevitable, with a number observing that "*P4P is going to be a way of life in the future.*" Hospitals viewed the RHQDAPU program as a means to help them gain experience with data collection, submission, and validation and to make quality improvements before P4P starts. A number of hospitals commented, "*We want to be prepared.*" Hospitals indicated they were "OK" with shifting from RHQDAPU directly to P4P. Several hospitals expressed a desire to structure an incentive program with two payment components: a P4R component to allow all hospitals to receive funds to recoup their data collection costs and a P4P component to reward differential performance.

Challenges in Engaging Physicians. Hospitals stated that they were not currently financially incentivizing physicians on the performance measures for which they were

being held accountable. Most observed that physician engagement was challenging and that, moving forward, it would be important to align physician incentives to ensure the right behavior occurred. A majority of hospitals, particularly large hospitals, indicated they could not do much to influence physician behavior and struggled with ways to ensure compliance on the performance measures. Frequently, the hospital CEOs with whom we spoke noted that “*doctor’s don’t like to practice cookbook medicine*” and “*don’t like to be told what to do.*” The problem of physician engagement was compounded occasionally when the performance measures on which the hospital was being asked to report were not in synch with current evidence-based medicine (i.e., as the evidence changes, reporting requirements frequently lag). A number of hospitals expressed the need to change gain-sharing laws so that hospitals could structure financial incentives internally for physicians, and that this would allow physicians to see “*what’s in it for them.*”

P4R and P4P Are Generating the Engagement of Hospital Leadership.

Hospitals were in widespread agreement that the P4R program had caused important changes in their organizations, noting that it has resulted in a more proactive focus on quality improvement and attention on performance at all levels of the organization. A common sentiment expressed was, “*Without P4R, the quality improvement effort would have been smaller and slower.*” This sentiment was also indicated by hospitals exposed to P4P programs. Hospitals noted that their hospital boards and leadership were now much more focused on quality, and that typically there was a monthly review of progress on the performance indicators during the hospital board meetings, something that had not occurred prior to the P4R program. Hospitals stated that their leadership and boards frequently reviewed the Hospital Compare website to see where their hospital stood relative to others in their community and nationally; they also noted, “*We don’t want to be in the bottom quartile.*”

Hospital Experiences with Premier PHQID

Among Premier hospitals that were voluntarily participating in PHQID, we found broad agreement that their decision to participate reflected a desire to “*get in at the start*

to hopefully shape it” and a recognition that “*P4P is coming, and it is a way to gain experience.*” Some of the Premier hospitals that were eligible to participate but had declined indicated that they were shadowing the PHQID project by collecting the same data and investing in quality improvement activities. They felt that it was important for them to do so to be prepared when P4P became a reality for all hospitals. Interestingly, among the subset of PHQID hospitals with which we spoke, many stated that the possibility of financial incentive was a negligible factor in their decision to participate in the demonstration.

While P4P and P4R Are Leading to Behavior Change Among Hospitals, the ROI Is Unclear. PHQID participants stated that the P4P demonstration is driving improvements in the care they provide but that it has required them to allocate significant staff and resources to meet program requirements. This sentiment was echoed by hospitals in the RHQDAPU program. Hospitals felt that incentive payments (actual or potential) did not offset costs they were incurring to participate. Among the hospitals in the RHQDAPU program, a number noted that the cost of participation exceeded the 0.4 percent update they could receive for reporting, although they noted this might change when CMS increased the update factor tied to public reporting to 2 percent. One hospital commented that “*you’ve got to make it worth people’s time to do these things.*” Several hospitals expressed the importance of having CMS help hospitals see the link between doing better on the quality measures and a positive ROI—such as reductions in costs, lengths of stay, and readmissions.

The PHQID Incentive Payment Structure Creates Cliff Effects and Penalizes Hospitals That Perform Well. The Premier demonstration payment structure provided financial rewards only to hospitals that performed in the top two deciles of performance, based on a relative comparison of performance among hospital participants in each year of the program. Across the board, hospital participants expressed dislike for the design of the incentive structure. They noted it created a cliff effect (all or nothing payment) by rewarding hospitals at or above the 80th percentile performance and not rewarding any hospital that fell below this cut point—even when there was no statistical difference in their performance. Hospitals felt they were being penalized unfairly under a relative

scoring method when most hospitals were scoring at or close to 100 percent—which occurred for several of the performance indicators that had effectively topped out. One hospital cited, as an example, that for aspirin at arrival, the top four decile groups had effectively achieved 100 percent compliance with the performance measure, yet only the top two deciles were paid incentive dollars. Several hospitals questioned the value of having hospitals expend substantial resources chasing the top tail of the performance distribution when performance scores were so tightly clustered to the top right end of the distribution, expressing a belief that the relative benefit to patients was small and that it effectively was causing hospitals to divert resources that could be deployed to lower-performing areas that were not incentivized.

Over time, as providers make improvements, the compression of performance scores toward the top end of the performance distribution (i.e., the ceiling effect) will present challenges to P4P program sponsors that seek to differentiate providers on a relative performance basis. Common remarks by hospitals included: *“All should get the bonus if they achieve top levels of performance,”* and *“Rewarding the top two deciles is meaningless when the scores are so compressed at the top end.”* Other hospital comments reflected frustration with the relative performance incentive structure, for example: *“Every time we do better the bar gets higher”* (the hospital noted that it was effectively 100 percent on some measures and got no incentive dollars); *“Funding [is] only for [the] top 20 percent of hospitals, so 80 percent are spending dollars to improve and getting nothing in return.”*

Another reason why hospitals expressed a dislike for using a relative incentive structure is that this approach creates uncertainty about what level of performance is required to win. One hospital said, *“The performance bar is constantly shifting up, and it is an unknown to hospitals.”* Only at the close of the year, after the hospitals are arrayed in the rank order of their performance, does a hospital know what level of performance was required to hit the 80th percentile of performance to win. Hospitals and their professional associations expressed a strong preference for using an absolute performance threshold as the basis for determining whether a hospital would receive an incentive payment. The absolute threshold was viewed as a preferred approach to structuring an

incentive payment because it is “*predictable*,” “*allows a hospital to know in advance what performance target [it] would need to hit*,” and “*allows all who meet the threshold to secure the bonus*.”

Hospitals also expressed support for establishing a lower threshold in order to be able to qualify for an incentive. It was noted that this threshold should “*increase as more institutions met the minimum bar*.” Our discussions found lukewarm support among individual hospitals for paying for improvement: “*Hospitals should meet a minimum standard of excellence to be allowed to care for patients, so you don’t want to pay for improvement that occurs below this threshold*.” Hospital associations, however, strongly supported paying on the basis of improvement.

At This Stage, It Is Unclear Whether PHQID Is Causing Unintended Consequences. While most hospitals stated they did not believe the focus on a limited set of performance measures has led to unintended consequences, such as ignoring other clinical areas, they did say that limited staff and financial resources had caused them to focus heavily on what was being measured and rewarded—providing support to those who claim financial incentives promote teaching to the test. Most hospitals said they either did not know whether negative consequences were occurring or were not specifically tracking them. One hospital remarked, “*If anything, PHQID has increased activity and focus, and other quality improvement investments are being made, such as EHRs, CPOE, and use of intensivists, which will drive improvements across the board, not just on those things being incentivized*.”

Hospital associations commented that they were aware of one unintended consequence associated with the “antibiotic timing” measure for pneumonia (i.e., percentage of pneumonia patients who have received the first dose of antibiotics within four hours after hospital arrival), which is a measure for PHQID and RHQDAPU. In an effort to do well on this measure, some hospitals may have been over-prescribing antibiotics to patients who did not have pneumonia, giving them the antibiotic within the four-hour window before a diagnosis of pneumonia could be confirmed. There is concern that the overuse of antibiotics will increase resistance to the drug in the future. As a result, this measure has been pulled from the measure set and is being respecified.

Hospitals, while unable to cite specific examples, expressed concern that the relative incentive structure could lead to such unintended consequences as gaming of the data or hospitals chasing the very top end of the performance distribution by increasing a performance rate from 98 percent compliance to 100 percent with little to no clinical benefit, just to secure the incentive dollars. Several hospitals stated that because hospital margins are very thin, hospitals will chase the dollars.

The Reporting Burden Is Significant. Hospitals emphasized that the reporting burden for hospitals to comply with PHQID and/or RHQDAPU is significant given that data collection is still largely a manual exercise requiring chart abstraction. This was found to be true even in larger institutions having more information technology (IT) resources. EHRs and CPOE are not yet designed to provide data to populate measures such as those in PHQID, RHQDAPU, or other nationally endorsed measurement sets. Most EHRs capture relevant information in text fields; so even when EHRs are available, a text search must be done to determine if an event occurred. Hospitals universally felt that the data collection burden should be an important selection criterion for P4R and P4P programs. There was also consensus on the need to align measures and measure specifications to minimize data collection and reporting burdens—although it was also noted that the problem was less about alignment of specifications and more about getting the various stakeholders to align on what they want to hold providers accountable for. However, it is important to note that even though CMS allowed sampling of patient records to minimize the hospital reporting burden, many large hospitals reported that they did not use the sampling method, citing a need to have 100 percent of the cases to do their quarterly quality improvement work with doctors. These hospitals stated that the small number of sampled cases showed results that were too variable and did not provide a reliable source of information to give to doctors.

The Problem of Small Numbers Exists. The problem of only a small number of patients meeting the measure criteria was also raised, primarily by small hospitals, including rural hospitals and CAHs. Estimates of performance based on a small number of events (i.e., patients who receive appropriate processes of care) are not stable and vary substantially from period to period, making the task of separating out the “signal” (true

performance) from the “noise” (random variation) a challenging one. Hospitals with small numbers of patients cited challenges in interpreting and using results that showed large variation from period to period. Among the smaller hospitals, there was agreement that “*we should only be measured on what we actually do.*” Smaller hospitals thought that CMS should work to construct measures that more readily apply to the care they provide, such as transfers. When asked whether hospitals would support the use of composites to help with the small-numbers problem, there was no strong signal of support. However, this response may have stemmed from a lack of understanding about how the composites might be constructed. There was, in contrast, strong support for risk adjustment to ensure comparability across hospitals.

Measures of Outpatient Hospital Services Are Not Being Used at This Stage. None of the hospitals or hospital associations with which we spoke reported measures of outpatient hospital services being included in any P4P or P4R program to which they had been exposed, although several of the hospitals exposed to the private-sector P4P program noted that its sponsor was beginning to discuss with hospitals how such measures might be developed. There was general agreement that services—visits, procedures, and tests—provided in the outpatient hospital setting represented a substantial portion of care for which there currently is no accountability. Hospitals noted that outpatient hospital services have been a huge revenue growth area, and some reported seeing “*much utilization that seems questionable.*” While hospitals recognized that a large amount of care is delivered in this setting, they cited many challenges with developing performance measures and holding hospitals accountable given that data are less standardized on the outpatient side, and the mix of services delivered in this setting varies substantially across institutions.

Support for Having a Robust Data Validation Process Is Strong. Hospitals universally agreed that data validation is a critical feature of P4P programs. Hospitals were concerned about possible gaming, especially if there is “*too much money on the table and people start panicking,*” and believed that an audit function was needed to guard against this behavior. An attestation-type approach to data validation, such as the process the Leapfrog Group uses, was not viewed as sufficiently rigorous for situations in

which money is tied to performance. Hospitals expressed frustration with the substantial lag in the current validation processes—minimally six to nine months for PHQID, and 12 months before RHQDAPU results are posted on Hospital Compare—which slows down the process for getting feedback for CQI and public reporting. Hospitals stated a need for more-frequent updates—within three months of data submission—with comparisons to peers/benchmarks for use in quality improvement activities.

Transparency of Performance Results Is Viewed as a Positive. Hospitals indicated that they thought public reporting of performance on the hospital measures was good and that it has forced their doctors to pay attention and get engaged. One hospital noted that *“an external force doing measurement and reporting is our key lever (other than relational) with doctors to get them to change their behavior.”* Another noted that *“it says someone is watching.”* Only a few hospitals said that *“reporting hasn’t been a factor in driving behavior changes.”* Most hospitals stated that public reporting of their results compared with those of their peers has garnered the attention of their hospital boards and stimulated investment in quality improvement, noting that *“no one wants to be at the bottom of the list.”* Hospitals preferred that if the RHQDAPU program evolved into a P4P program, a pilot or dry-run period of data collection occur prior to public reporting and payouts.

Although hospital leadership and physicians are internally paying attention to the comparative results, hospitals seemed to be unsure about whether consumers really use the information. Many hospitals thought that the CMS Hospital Compare website should be simplified to make it easier for consumers to use. There was no consensus among hospitals about what would be the appropriate comparison group of hospitals or whether one is even needed for public reporting of results. One hospital stated: *“The consuming public needs to know if a hospital will provide adequate care, so the focus should be on whether the hospital hits a threshold target [rather than] comparing one hospital to another.”* Another hospital thought that regional comparisons would be helpful to consumers *“who won’t be traveling to other states for care.”*

Hospitals Are Encountering Certain Challenges. Many hospitals stated that it was difficult to get physicians to change their behavior regarding actions called for in the

performance measures and that they felt as though they were serving as a go-between for CMS and the physician. Hospitals thought they had little leverage to affect physician behavior other than having good relationships. The current prohibition on gain sharing precludes hospitals from structuring provider financial incentives within their organizations, thus hindering their ability to motivate physicians to engage in the P4R and P4P programs (“*A slow process until MD incentives are also aligned.*” “*Physician and hospital P4P programs shouldn’t be separate*”).

Having to work with and win over doctors was a common theme in our discussions with hospitals (“*Doctors don’t like hospitals telling them what to do.*” “*Doctor’s don’t like to practice cookbook medicine*”).

Some hospitals reported that in response to the challenges of engaging physicians, they had developed solutions to force behavior change, such as creating admission and discharge forms that prompt doctors for information and/or to do required things, creating standing clinical protocols, and structuring clinical treatment paths differently. Hospitals appeared to be developing unique interventions rather than implementing a one-size-fits-all approach to driving improvements in care. It was noted that making P4P and quality improvement work requires a lot of coordination across departments.

Hospitals also noted that involvement in these programs requires a lot of staff resources for data collection and validation and quality improvement. Several remarked that to succeed in these programs, a hospital needs infrastructure and multidisciplinary teams, two things not available in smaller community hospitals and hospitals in rural areas, where there are no dedicated staff to perform these functions and “*the CEO is often wearing several hats within the organization.*”

On the subject of data submissions and the validation process, hospitals expressed broad appreciation for the important “assistance” role that Premier played as a “go-to” entity. The feeling was that Premier provided an important support function related to a hospital’s ability to comply with the program requirements.

Hospitals cited struggles faced because of ongoing changes in the evidence without corresponding changes in what hospitals are held accountable for. They reported that their physicians had made changes in practice consistent with new evidence, even

though the hospitals were still required to comply with measure specifications that reflected out-of-date evidence. Hospitals urged that P4R and P4P program sponsors work to address, in a timely manner, changes in the evidence and what hospitals are held accountable for.

Advice Offered by Hospitals Regarding P4P Program Designs

The key recommendations that hospitals had for anyone considering designing and implementing a P4P program were as follows:

- Reward everyone that does well. Avoid setting up a reward structure that only pays out top deciles when measures are compressed at top end.
- Do not pay based on improvement or, if you do, set a minimum threshold of performance and only pay for improvement above that minimum.
- Provide regular performance feedback for quality improvement purposes. Monthly feedback is most helpful to those on the front line of the organization who are trying to make change. Hospitals expressed a desire to get feedback that shows a particular hospital's percentile score with the raw score and comparison benchmarks (in real time).
- Focus on selecting measures for core areas where expenditures and patient volume are high.
- Provide support and technical assistance, especially to small hospitals and CAHs, since participation requirements can be significant.
- Involve hospitals directly in planning and implementation (*"They know what really happens in a hospital."*). Some hospitals felt that national associations (e.g., AHA, FHA) were adequate representatives for hospitals' concerns, but small, rural, and CAH hospitals felt that state hospital associations from states with a substantial rural provider population might better represent their particular issues.
- For small hospitals, limit what is measured to what they do—do not hold them accountable for things they do not do. Allow smaller hospitals to choose from

a smaller number of clinical conditions in order to make program participation more manageable for them.

- Allow hospitals to directly incentivize their physicians and be sure to align physician measures and incentives with hospital measures and incentives. Change restrictions on gain sharing so that hospitals can provide financial incentives to their doctors.
- Focus hospital measurement on things the hospital has control over (e.g., infection rates, turnaround time on tests and procedures).
- Coordinate and align with other programs/hospital reporting requirements.
- Use evidence-based measures that are standardized and consensus based to reduce hospital pushback (e.g., that are endorsed by NQF and HQA). Educate physicians about measures being evidence based in order to get buy-in, potentially working through such professional journals as the *Journal of the American Medical Association (JAMA)* and the *New England Journal of Medicine (NEJM)*.
- Expand measurement beyond the CMS RHQDAPU areas in which hospitals are already doing well to include measures of outcomes, cost/efficiency, transitions in care, medication management, patient experience related to safety, and outpatient hospital services.
- Pilot new measures prior to payout and reporting.
- Minimize hospital burden by selecting a “reasonable” number of measures to track and by aligning with other hospital reporting requirements.
- Support risk adjustment to ensure comparability and to minimize possible unintended consequences of risk selection.
- Use all-payer data to score hospitals to avoid the small-numbers problem.
- Validate the data to prevent gaming.
- Consider the important role that data vendors can play by supporting hospitals with data submissions and validation.

- Create a standardized program but consider regional approaches to allow experimentation (“*[The] right design isn’t known today, and we need to learn as we go*”).

V. SUMMARY OF FINDINGS FROM ENVIRONMENTAL SCAN

Mounting cost pressures and substantial deficits in the quality of care within the U.S. health care system have led policy makers to consider options for system reform to drive improvements. Value-based purchasing is one reform option being examined and tested by payers in the public and private sectors, and it includes both financial (e.g., P4P) and non-financial (e.g., transparency of performance scores) incentives designed to change the behavior of providers.

The Deficit Reduction Act of 2005 (Public Law 109-171, Section 5001(b)) created a statutory mandate for the Secretary to develop a VBP plan for Medicare hospital services commencing FY 2009. This mandate was delegated to the CMS Hospital VBP Workgroup. This environmental scan was conducted to inform the development of the VBP plan for Medicare hospital services. Our scan comprised a review of the literature and key informant discussions with a wide array of individuals who could provide a picture of the current state-of-the-art in hospital pay for performance, including 27 program sponsors, 28 hospitals, 7 hospital associations, 5 data support vendors, and a number of individuals with expertise in rural hospital issues. As part of our discussions, we also examined the experiences of hospitals participating in the Medicare RHQDAPU pay-for-reporting program.

Among the key findings of this review is that hospital P4P has been implemented by more than 40 sponsors, in some cases for more than three-to-five years. Little empirical evidence has emerged, however, from these initiatives to gauge the impact of hospital P4P in meeting a program sponsor's objectives. This is primarily a function of the absence of formal evaluation occurring in most P4P programs and the challenges of conducting evaluation in real-world applications that lack comparison groups to assess the impact of the P4P intervention. The strongest evidence on the impact of hospital P4P to date has been shown through the Premier evaluation of the Premier Hospital Quality Incentive Demonstration (PHQID) and the Lindenauer study of the impact of PHQID relative to the Medicare pay-for-reporting program. These studies suggest the additional effects of P4P are somewhat modest relative to public reporting and other quality

interventions that are occurring simultaneously. The literature suggests, however, that multifaceted interventions will be most effective at producing sustained improvements in patient care (Grol et al 2002; Grol and Grimshaw 2003). Drawing from the theoretical literature on the use of incentives, it appears that incentives can be effective in changing behavior, and that how the incentives are structured will determine the type and magnitude of the behavioral response.

In our hospital and P4P program sponsor discussions, there was an expressed desire to allow experimentation to create models where learning could occur, which could help inform design structures. The discussants anticipate that the results of P4P and specific design options may differ as a function of the varying structure of local health care markets.

Given that P4P is a newly emerging reform tool and that little information is currently available about the impact of P4P or the influence of various design structures on P4P outcomes, P4P programs should incorporate evaluation and ongoing monitoring into their design as a means of building a knowledge base. The collection and broad dissemination of this type of information will be critical to future efforts to construct P4P programs so that they can meet their programmatic objectives. Funding will be necessary to support program evaluation, and the evaluation work needs to be sustained over multiple years to fully assess impact and monitor for unintended consequences.

The key design and implementation lessons that emerged from our discussions with program sponsors, hospitals, and data vendors included:

- **Measures**—Hospitals expressed concerns about growing data collection and reporting burdens across the various P4P programs and reporting initiatives being developed by an array of sponsors, whose efforts are not fully aligned. Hospitals expressed a strong desire for measures to be aligned, for reporting efforts to be coordinated, and for use of evidence-based standardized measures to minimize physician pushback. While P4P program sponsors desire to expand the number and types of performance measures to ensure a more comprehensive picture of hospital quality, hospitals stated a desire for a more limited set of measures on which they could focus quality improvement efforts. Given the limitations in the

number and type of measures currently available for use in pay for performance and public reporting, resources will be required to support additional measure development and testing as well as the development of methods to create composites.

- **Payment structures**—There is consensus among hospitals that payment structures should use absolute thresholds and reward all good performers, rather than providing incentives on a relative-performance basis, for example only to the top 10 or 20 percent of hospitals participating in a P4P program. This was seen as critical when the measures of performance used have scores that “top out,” reflecting little meaningful difference in the performance across hospitals, as has occurred for several process-of-care measures (e.g. for care of acute myocardial infarction). Another approach that could avoid the payment issues associated with topped out measures is to use the appropriate care composite, which reduces the ceiling effect, as the basis for payment rather than individual measures. Programs sponsors felt strongly that performance improvement as well as attainment of specific benchmarks should be included as a component of the payment structure, at least in the early years of the program, in order to engage all hospitals in the P4P program. Hospitals also noted the difficulty of getting physicians to change their behavior absent aligned incentives on the physician side, and called for program sponsors to create parallel physician incentives focused on inpatient care for the same conditions used in hospital programs. Physicians would also be more likely to support P4P programs that did not place an additional burden on physicians in terms of data collection or documentation.
- **Data infrastructure**—Current validation efforts are weak, and program sponsors and hospitals acknowledged the need to strengthen validation as more money is put at risk in P4P programs. Hospitals indicated the need for technical support to comply with P4P program requirements, citing the important role played by QIOs and data vendors in this regard. Health information systems require modification moving forward to capture the data elements used to produce performance measures, and absent this investment, hospitals will continue to have to extract

- **Public reporting**—Hospitals indicate they do pay attention to how their institution looks publicly and that public reporting has forced their boards to more closely monitor quality and provide resources for quality improvement. Both program sponsors and hospitals cited a need for simplification of performance information presented on consumer websites, such as the CMS Hospital Compare website, to facilitate consumer understanding and use of the information.
- **Engagement strategies**—Program sponsors noted the importance of engaging hospitals in the planning and execution of P4P programs to encourage a more collaborative versus payer-driven approach to implementing this payment reform. Engagement strategies included involving providers in the measures selection process and program design more broadly, and in ongoing planning as the program evolves over time.

Our discussions also uncovered a number of program implementation challenges that merit consideration during program design and implementation. One challenge that affects a sizeable number of hospitals is the problem of having only a small number of events or cases to report for one or more measures; a small number of events to score leads to unstable estimates of performance to use in performance-based incentive payments. While this is a more acute problem for small and rural hospitals with a small number of patients per year, the problem can also occur for medium- and large-size hospitals depending on their service mix, details of measure specifications, and the use of sampling during data collection. Use of all-payer data, collecting data over extended periods of time, use of composite measures, and identifying measures relevant to smaller providers are approaches that can help to mitigate the small numbers problem.

The data collection burden, which affects how many measures a P4P program can reasonably require a hospital to collect and report, creates challenges for efforts to comprehensively assess the performance of hospitals. The more comprehensive the

measure set used, the greater the burden on hospitals, given existing information technologies. Current information systems are not equipped to capture and easily retrieve the clinical information used to create performance measures, nor are they structured to enable routine monitoring of quality of care. Until health information systems are upgraded to capture this information, program sponsors will be constrained in the number and breadth of measures they can expect hospitals to collect and report. P4P programs are also challenged with an acute need to ensure the integrity of the data used to score hospitals and make differential payments, which requires resources for data validation. Allocating sufficient resources to validation work is critical for program credibility, and today only limited resources are being used for data validation within P4P programs. Most hospitals stated that the current level of validation is insufficient, given the potential to shift large sums of money within the system.

P4P programs have the potential to drive system improvements. The success of these programs in meeting improvement goals will be affected by their design, implementation, and allocating sufficient resources to engage in the necessary day-to-day operations, program monitoring and impact evaluating, and ongoing modification. Given the limited knowledge base, it is critical that P4P programs include evaluation in their design to generate the knowledge to support smart program design and efficient use of resources.

Hospitals understand that P4P is likely to be part of their future and generally seem supportive of the concept. They face a number of challenges to their ability to successfully participate in these programs, including lack of physician engagement, inadequate information infrastructure that necessitates the manual collection of data from charts, and potentially conflicting signals from various organizations measuring hospital performance. These implementation challenges should be carefully considered in the design of any hospital P4P program.

APPENDIX A: DESIGN ISSUES EXPLORED AS PART OF THE ENVIRONMENTAL SCAN

This appendix lists the complete set of design issues that were identified by ASPE and CMS as being of interest for exploring through the environmental scan work.

OVERVIEW

- What are the goals of existing pay-for-performance (P4P) programs and demonstrations in the hospital setting?
- What should Medicare's goals be for P4P in the hospital setting?
- What is the most effective way to transition from pay-for-reporting (P4R) to P4P? What assistance should CMS offer to providers in the implementation of P4P?
- What are the lessons learned by organizations with P4P and P4R programs in practice or participating in demonstrations? How do these programs demonstrate that such programs improve both quality of care and the efficiency of health care delivery?
- How are hospitals included in the design and implementation of P4P and P4R programs?
- What mechanisms are used to communicate with hospitals about the program, and what lessons have been learned about engaging providers?
- Is participation voluntary or mandated?
- If participation is voluntary, what inducements for participation are being used, and how effective are they at encouraging participation?
- What mechanisms are put in place to monitor for unintended consequences, both for clinical care and data quality/gaming?
- What is the return on investment (ROI) for P4P and how should it be calculated?
- Should Medicare P4P be based on all adult patients, as hospital public reporting is currently structured, or only patients eligible for Medicare?
- How can Medicare recognize the unique challenges faced by rural and critical access hospitals (e.g., small patient volumes, limited staff resources) in the design and implementation of P4P?
- What choices in measure selection, payment methodology, and coordination and communication can best support state and private purchasers engaged in P4P while also reducing the burden on providers?
- How should other types of hospitals, beyond subsection (d) hospitals, be integrated into P4P in the future?
- How should outpatient hospital services be integrated in the future?

MEASURES

- Define the set of services provided in the outpatient hospital setting to identify what could be measured and potentially rewarded.
- What measures of performance (e.g., clinical effectiveness, efficiency, patient experience, care coordination/transitions) are currently being used for both inpatient and outpatient hospital care in practice, demonstrations, etc.?
- How do measures of performance used in practice and in demonstrations differ from and align with recommendations concerning P4P and public reporting from such sources as the IOM, MedPAC, JCAHO, AHRQ?
- What are the benefits of and barriers to the use of different types of measures in a Medicare hospital P4P plan, including process (e.g., current HCAHPS measures), outcomes (e.g., mortality), patient survey (e.g., HCAHPS), administrative (e.g., AHRQ PSIs), and structural measures (e.g., the structural Leapfrog Group measures recommended in DRA Section 5001(a) as the “starter set” of hospital measures as defined in the IOM report “Performance Measurement: Accelerating Improvement”)?
- What criteria are existing hospital P4P programs and public reporting activities using to select measures? What are the salient differences, if any, in the criteria used in P4P and public reporting programs? What standards are used in these programs to assess the extent to which a measure is associated with improved processes or outcomes of care?
- How are programs addressing methodological issues around P4P, including level of aggregation of measures (i.e., composite scoring, weighting), establishment of benchmarks versus thresholds versus targets, risk adjustment, and opportunities for gaming?
- How should the burden of data collection factor in as a criterion for selection of measures or topics?
- How do existing P4P or public reporting systems assess the accuracy of the data they receive? Do they validate a provider’s general ability to provide accurate data, or do they audit or otherwise certify the accuracy of specific data transmissions? How does the quality assurance strategy affect the selection of particular measures or topics?
- What are the process, criteria, and timeline for modification/maintenance of P4P measures, including adding, changing, retiring, rotating, or deleting measures in a P4P environment and giving hospitals and other stakeholders adequate notice of measure modifications? How can maximum flexibility be built into the process to allow for quick response to new evidence in order to modify both the individual measures and the associated payment incentives?
- What approaches are more or less successful for involving stakeholders in the identification, maintenance, and future expansion of P4P and public reporting measure sets (e.g., HCAHPS, JCAHO, NQF, hospital systems, specialty societies, individual hospitals)?

- How should “new” measures be introduced in P4P? Is public reporting (with or without an incentive) a necessary precursor or transition step for all measures used in a P4P approach?
- If public reporting is not a prerequisite, should the process include a dry run (a period during which hospitals gain experience with reporting the measure prior to its use for P4P)?
- What are the longer-term needs for new areas of measure development? Issues to be considered include how to create measures that address both the inpatient and the outpatient hospital setting; patient safety, overuse, medication use, appropriateness, readmissions, complications, efficiency, equity, coordination of care across settings; and other identified “measure gaps” pertinent to P4P.
- What processes need to be established to assure alignment and convergence of standardized measures of hospital performance across the hospital industry?
- What can other P4P arrangements suggest about how a Medicare hospital P4P and public reporting plan could align with and promote similar objectives in other settings (e.g., physician practice, post-acute settings)?

DATA

- What data collection, data management, reporting infrastructure, and data outreach were required to implement existing P4P programs (e.g., sampling methodology, storage capacity)?
- How do current P4P programs address data collection issues, including sampling and minimizing burden, such as
 - The alignment process with JCAHO (including warehouse edits and abstraction tool skip patterns) so that there continues to be a single abstraction of quality data for hospitals to receive their accreditation and CMS quality data payment.
 - Modifying reporting deadlines to better facilitate continuous quality data submission for concurrent abstraction hospitals.
 - Evaluating sampling requirements to ensure reliable data while minimizing burden.
 - The use of composite measures
- How can the lag from date of service to public reporting be minimized?
- What plans are there for receiving data directly from electronic health records (EHRs)?
- How should the data be safeguarded?
- How are data security and privacy issues balanced with restricted access to clinical warehouse data for analysis and modeling?
- What roles are currently served by and envisioned for various tools, including CART (the Quality Improvement Organization’s [QIO’s] Clinical Abstracting and Reporting Tool) and QnetExchange (the QIO data portal)?
- What access is required/envisioned for QIO data?

- How do the confidentiality requirements associated with data reported to CMS QIOs affect the uses to which hospital data reported for P4P can be put? Can the DHHS/CMS share this data with other payers, and can the data of other payers be integrated into the data set or calculation of rates? What entity controls access to and use of the data?
- How should the validation for P4P be structured to maximize effectiveness while minimizing costs? How should validation methodology assure abstraction reliability, adherence to sampling methodology, and submission completeness? How will measure-specific characteristics, such as relative variability in measure rates by hospital, be incorporated into validation sample sizes?

PAYMENT MECHANISMS

- What types of incentives, financial or non-financial, currently exist or are under consideration (e.g., financial, public recognition, public reporting, confidential peer comparisons, systems support)?
- How effective are different types of incentives at influencing provider behavior?
- Should incentives be based on thresholds, improvement, and/or high achievement? If based on relative performance, what characteristics define the relative peer group for comparison?
- What types of hospital providers are eligible for the rewards?
- What are the methods of delivery of financial rewards (e.g., differential, lump sum)?
- What is the timeframe for reward delivery (e.g., annual, quarterly)?
- For financial awards applied to service payments, are they applied to all services, measured services, and/or related services?
- What is the source of funding?
- What levels (fixed dollar, percent of payments) and types (negative versus positive) of financial incentives have been used or are under consideration? For those they have been used, what is the relationship between the levels and types of incentives and provider behavior?
- If applicable, how have operational issues (e.g., claims processing) impacted P4P programs?
- What mechanisms currently foster program integrity? What program integrity issues have occurred? What are potential program integrity issues initially and over time?

PUBLIC REPORTING

- What hospital-quality public reporting systems are currently available, and what is the evidence of their use and impact? What features (in terms of both design of the report and publicity associated with the report) of those systems are associated with greater impact?
- How do private and state purchasers address the policy issues with which CMS has struggled, and what lessons can be learned about these issues:

- How should reports simplify data and make them easier to use? Should reports be created by rank-ordering by performance? And should they use symbols, bars, or numerical rates? What are the most effective ways of conveying confidence intervals and data uncertainty to the general public, health care providers, and hospital quality improvement staff? Are there some hospital quality measures that have utility only for quality monitoring and improvement, only through financial incentives, or only through public reporting? Do some publicly reported measures (e.g., outcome measures) have significant effect on hospital quality improvement activity without being tied to financial incentives, while others (e.g., process measures) have less effect unless they are tied to financial incentives?
- If hospitals are penalized for not improving above a threshold (along the lines of the theoretical penalty that the Premier demo will be imposing on underperforming hospitals), should CMS publicly report and highlight the fact of the penalty?
- Should CMS report improvement in quality, performance above certain benchmarks in quality, or relative ranking among peers on quality measures—or all of the above? What evidence do we have that one type of public reports has greater impact than other types?
- What cost measures are most important to display for different audiences? Who does or would use cost data for decisionmaking, and how can such data be more effectively displayed for that audience?
- How do we best display and explain efficiency measures? How do different audiences interpret these measures? Given evidence that consumers may misinterpret such measures (e.g., longer lengths of stay mean “this hospital cares more about their patients than other hospitals do”), what are the most effective ways of explaining such measures to the public?
- How can we display efficiency measures and absolute costs together most effectively?
- How do token financial incentives to patients impact their understanding and weighting of quality and efficiency measures? For example, would co-pay discounts based on quality scores increase the awareness and credibility of quality measures among patients?
- How can CMS reach all of its customers, from beneficiaries to providers to researchers? How does CMS meet the needs of different audiences and the different uses to which they put the data? How does CMS provide transparency and access to data while ensuring adequate protections for privacy and not overwhelming CMS’ data management capabilities?
- How should CMS and DHHS portray the hospital P4P program to the public to engage the interest and support of consumers and the general public? What reactions from the provider community can be anticipated and planned for?

CROSS-CUTTING THEMES

- How can hospital P4P be integrated into the Medicare purchasing environment?
- What is the evidence of the impact of P4P programs on changing provider performance?
- What features are necessary for the sustainability of programs?
- What steps have successful hospital P4P programs used to partner with and engage other stakeholders? What implications do those efforts have for the Listening Sessions in this contract, discussions with the HQA, and other collaborations with partner organizations?
- How does hospital P4P improve Medicare's position as a value-based purchaser?
- How do we incorporate patient safety themes in our approach to P4P?
- How do we integrate P4P and the use of EHRs?
- How can hospital P4P enhance the evidence base for quality improvement and not interfere with innovation?
- How do we minimize burden for providers and CMS?
- What actions are needed to translate P4P programs into the goals of quality improvement and efficiency?

OUTPATIENT SETTING

- What is the scope of outpatient hospital services, and which of these services could be initially targeted for performance measurement and potential reward?
- Are there programs currently under way to align reimbursement with value-based purchasing (VBP) in the outpatient hospital setting?
- Are there measures currently available that could be applied and/or modified in the context of developing a Medicare outpatient hospital P4P program in the near term? If yes, what are they?
- What are the gaps in available measures, and what strategy would be required to fill in these gaps to create a robust set of measures for use longer term in a P4P program in the outpatient hospital setting?
- Are there unique issues of data infrastructure, payment methodology, and/or public reporting in the outpatient setting compared with the inpatient setting? If so, what are these issues, how do they impact the development of a P4P program for the outpatient setting, and how can they be resolved?
- What are the challenges of CMS-stakeholder collaboration in the outpatient hospital setting compared with the inpatient setting? How should they be addressed?

APPENDIX B: SUMMARY OF PAY-FOR-PERFORMANCE DESIGN PRINCIPLES

This appendix builds on the summary of P4P design principles and recommendations presented in Chapter 1 of this report. Here we present and summarize the P4P design principles established by 26 organizations representing a variety of stakeholders, including purchasers, health care providers, policy organizations, accreditation organizations, health plans, and consumers. Table B.1 displays the P4P design principles for each of the 26 organizations. Table B.2 tallies the principles and recommendations across recommendations.

Table B.1. P4P Principles and Recommendations from Stakeholders

P4P Design Principles/Recommendations	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
<i>Medicare Specific</i>												
P4P in Medicare should be implemented using a phased approach that varies by setting, reward amount, and measures			X									
Medicare should fund the program by setting aside a small share of payments in a budget-neutral approach		X										
Congress should derive initial funding (3–5 years) largely from existing funds by creating provider-specific pools from a reduction in base Medicare funding for each class of providers			X									

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
P4P Design Principles/Recommendations												
A consolidated pool should be formed from which all providers are rewarded when measures allowing for shared accountability are developed			X									
A Medicare P4P program must not be budget neutral or subject to artificial Medicare payment volume controls												
Medicare incentives should be financed with a new, dedicated stream of funding											X	
Medicare should distribute all payments that are set aside to providers achieving quality criteria		X										
Medicare should establish a process for continual evolution of measures		X										
A Medicare P4P program should be phased in gradually starting with reporting on structural measures and moving to enhanced payment based on evidence-based clinical measures												
Medicare should initially reward care that is of high clinical quality, patient centered, and efficient			X									

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Medicare should consider expanding the proportion of payment based on performance over time							X					
Medicare should initially reward both providers who improve performance significantly and providers who achieve high performance			X									
Medicare should offer incentives to providers for the submission of performance data, and these data should be publicly available in ways that are meaningful and understandable to consumers			X									
The program should be designed such that virtually all Medicare providers submit performance measures for public reporting and participate in P4P as soon as possible			X									
CMS should design the program to include components that promote, recognize, and reward care coordination across providers			X									
CMS should implement a monitoring and evaluation system for the program			X									

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
A Medicare P4P program must be pilot tested across settings and specialties and phased in over an appropriate period												
Incentives should eventually apply to all Medicare providers, including FFS and Medicare Advantage											X	
Metrics												
Programs should utilize accepted, evidence-based measures	X	X	X	X		X	X		X	X		
Measures should be pilot tested, validated, and vetted through a process that includes public comment and phased in				X								
The measurement set should include measures of clinical quality, patient experience, and infrastructure						X						
Measures need to be prioritized to address areas that are important to patients (such as those that prevent deaths, complications, and discomfort), as well as those that improve satisfaction, outcomes, and experience with care				X								

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Incentives should be based on existing measures and should emphasize clinical effectiveness											X	
Measures adopted should be developed by nationally recognized measurement organizations and recommended by consensus-building organizations				X			X					
Metrics should be high volume, high gravity, and strongly evidence based; have a gap between current and ideal practice and good prospects for quality improvement; and have measurement reliability, validity, and feasibility									X			
Program designers should include a sufficient number of metrics across a spectrum of health promotion activities to provide a balanced view of performance									X	X		
The development, validation, selection, and refinement of measures should be a transparent process that has broad consensus among stakeholders				X					X			

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
The development and selection of metrics should include participation by the patient community as well as by physicians and other providers									X			
Distinct standards should be developed to evaluate performance relative to the most vulnerable patients: frail elderly and patients with chronic, debilitating, or life-threatening illness												
Process measures, such as those used by the HQA, should be used												
Process or intermediate outcome measures are preferred unless robust, well-accepted methods of risk adjustment can be applied to outcome measures												
The focus should be on structure and process measures until evidence-based outcome measures are developed												
Structure, process, and outcome measures should be utilized							X		X			
Outcome measures are the highest priority because of their central importance to patients				X								

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Outcome measures must be subject to the best available risk adjustment for patient demographics, severity of illness, and co-morbidities						X						
Metrics should be selected from the following domains: patient centeredness, effectiveness, safety, and efficiency				X					X			
Metrics should include efficiency measures							X					
Efficiency measures should only be used when both the cost and the quality of a particular treatment are considered				X								
When measuring quality, focus on misuse and overuse as well as underuse							X					
Provide positive provider incentives for adoption and utilization of IT	X		X	X			X	X	X	X		

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Programs implemented by either the public or the private sector involving HIT should incentivize only those applications and systems that are standards based to enable interoperability and connectivity, and should address the transmission of data to the point of care	X							X				
Programs should move from an individual disease management approach to cross-cutting measures						X						
Metrics should be stable over time												
Metrics should be kept current to reflect changes in clinical practice												
Each measure should remain in the set for at least three years but should be evaluated annually to adjust weighting and specifications as necessary						X						
Local measures should closely follow national metrics as long as they are reportable from electronic data sets										X		
To prevent physician de-selection of patients, programs should use risk adjustment methods	X	X							X			

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
To ensure fairness, performance data must be fully adjusted for sample size and case mix composition, including age/sex distribution, severity of illness, number of co-morbid conditions, patient compliance, and other features of the practice or patient population that may influence the results				X					X			
The responsibility for developing, maintaining, and revising measures must reside with the specialty organizations representing the providers in whose scope of practice the measure resides												
Measures should be selected to ensure that all hospitals have an opportunity to participate and succeed												
Measures should be uniform across all providers of imaging services and across payers												
Measures used for P4P should meet higher standards than measures designed for other purposes				X								
Programs should reward accreditation or have an equivalent mechanism that rewards continuous attention to all clinical and support systems and processes	X											

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	ICAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
<i>Data Collection, Reporting, Feedback</i>												
Data should be collected without undue burden on providers	X	X							X			X
IT tools should be used whenever possible for data acquisition												
Programs must reimburse physicians for any administrative burden for collecting and reporting data												
Allow physicians to review, comment on, and appeal results prior to payment or reporting												
Programs should have a mix of financial and non-financial incentives (e.g., public reporting)	X				X		X		X			
Physician performance data must remain confidential and not subject to discovery in legal proceedings												
Public reporting/recognition is essential			X	X	X	X				X		
Performance data feedback should provide comparisons to peers and benchmarks												
Educational feedback should be provided to providers	X			X								

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Physicians must have timely access to the comparative performance database to which they have contributed data, including the ability to benchmark their data												
Programs should favor the use of clinical data over claims-based data												
Programs should use administrative data and data from medical records												
Measures should be feasible to collect using administrative data						X						
Performance data should be audited	X			X								
Programs should use an auditable data collection method tested for reliability and accuracy				X								
Metric assessments and payments should be made as frequently as possible to better align rewards with performance	X								X			
Hospital bonuses should be calculated every 6 months based on activity in the previous 6 months					X							
Data reporting must not violate patient privacy												

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
P4P assessments should be done with sample sizes (denominators) large enough to produce statistically significant results									X			
Incentives												
Reimbursement must be aligned with the practice of high-quality, safe health care	X	X							X			
Incentives should be based on rewards, not penalties									X			
Hospital rewards should be based on 50/50 sharing of savings from improvement					X							
Programs should reward providers based on improving care and exceeding benchmarks		X		X	X				X	X	X	
A sliding scale of rewards should be established to allow for recognition of gradations in quality	X											
Programs must not reward physicians/hospitals based on rankings that compare them with other physicians/hospitals in the program				X								
Payments must exceed the total cost of implementation, including data collection and reporting costs												

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Incentives must be significant enough to drive desired behaviors and support CQI				X		X						X
Mechanisms must be established to allow performance awards for physician behaviors in hospital settings that produce cost savings												
General Program Design												
Funding for P4P initiatives should come from additional resources, not a redistribution of resources												
Top performers should be eligible for market share through patient shift					X							
Programs should offer voluntary physician participation												
Physicians and/or hospitals should be involved in the program design				X								X
Programs should encourage strong alignment between practitioner and provider goals	X											
Providers must have the opportunity to understand the measures, analytical methodology, and use of data for public reporting before participating in a P4P program				X								

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	JCAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Most providers should be able to demonstrate improved performance		X							X			
When selecting areas of clinical focus/measures, programs should strongly consider consistency with national and regional efforts	X								X	X		
Programs should be consolidated across employers and health plans to make the bonuses meaningful and the program more manageable for physicians						X						
Programs should be designed to include practices of all sizes and levels of IT capabilities												
Physician organizations rather than individual physicians should be the accountable entity in P4P programs									X			
Initiatives need to be flexible enough to assess performance at both the individual and the group level												
Accountability must occur at the individual physician level				X								
Payments should recognize systemic drivers of quality in units broader than individual provider organizations and practitioner groups	X											

	HEALTH CARE ORGANIZATIONS										HEALTH PLANS	
P4P Design Principles/Recommendations	ICAHO	MedPAC	IOM	NQF Conference	Leapfrog	IHA	Natl. Business Group on Health	eHealth Initiative Fdn.	Healthways/ Johns Hopkins	Pacific Business Group on Health	Alliance of Comm. Health Plans	AHIP
Programs should be designed to acknowledge the united approach (team approaches, integration of services, continuity of care)	X			X								
Fair and accurate models for attributing care when multiple physicians treat the same patient must be implemented												
The results of P4P programs should not be used against physicians in health plan credentialing, licensure, or certification												
The data or the program should be adjusted for patient non-compliance									X			
Programs should incorporate periodic objective evaluations of impacts and make adjustments	X			X								
As P4P methodologies develop, patient access to quality care should be facilitated and not impeded by reduced reimbursement												
Programs should invest in sub-threshold performers who are committed to improvement	X			X								

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
<i>Medicare Specific</i>														
P4P in Medicare should be implemented using a phased approach that varies by setting, amount of reward, and measures														
Medicare should fund the program by setting aside a small share of payments in a budget-neutral approach														
Congress should derive initial funding (3–5 years) largely from existing funds by creating provider-specific pools from a reduction in base Medicare funding for each class of providers														
A consolidated pool should be formed from which all providers are rewarded when measures that allow for shared accountability are developed														
A Medicare P4P program must not be budget neutral or subject to artificial Medicare payment volume controls						X				X				
Medicare incentives should be financed with a new, dedicated stream of funding														
Medicare should distribute all payments that are set aside to providers achieving quality criteria														

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
Medicare should establish a process for continual evolution of measures														
A Medicare P4P program should be phased in gradually starting with reporting on structural measures and moving to enhanced payment based on evidence-based clinical measures			X											
Medicare should initially reward care that is of high clinical quality, patient centered, and efficient														
Medicare should consider expanding the proportion of payment based on performance over time														
Medicare should initially reward both providers who improve performance significantly and providers who achieve high performance														
Medicare should offer incentives to providers for the submission of performance data, and these data should be publicly available in ways that are meaningful and understandable to consumers														

	PHYSICIAN GROUPS										HOSPITAL GROUPS		PATIENT GROUPS	
P4P Design Principles/Recommendations	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
The program should be designed such that virtually all Medicare providers submit performance measures for public reporting and participate in P4P as soon as possible														
CMS should design the program to include components that promote, recognize, and reward care coordination across providers														
CMS should implement a monitoring and evaluation system for the program														
A Medicare P4P program must be pilot tested across settings and specialties and phased in over an appropriate period										X				
Incentives should eventually apply to all Medicare providers, including FFS and Medicare Advantage														
Metrics														
Programs should utilize accepted, evidence-based measures	X	X	X	X	X	X		X	X	X	X	X		
Measures should be pilot tested, validated, and vetted through a process that includes public comment and phased-in														

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
The measurement set should include measures of clinical quality, patient experience, and infrastructure														
Measures need to be prioritized to address areas that are important to patients (such as those that prevent deaths, complications, and discomfort), as well as those that improve satisfaction, outcomes, and experience with care														
Incentives should be based on existing measures and should emphasize clinical effectiveness														
Measures adopted should be developed by nationally recognized measurement organizations and recommended by consensus-building organizations	X									X				
Metrics should be high volume, high gravity, and strongly evidence based; have a gap between current and ideal practice and good prospects for quality improvement; and have measurement reliability, validity, and feasibility														
Program designers should include a sufficient number of metrics across a spectrum of health promotion activities to provide a balanced view of performance														

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
The development, validation, selection, and refinement of measures should be a transparent process that has broad consensus among stakeholders			X											
The development and selection of metrics should include participation by the patient community as well as by physicians and other providers														X
Distinct standards should be developed to evaluate performance relative to the most-vulnerable patients: frail elderly and patients with chronic, debilitating, or life-threatening illness														X
Process measures, such as those used by the HQA, should be used											X			
Process or intermediate outcome measures are preferred unless robust, well-accepted methods of risk adjustment can be applied to outcome measures								X						
The focus should be on structure and process measures until evidence-based outcome measures are developed										X				
Structure, process, and outcome measures should be utilized			X		X		X							

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
Outcome measures are the highest priority because of their central importance to patients														
Outcome measures must be subject to the best available risk adjustment for patient demographics, severity of illness, and co-morbidities	X				X				X					
Metrics should be selected from the following domains: patient centeredness, effectiveness, safety, and efficiency													X	
Metrics should include efficiency measures					X									
Efficiency measures should only be used when both the cost and the quality of a particular treatment are considered			X								X			
When measuring quality, focus on misuse and overuse as well as underuse					X									
Provide positive provider incentives for adoption and utilization of IT	X	X		X	X	X	X			X				X

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
Programs implemented by either the public or the private sector involving HIT should incentivize only those applications and systems that are standards based to enable interoperability and connectivity, and should address the transmission of data to the point of care														
Programs should move from an individual disease management approach to cross-cutting measures														
Metrics should be stable over time	X		X											
Metrics should be kept current to reflect changes in clinical practice								X		X	X	X		
Each measure should remain in the set for at least three years, but should be evaluated annually to adjust weighting and specifications as necessary														
Local measures should closely follow national metrics as long as they are reportable from electronic data sets														
To prevent physician de-selection of patients, programs should use risk adjustment methods	X	X	X		X	X				X				

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
P4P Design Principles/Recommendations	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
To ensure fairness, performance data must be fully adjusted for sample size and case mix composition, including age/sex distribution, severity of illness, number of co-morbid conditions, patient compliance, and other features of the practice or patient population that may influence the results	X	X	X			X				X		X		
The responsibility for developing, maintaining, and revising measures must reside with the specialty organizations representing the providers in whose scope of practice the measure resides								X	X					
Measures should be selected to ensure that all hospitals have an opportunity to participate and succeed											X			
Measures should be uniform across all providers of imaging services and across payers									X					
Measures used for P4P should meet higher standards than measures designed for other purposes														

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
Programs should reward accreditation or have an equivalent mechanism that rewards continuous attention to all clinical and support systems and processes														
Data Collection, Reporting, Feedback														
Data should be collected without undue burden on providers	X	X	X					X	X		X		X	X
IT tools should be used whenever possible for data acquisition			X											
Programs must reimburse physicians for any administrative burden for collecting and reporting data	X	X				X				X				
Allow physicians to review, comment on, and appeal results prior to payment or reporting	X		X	X		X				X		X		
Programs should have a mix of financial and non-financial incentives (e.g., public reporting)							X							
Physician performance data must remain confidential and not subject to discovery in legal proceedings										X				
Public reporting/recognition is essential														X

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
Performance data feedback should provide comparisons to peers and benchmarks		X												
Educational feedback should be provided to providers	X		X											
Physicians must have timely access to the comparative performance database to which they have contributed data, including the ability to benchmark their data								X						
Programs should favor the use of clinical data over claims-based data					X									
Programs should use administrative data and data from medical records	X													
Measures should be feasible to collect using administrative data														
Performance data should be audited		X			X					X				
Programs should use an auditable data collection method that is tested for reliability and accuracy														
Metric assessments and payments should be made as frequently as possible to better align rewards with performance		X												
Hospital bonuses should be calculated every 6 months based on activity in the previous 6 months.														

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
Data reporting must not violate patient privacy	X		X											
P4P assessments should be done with sample sizes (denominators) large enough to produce statistically significant results	X	X	X											
Incentives														
Align reimbursement with the practice of high quality, safe health care	X	X	X		X	X		X		X	X	X		X
Incentives should be based on rewards, not penalties	X	X	X		X					X	X		X	X
Hospital rewards should be based on a 50/50 sharing of savings from improvement														
Programs should reward providers based on improving care and exceeding benchmarks	X	X	X				X			X	X			
A sliding scale of rewards should be established to allow for recognition of gradations in quality														
Programs must not reward physicians/hospitals based on rankings that compare them with other physicians/hospitals in the program	X										X		X	
Payments must exceed the total cost of implementation, including data collection and reporting costs								X						

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
P4P Design Principles/Recommendations	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
Incentives must be significant enough to drive desired behaviors and support continuous quality improvement			X	X				X			X			
Mechanisms must be established to allow performance awards for physician behaviors in hospital settings that produce cost savings										X				
General Program Design														
Funding for P4P initiatives should come from additional resources, not a redistribution of resources				X							X			
Top performers should be eligible for market share through patient shift														
Programs should offer voluntary physician participation	X	X				X				X				
Physicians and/or hospitals should be involved in the program design	X	X	X			X				X				
Programs should encourage strong alignment between practitioner and provider goals				X							X		X	
Providers must have the opportunity to understand the measures and analytical methodology and use of data for public reporting before participating in a P4P program								X						

P4P Design Principles/Recommendations	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
Most providers should be able to demonstrate improved performance-focus on areas needing improvement	X													
When selecting areas of clinical focus/measures, programs should strongly consider consistency with national and regional efforts								X			X			
Programs should be consolidated across employers and health plans to make the bonuses meaningful and the program more manageable for physicians		X												
Programs should be designed to include practices of all sizes and levels of IT capabilities	X	X												
Physician organizations rather than individual physicians should be the accountable entity in PFP programs	X				X									
Initiatives need to be flexible enough to assess performance at both the individual and the group level												X		
Accountability must occur at the individual physician level														
Payments should recognize systemic drivers of quality in units broader than individual provider organizations and practitioner groups														

	PHYSICIAN GROUPS										HOSPITAL GROUPS			PATIENT GROUPS
	AMA	AAFP	ACP	Mass. Medical Society	ACC Fdn.	MGMA	AMGA	American Society of Anesth.	ACR	Surgical Specialty Orgs*	AHA	AAMC	Conn. Hospital Assoc.	National Patient Advocacy Foundation
P4P Design Principles/Recommendations														
Programs should be designed to acknowledge the united approach (team approaches, integration of services, continuity of care)	X				X		X	X						
Fair and accurate models for attributing care when multiple physicians treat the same patient must be implemented												X		
The results of P4P programs should not be used against physicians in health plan credentialing, licensure, or certification	X	X	X											
The data or the program should be adjusted for patient non-compliance	X	X	X											
Programs should incorporate periodic objective evaluations of impacts and make adjustments			X	X	X									
As P4P methodologies develop, patient access to quality care should be facilitated and not impeded by reduced reimbursement														X
Programs should invest in sub-threshold performers who are committed to improvement														

*American Academy of Ophthalmology, American Academy of Otolaryngology, American Association of Neurological Surgeons, American Association of Orthopedic Surgeons, American College of Surgeons, American Society of Cataract and Refractive Surgery, American Society of Plastic Surgeons, American Urological Association, Congress of Neurological Surgeons, Society for Vascular Surgery, Society of American Gastrointestinal and Endoscopic Surgeons, Society of Gynecologic Oncologists, Society of Surgical Oncology, and The Society of Thoracic Surgeons.

Table B.2. Summary of P4P Design Principles and Recommendations

Principles and Recommendations	Number of Orgs Supporting (n=26)
<p>Metrics for P4P Programs:</p> <ul style="list-style-type: none"> • Evidence based 19 • Risk adjust to mitigate impact of patient non-compliance, avoid physician de-selection of patients, and ensure fairness 11 • Comprehensive in scope 5 • The development, validation, and selection of measures should include all stakeholders 5 • Recommended by consensus-building organizations 4 • Keep current to reflect changes in clinical practice 4 • Focus on clinical areas needing improvement 3 • Stable over time 2 • Focus on misuse and overuse as well as underuse 2 • Developed, maintained, and revised by specialty organizations 2 • Include the patient community in the selection process 1 • Should meet higher standards than metrics used for other purposes 1 • Select such that all hospitals may participate 1 • Evaluate performance relative to the most-vulnerable patients (frail elderly and patients with chronic, debilitating, or life-threatening illness) 1 • Move from an individual disease management approach to cross-cutting measures 1 • Reward accreditation or similar process 8 <p><i>Process measures</i></p> <ul style="list-style-type: none"> • Should be included in P4P programs 11 <p><i>Outcome measures</i></p> <ul style="list-style-type: none"> • Risk adjust 7 • Should be included in P4P programs 2 • Are not sufficiently developed 1 • Give the highest priority 15 <p><i>Structural measures</i></p> <ul style="list-style-type: none"> • Should be included in P4P programs 15 • Should include HIT adoption and utilization measures 2 • Should require HIT systems to be standards based and provide data at the point of care 5 <p><i>Efficiency measures</i></p> <ul style="list-style-type: none"> • Should be included in P4P programs 3 • Use only when both the cost and the quality of a treatment are considered 5 	

<p><i>Patient experience measures</i></p> <ul style="list-style-type: none"> • Should be included in P4P programs 	
<p>Data Collection, Reporting, Feedback:</p> <ul style="list-style-type: none"> • Avoid undue burden on providers • Include public reporting • Allow providers to review, comment on, and appeal results prior to payment or reporting • Audit performance data • Sample sizes must be large enough to produce statistically significant results • Assess performance and make payments as frequently as possible to align rewards and performance • Data reporting must not violate patient privacy • Give providers feedback with benchmarking data • Favor the use of clinical data over administrative data • Use both clinical data and administrative data • Choose measures that are feasible to collect using administrative data • Performance data must remain confidential and not subject to discovery in legal proceedings 	<p>12</p> <p>8</p> <p>6</p> <p>5</p> <p>4</p> <p>3</p> <p>3</p> <p>2</p> <p>1</p> <p>1</p> <p>1</p> <p>1</p>
<p>Incentives:</p> <ul style="list-style-type: none"> • Reward high-quality, safe health care • Base rewards on improving care and exceeding benchmarks • Base incentives on rewards, not penalties • Provide incentives significant enough to drive desired behaviors and support improvement • Payment must exceed the cost of implementation (collecting and reporting data) • Do not base incentives on provider ranking • Establish gain-sharing mechanisms • Base hospital rewards on a 50/50 shared savings with payers • Top performers should be eligible for increased market share through patient shift (steering/tiering) • Establish a sliding scale of rewards to recognize gradations in quality 	<p>13</p> <p>12</p> <p>9</p> <p>7</p> <p>5</p> <p>4</p> <p>1</p> <p>1</p> <p>1</p> <p>1</p>
<p>General Program Design:</p> <ul style="list-style-type: none"> • Providers should be involved in the program design • Acknowledge team approaches, integration of services, care coordination • Consider consistency with national and regional efforts • Incorporate periodic evaluation of impacts and make adjustments • Encourage strong alignment of physicians and hospitals • Programs should be voluntary • Give providers an opportunity to understand the measures, methodology, and reporting requirements before they participate in P4P • Invest in sub-threshold performers who are committed to improvement • Funding should come from additional resources, not a redistribution of resources 	<p>7</p> <p>6</p> <p>5</p> <p>5</p> <p>4</p> <p>2</p> <p>2</p> <p>2</p> <p>2</p> <p>2</p>

<ul style="list-style-type: none"> • Include providers of all sizes and levels of IT capabilities • Consolidate programs across employers and health plans • Design to mitigate the impact of patient non-compliance • Patient access should not be impeded by reduced reimbursement • Implement fair and accurate attribution rules for providers 	<p>2 1 1 1</p>
Medicare-Specific Recommendations:	(n=7)
<ul style="list-style-type: none"> • Program should not be budget neutral • Program should be budget neutral • Use a phased approach • Reward care that is of high clinical quality, patient centered, and efficient • Reward improvement and high performance • Require public reporting • Reward care coordination • Include a monitoring and evaluation system • Provide incentives for FFS and Medicare Advantage providers • Establish a process for continual evolution of measures • Distribute all funds that are set aside to providers achieving quality criteria • Consider expanding the proportion of payment based on performance over time • Pilot test across settings 	<p>3 2 2 1 1 1 1 1 1 1 1 1</p>

APPENDIX C: INPATIENT HOSPITAL MEASURES

Measure	Organizations Collecting/Utilizing Measures															NQF En-dorsed	IOM Domain
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GWTC ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRO ¹³	CDC ¹⁴		
AMI:																	
Aspirin at Arrival	X	X	X	X	X				X		X					X	Effective
Aspirin at Discharge	X	X	X	X	X				X		X					X	Effective
ACEI or ARB for LVSD	X	X	X	X	X				X	X	X					X	Effective
Smoking Cessation Advice/Counseling	X	X	X	X	X				X	X	X					X	Effective
Beta Blocker at Discharge	X	X	X	X	X				X	X						X	Effective
Beta Blocker at Arrival	X	X	X	X	X				X		X					X	Effective
Mean Time to Thrombolysis/Fibrinolysis	X	X															Effective
Thrombolytic/Fibrinolytic Received Within 30 Minutes of Arrival	X	X	X	X	X				X							X	Effective
Mean Time to PCI	X	X															Effective
PCI Within 120 Minutes of	X	X	X	X	X				X							X	Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Arrival																	
Reperfusion Within 90 Minutes of Arrival											X					X	Effective
Inpatient Mortality	X				X									X		X	Safe
30-Day Mortality (Medicare patients)		X	X	X												X	Safe
PCI Volume								X								X	Safe
PCI Mortality								X								X	Safe
Heart Failure:																	
Discharge Instructions	X	X	X	X	X				X	X						X	Effective, Patient Ctrd.
LVF Assessment	X	X	X	X	X				X	X						X	Effective
ACEI or ARB for LVSD	X	X	X	X	X				X	X						X	Effective
Smoking Cessation Advice	X	X	X	X	X				X	X						X	Effective, Patient Ctrd.

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/IG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Beta Blocker at Discharge										X							Effective
Inpatient Mortality														X			Safe
30-Day Mortality (Medicare Patients)		X	X	X												X	Safe
30-Day All-Cause Risk Standardized Readmission		X															Effective
Pneumonia:																	
Oxygenation Assessment	X	X	X	X	X											X	Effective
Pneumococcal Vaccination	X	X	X	X	X											X	Effective
Blood Cultures Within 24 Hours Prior to or After Arrival—ICU Patients	X	X															Effective
Blood Culture Before First Antibiotic Received	X	X	X	X	X											X	Effective
Smoking Cessation Advice	X	X	X	X	X											X	Effective, Patient Ctrd.
Antibiotic Timing (Median)	X																Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Initial Antibiotic Received Within 8 Hours of Arrival	X																Effective
Initial Antibiotic Received Within 4 Hours of Arrival	X	X	X	X	X											X	Effective
Initial Antibiotic Selection for CAP in Immunocompetent Patient		X	X	X												X	Effective
Initial Antibiotic Selection for CAP in Immunocompetent—ICU Patient	X				X											X	Effective
Initial Antibiotic Selection for CAP in Immunocompetent—Non-ICU Patient	X				X											X	Effective
Influenza Vaccination	X	X	X	X	X											X	Effective
Inpatient Mortality													X			X	Effective
30-Day Pneumonia Mortality		X	X													X	Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Pregnancy and Related Conditions:																	
VBAC	X															X	Effective
Inpatient Neonatal Mortality	X																Safe
3rd or 4th Degree Laceration	X															X	Safe
Birth Trauma-Injury to Neonate														X			Safe
Obstetric Trauma—Vaginal Delivery with Instrument														X			Safe
Obstetric Trauma—Vaginal Delivery Without Instrument														X			Safe
Obstetric Trauma—Cesarean Delivery														X			Safe
Surgical Care Improvement/ Surgical Infection Prevention:																	
Prophylactic Antibiotic Received Within 1 Hour Prior to Incision—Overall Rate	X	X	X	X		X					X					X	Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Prophylactic Antibiotic Received Within 1 Hour Prior to Incision—Hip Arthroplasty	X				X	X										X	Effective
Prophylactic Antibiotic Received Within 1 Hour Prior to Incision—Knee Arthroplasty	X				X	X										X	Effective
Prophylactic Antibiotic Received Within 1 Hour Prior to Incision—Colon Surgery	X					X										X	Effective
Prophylactic Antibiotic Received Within 1 Hour Prior to Incision—Hysterectomy	X					X										X	Effective
Prophylactic Antibiotic Received Within 1 Hour Prior to Incision—Vascular Surgery	X					X										X	Effective
Prophylactic Antibiotic Selection for Surgical Patients—Overall Rate	X	X	X	X		X					X					X	Effective
Prophylactic Antibiotic Selection—Hip Arthroplasty	X				X	X										X	Effective
Prophylactic Antibiotic	X				X	X										X	Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Selection—Knee Arthroplasty																	
Prophylactic Antibiotic Selection—Colon Surgery	X					X										X	Effective
Prophylactic Antibiotic Selection—Hysterectomy	X					X										X	Effective
Prophylactic Antibiotic Selection—Vascular Surgery	X					X										X	Effective
Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time—Overall Rate	X	X	X	X		X					X					X	Effective
Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time—Hip Arthroplasty	X				X	X										X	Effective
Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time—Knee Arthroplasty	X				X	X										X	Effective
Prophylactic Antibiotics Discontinued Within 24 Hours	X					X										X	Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
After Surgery End Time— Colon Surgery																	
Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time— Hysterectomy	X					X										X	Effective
Prophylactic Antibiotics Discontinued Within 24 Hours After Surgery End Time— Vascular Surgery	X					X										X	Effective
Recommended VTE Prophylaxis Ordered	X	X	X	X	X ¹⁵	X										X	Effective
Recommended VTE Prophylaxis Received Within 24 Hours Prior to or After Surgery	X	X	X	X	X ¹⁵	X										X	Effective
Cardiac Surgery Patients with Controlled 6 AM Post- Operative Serum Glucose	X	X				X				X							Effective
Surgery Patients with Appropriate Hair Removal						X				X							Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/G ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Colorectal Surgery Patients with Immediate Post-Operative Normothermia						X					X						Effective
Surgery Patients on Beta Blockers Prior to Admission Who Received a Beta Blocker During the Perioperative Period						X											Effective
Mortality Within 30 Days of Surgery	X	X				X											Effective, Safe
ICU:																	
Ventilator-Associated Pneumonia Prevention—Patient Positioning	X																Effective
Ventilator Bundle										X							Effective
Stress Ulcer Disease Prophylaxis	X																Effective
DVT Prophylaxis	X									X							Effective
Central Line Associated Blood Stream Infection	X													X	X		Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/G ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Central Line Bundle Compliance											X						Effective
Central Line Insertion Adherence Practices															X		Effective
Urinary Catheter–Associated Urinary Tract Infection																	Effective
Severe Sepsis/Septic Shock: Activate Drotrecogin Alfa											X						Effective
Severe Sepsis/Septic Shock: Low Dose Glucocorticoid											X						Effective
Severe Sepsis/Septic Shock: Blood Cultures Collected											X						Effective
Severe Sepsis: Central Venous Oxygen Saturation																	Effective
Severe Sepsis: Central Venous Pressure											X						Effective
Severe Sepsis/Septic Shock: Glucose Values											X						Effective
Severe Sepsis/Septic Shock: Median Inspiratory Plateau											X						Effective

Measure	Organizations Collecting/Utilizing Measures															NQF En-dorsed	IOM Domain
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/G ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³	CDC ¹⁴		
Pressures																	
Severe Sepsis/Septic Shock: Median Time to Broad Spectrum Antibiotic											X						Effective
Blood Cultures Performed Within 24 Hours Prior to or After Arrival for Patients Transferred to ICU	X	X															Effective
ICU Length of Stay	X																Effective, Efficient
Hospital Mortality for ICU Patients	X																Effective, Safe
Stroke:																	
Deep Vein Thrombosis (DVT) Prophylaxis (Ischemic)	X									X							Effective
DVT Prophylaxis for Intracranial Hemorrhage																	Effective
Discharged on Antithrombotics (Ischemic, TIA)	X																Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Discharged on Antiplatelet Therapy										X							Effective
Patients with Atrial Fibrillation Receiving Anticoagulation Therapy (Ischemic)	X									X							Effective
Tissue Plasminogen Activator (t-PA) Considered (Ischemic, TIA)	X																Effective
Antithrombotic Medication Within 48 Hours of Hospitalization (Ischemic, TIA)	X									X							Effective
Lipid Profile (Ischemic, TIA)	X																Effective
Screen for Dysphagia (Ischemic, Hemorrhagic, TIA)	X									X							Effective
Stroke Education (Ischemic, Hemorrhagic, TIA)	X																Effective, Patient Ctrd.
Smoking Cessation (Ischemic, Hemorrhagic, TIA)	X									X							Effective, Patient Ctrd.

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Plan for Rehabilitation Considered (Ischemic, Hemorrhagic)	X																Effective, Patient Ctrd.
Lipids Measured									X								Effective
Blood Pressure Management									X								Effective
Non-Invasive Cartoid Imaging Reports																	Effective
CT or MRI Reports									X								Effective
Avoidance of Intravenous Heparin																	Effective
Acute Stroke In-Hospital Mortality Rates													X				Safe
Cardiac Surgery:																	
Participation in a Systematic Database for Cardiac Surgery (STS)						X										X	Effective
Surgical Volume—Isolated CABG		X														X	Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Surgical Volume—Valve Surgery		X														X	Safe
Surgical Volume—CABG + Valve Surgery		X														X	Safe
Prophylactic Antibiotic Within 1 Hour Prior to Surgical Incision—CABG					X	X										X	Effective
Prophylactic Antibiotic Within 1 Hour Prior to Surgical Incision—Other Cardiac Surgery	X					X										X	Effective
Selection of Antibiotic—CABG	X				X	X										X	Effective
Selection of Antibiotic—Other Cardiac Surgery	X					X										X	Effective
Prophylactic Antibiotics Discontinued Within 48 Hours After Surgery End Time—CABG	X				X	X											
Prophylactic Antibiotics Discontinued Within 48 Hours After Surgery End Time—	X					X										X	Effective

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GWIG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Other Cardiac Surgery																	
Pre-Operative Beta Blockade—CABG						X										X	Effective
Use of Internal Mammary Artery—CABG		X			X											X	Effective
Aspirin at Discharge—CABG					X						X						
Post-Operative Hemorrhage or Hematoma—CABG					X								X				
Post-Operative Physiologic and Metabolic Derangement					X								X				
Prolonged Intubation—CABG						X										X	Effective
Deep Sternal Wound Infection Rate—CABG						X										X	Safe
Stroke/Cerebrovascular Accident—CABG						X										X	Safe
Post-Operative Renal Insufficiency—CABG						X										X	Safe
Surgical Re-exploration—CABG						X										X	Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Anti-Platelet Medication at Discharge—CABG						X										X	Effective
Beta Blockade at Discharge—CABG						X										X	Effective
Anti-Lipid lipid Treatment at Discharge—CABG						X										X	Effective
Risk-Adjusted Inpatient Operative Mortality—CABG																X	Safe
Risk-Adjusted Operative Mortality—CABG						X										X	Safe
Risk-Adjusted Operative Mortality for AVR						X										X	Safe
Risk-Adjusted Operative Mortality for MVR						X										X	Safe
Risk-Adjusted Operative Mortality for MVR + CABG						X										X	Safe
Risk-Adjusted Operative Mortality for AVR + CABG						X										X	Safe
CABG Inpatient Morality Rate					X									X			Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
PTCA Mortality Rate														X			Safe
Carotid Endarterectomy Mortality Rate														X			Safe
Bilateral Cardiac Catheterization Rate														X			Safe
Surgery (Non-Cardiac):																	
Complications of Anesthesia														X			Safe
Failure to Rescue	X													X			Safe
Foreign Body Left in During Procedure														X			Safe
Post-Operative Hip Fracture														X			Safe
Post-Operative Hemorrhage or Hematoma					X ¹⁵									X			Safe
Post-Operative Physiologic and Metabolic Derangements					X ¹⁵									X			Safe
Readmissions 30 Days Post-Discharge					X ¹⁵												Safe, Efficient

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Post-Operative Respiratory Failure														X			Safe
Post-Operative Pulmonary Embolism or Deep Vein Thrombosis														X			Safe
Post-Operative Sepsis														X			Safe
Post-Operative Wound Dehiscence														X			Safe
Hip Replacement Mortality Rate														X			Safe
Esophageal Resection Mortality Rate														X			Safe
Pancreatic Resection Mortality Rate														X			Safe
AAA Repair Mortality Rate														X			Safe
Incidental Appendectomy Among Elderly Rate														X			Safe
Laparoscopic Cholecystectomy Rate														X			Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/G ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Surgical Site Infection													X		X		Safe
Surgical Wound Disruption													X				Safe
Other Surgical Wound Occurrence													X				Safe
Pneumonia Post-Surgery													X				Safe
Unplanned Intubation													X				Safe
Pulmonary Embolism													X				Safe
On Ventilator > 48 Hours													X				Safe
Other Respiratory Occurrences													X				Safe
Progressive Renal Insufficiency													X				Safe
Acute Renal Failure													X				Safe
Urinary Tract Infection													X				Safe
Other Urinary Tract Occurrence													X				Safe
CVA/Stroke													X				Safe
Coma													X				Safe
Peripheral Nerve Injury													X				Safe
Other CNS Occurrence													X				Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Cardiac Arrest Requiring CPR													X				Safe
Myocardial Infarction													X				Safe
Other Cardiac Occurrence													X				Safe
Bleeding Requiring > 4 Units PRBC/Whole Blood Transfusions Within the First 72 Hours Post-Operative													X				Safe
Surgical Graft/Prosthesis/Flap Failure													X				Safe
DVT/Thrombophlebitis													X				Safe
Systemic Sepsis (SIRS)													X				Safe
Systemic Sepsis (Sepsis)													X				Safe
Systemic Sepsis (Septic Shock)													X				Safe
Other Occurrences													X				Safe
Return to the Operating Room Within 30 Days of Surgery													X				Safe
Death Within 30 Days of Surgery													X				Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Death Greater Than 30 Days After Surgery in Acute Care													X				Safe
Venous Thromboembolism (VTE):																	
Risk Assessment/Prophylaxis Within 24 Hours of Admission	X																Effective
Risk Assessment/Prophylaxis Within 24 Hours of Transfer to ICU	X																Effective
Documentation of Inferior Vena Cava Filter Indication	X																Effective
VTE Patients with Overlap Therapy	X																Effective
VTE Patients Receiving Heparin-Platelet Count Monitoring	X																Effective
VTE Discharge Instructions	X																Effective
Incidence of Potentially Preventable Hospital-Acquired	X																Effective, Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
VTE																	
VTE 30-Day Hospital Readmission (ICSI)																	Effective, Efficient
Cancer:																	
Patients with Early Stage Breast Cancer Who Have Evaluation of the Axilla																	Effective
College of American Pathologists Breast Cancer Protocol																	Effective
Colon Cancer: Surgical Resection Includes at Least 12 Nodes																	Effective
College of American Pathologists Colon and Rectum Protocol																	Effective
Completeness of Pathologic Reporting																	Effective

Measure	Organizations Collecting/Utilizing Measures															NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/TG ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³	CDC ¹⁴			
Nursing/General Care:																		
Death Among Surgery Inpatients with Treatable Serious Complications	X																X	Safe
Pressure Ulcer Prevalence	X													X			X	Safe
Falls Prevalence	X																X	Safe
Falls with Injury	X																X	Safe
Restraint Prevalence (Vest and Limb Only)	X																X	Safe
Influenza Vaccination for Healthcare Workers	X	X																Safe
Patient Safety (Non-Surgical):																		
Death in Low Mortality DRGs														X				Safe
Decubitis Ulcers														X				Safe
Failure to Rescue														X				Safe
Iatrogenic Pneumothorax														X				Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/G ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Selected Infections due to Medical Care														X			Safe
Transfusion Reaction														X			Safe
GI Hemorrhage In-Hospital Mortality Rate														X			Safe
Hip Fracture In-Hospital Mortality Rate														X			Safe
Structural:																	
Nursing Care Hours per Patient Day	X															X	Effective, Safe
Nursing Skill Mix (RN, LVN, LPN, UAP, and Contract)	X															X	Effective, Safe
Nursing Practice Environment	X															X	Safety
Nursing Voluntary Turnover	X															X	Safety
Computer Physician Order Entry											X					X	Safe
ICU Physician Staffing (Intensivist)											X					X	Safe

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³		
Evidence-Based Hospital Referral											X				X	Safe
NQF Safe Practices											X				X	Safe
Psychiatric Services:																
Assessment of Violence Risk, Substance Use Disorder, Trauma, and Patient Strengths	X															Effective, Safe, Patient Ctrd.
Hours of Restraint Use	X															Safe
Hours of Seclusion Use	X															Safe
Patients Discharged on Multiple Antipsychotic Medications	X															Effective, Safe
Discharge Assessment and Aftercare Recommendations Sent to Next Level of Care upon Discharge	X															Effective, Patient Ctrd.

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain		
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴	
Care Coordination:																		
3 Item Care Transition		X															X	Effective, Patient Ctrd.
Patient Experience:																		
Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS)	X	X	X	X												X		Patient Ctrd.
Cross-Cutting Length of Stay/Readmission:																		
Inpatient Hospital Average Length of Stay by Medical Service (PacifiCare)																		Efficient
Risk-Adjusted Average Length of Inpatient Stay (CareScience)																		
Severity-Standardized Average												X						

Measure	Organizations Collecting/Utilizing Measures														NQF En-dorsed	IOM Domain	
	Joint Commission	CMS ¹	HQA ²	CMS-RHQDAPU ³	Premier ⁴	SCIP ⁵	STS ⁶	ACC ⁷	ACE ⁸	GW/G ⁹	IHI ¹⁰	Leapfrog ¹¹	NSQIP ¹²	AHRQ ¹³			CDC ¹⁴
Length of Stay, Routine Care																	
Severity-Standardized Average Length of Stay, Special Care												X					
14 Day All-Cause Readmission Rate												X					
Inpatient Readmission Rate by Medical Diagnosis (Pacifcare)																	

¹ Center for Medicare and Medicaid Services

² Hospital Quality Alliance

³ Reporting Hospital Quality Data for Annual Payment Update

⁴ Premier Hospital Quality Incentive Demonstration

⁵ Surgical Care Improvement Project

⁶ The Society of Thoracic Surgeons

⁷ American College of Cardiology

⁸ Alliance for Cardiac Care Excellence

⁹ Get With the Guidelines

¹⁰ Institute for Healthcare Improvement

¹¹ The Leapfrog Group

¹² National Surgical Quality Improvement Program

¹³ Agency for Healthcare Research and Quality

¹⁴ Center for Disease Control

¹⁵ These Premier measures apply only to Hip and Knee Replacement.

APPENDIX D: LIST OF ORGANIZATIONS PARTICIPATING IN THE ENVIRONMENTAL SCAN

Hospital P4P and Public Reporting Program Sponsors

Anthem, National office	Horizon BCBS, NJ
Anthem, VA	Independent Health
Blue Cross Blue Shield, HI	Kaiser Permanente, National and
Blue Cross Blue Shield, IL	Northern and Southern CA offices
Blue Cross Blue Shield, MA	Leapfrog Group (Hospital Rewards
Blue Cross Blue Shield, MI	program)
Blue Shield, Northeastern NY	Maine Health Management Coalition
The Employer Healthcare Alliance	PacifiCare/United Healthcare
Cooperative (“The Alliance”)	Premier Health System
Employers’ Coalition on Health	Priority Health
Excellus/Univera	Providence Health Plan
Fallon Community Health Plan	Regence Blue Shield
Harvard Pilgrim Health Plan	Tufts Health Plan
Health Partners	The Veterans Administration
Highmark BCBS	Anonymous program sponsor (1)

Hospitals and Health Systems

Amsterdam Memorial Hospital, Amsterdam, NY	Geisinger Health System, Danville, PA
Baptist Health System of East TN, Knoxville, TN	Hackensack University Medical Center, Hackensack, NJ
Bleckley Memorial Hospital, Cochran, GA	Henry Ford Health System, Detroit, MI
Crenshaw Community Hospital, Luverne, AL	Hopi Health Care Center, Polacca, AZ
Fairchild Medical Center, Yreka, CA	Kaiser Permanente, CA
Foote Memorial Hospital, Jackson, MI	McLeod Medical Center, Florence, SC
Franklin Medical Center, Greenfield, MA	Mercy Medical Center, Centerville, IA
	Park Nicollet, St. Louis Park, MN

Rice County District One Hospital,
Faribault, MN
San Luis Valley Regional Medical Center,
Alamosa, CO
South Central Regional Medical Center,
Laurel, MS
Southwestern General Hospital, El Paso, TX
Spruce Pine Community Hospital, Spruce
Pine, NC
St. John Health System, Warren, MI

St. Joseph Hospital, Polson, MT
St. Jude Medical Center, Fullerton, CA
Trinity Health System, 20 hospitals in 7
states
Walla Walla General Hospital, Walla Walla,
WA
White River Medical Center, Batesville, AR
William Beaumont Hospital, Royal Oak, MI
Anonymous hospitals (2)

Hospital Associations

American Hospital Association
Association of American Medical Colleges
Catholic Health Association
Federation of American Hospitals

National Association of Children's Hospitals
& Related Institutions
North Carolina Hospital Association
South Dakota Hospital Association
Voluntary Hospital Association

Data Vendors

Hospital Corporation of America
Illinois Hospital Association
Maryland Hospital Association

Premier Health System
Quantros
Thomson Healthcare

Other Organizations

Cypress Healthcare
Kansas Department of Health and
Environment, Office of Local and Rural
Health
Health Resources and Services
Administration, Office of Rural Health
Policy
National Rural Health Association
Stratis Health (Minnesota QIO)
Stroudwater Associates

Upper Midwest Rural Health Research
Center

REFERENCES

- Asch SM, Kerr EA, Keesey J, Adams JL, Setodji CM, Malik S, McGlynn EA. (2006) Who Is at Greatest Risk for Receiving Poor-Quality Health Care? *New England Journal of Medicine* 354(11):1147–1156.
- Asch B, Warner J. (1996) Incentive Systems: Theory and Evidence. In Lewin D, Mitchell D, Zaidi M (eds), *The Human Resource Management Handbook*, Part One. Greenwich, CA: JAI Press, 175–215.
- Barnato AE, Lucas FL, Staiger D, Wennberg DE, Chandra A. (2005) Hospital-Level Racial Disparities in Acute Myocardial Infarction Treatment and Outcomes. *Medical Care* 43:308–319.
- Bazerman MH, Baron J, Skonk K. (2001) *You Can't Enlarge the Pie*. Cambridge, MA: Basic Books.
- Berthiaume JT, Chung RS, Ryskina KL, Walsh J, Legorreta AP. (2006) Aligning Financial Incentives with Quality of Care in the Hospital Setting. *Journal for Healthcare Quality* 28(2):36–44, 51.
- Berthiaume JT, Tyler PA, Ng-Osorio J, LaBresh KA. (2004) Aligning Financial Incentives with “Get with the Guidelines” to Improve Cardiovascular Care. *American Journal of Managed Care* 10(7 Pt 2):501–504.
- Berwick DM. (1995). The Toxicity of Pay for Performance. *Quality Management in Health Care* 4(1):27–33.
- Birkmeyer NJ, Birkmeyer JD. (2006) Strategies for Improving Surgical Quality—Should Payers Reward Excellence or Effort? *New England Journal of Medicine* 354(8):864–870.
- Cameron J, Banko KM, Pierce WD. (2001) Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues. *The Behavior Analyst* 24(1):1–44.

- Casalino LP, Elster A. (2007) Will Pay-for-Performance and Quality Reporting Affect Health Care Disparities? *Health Affairs* 26:w405–w414.
- CMS. (2007a) *CMS Announces Payment Reforms for Inpatient Hospital Services in 2008*. As of August 1, 2007: <http://www.cms.gov/apps/media/pressrelease.asp>
- CMS. (2007b) *National Health Expenditure Prospectus 2006-2016*. Office of the Actuary. As of July, 2007:
<http://www.cms.hhs.gov/nationalhealthexpenddata/downloads/proj2006.pdf>
- Davies HT. (2001) Public Release of Performance Data and Quality Improvement: Internal Responses to External Data by US Health Care Providers. *Quality in Health Care* 10(2):104-10.
- Deci EL, Koestner R, Ryan RN. (1999) A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. *Psychological Bulletin* 125(6):627–668; discussion 692–700.
- Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantellis E, Hiroeh U, Roland M. (2006) Pay-for-Performance Programs in Family Practices in the United Kingdom. *New England Journal of Medicine* 355(4):375–384.
- Fisher ES, Staiger DO, Bynum JPW, Gottlieb DJ. (2007) Creating Accountable Care Organizations: The Extended Hospital Medical Staff. *Health Affairs* 26(1):w44–w57.
- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. (2003) The Implications of Regional Variations in Medicare Spending. Part 1: The Content, Quality, and Accessibility of Care. *Annals of Internal Medicine* 138(4):273–287.
- Fisher ES, Wennberg DE, Stukel TA, Gottlieb DJ, Lucas FL, Pinder EL. (2003) The Implications of Regional Variations in Medicare Spending. Part 2: Health Outcomes and Satisfaction with Care. *Annals of Internal Medicine* 138(4):288–299.

- Freedman JL, Cunningham JA, Krismer K. (1992) Inferred Values and the Reverse-Incentive Effect in Induced Compliance. *Journal of Personality and Social Psychology* 62(3):357–368.
- Glickman SW, Ou F, Delong ER, Roe MT, Lytle BL, Mulgund J, Rumsfeld JS, Gibler WB, Ohman EM, Schulman KA, Peterson ED. (2007) Pay for Performance, Quality of Care, and Outcomes in Acute Myocardial Infarction. *Journal of the American Medical Association* 297:2373–2380.
- Gneezy U, Rustichini A. (2000) Pay Enough or Don't Pay at All. *The Quarterly Journal of Economics* 115(3):791–810.
- Grol R, Baker R, Moss F. (2002) Quality Improvement Research: Understanding the Science of Change in Health Care. *Quality and Safety in Health Care* 11:110–111.
- Grol R and Grimshaw J. (2003) From Best Evidence to Best Practice: Effective Implementation of Change in Patients' Care. *The Lancet* 362:1225–1230.
- Grossbart SR. (2006) What's the Return? Assessing the Effect of "Pay-for-Performance" Initiatives on the Quality of Care Delivery. *Medical Care Research and Review* 63(1 Suppl):29S–48S.
- Heath C, Larrick RP, Wu G. (1999) Goals as Reference Points. *Cognitive Psychology* 38:79–109.
- Heffler S, Smith S, Keehan S, Borger C, Clemens MK, Truffer C. (2005) Trends: U.S. Health Spending Projections for 2004–2014: What Do They Portend For The Federal Growth Initiative? *Health Affairs* 24(2):465–472.
- Holmstrom B, Milgrom P. (1991) Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, and Organization* 7:24–52.
- Institute of Medicine. (2006) *Rewarding Provider Performance: Aligning Incentives in Medicare*. Washington, DC: National Academy Press.

- Institute of Medicine. (2001) *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.
- Jha AK, Li Z, Orav EJ, Epstein AM. (2007) Where Do Elderly Blacks Receive Hospital Care? The Concentration and Quality of Hospitals That Care for Elderly Black Americans. *Archives of Internal Medicine* 167:1177–1182.
- Kahneman D, Knetsch JL, Thaler R. (1986) Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *American Economic Review* 76.
- Kahneman D, Tversky A. (1979) Prospect Theory: An Analysis of Decision Under Risk. *Econometrica* 47(2):263–292.
- Kivetz R, Urminsky O, Zheng Y. (2006) The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention. *Journal of Marketing Research* 43(1):39–58.
- Leapfrog Group. (2007) *Incentives and Rewards Compendium*. As of July 30, 2007: <http://ir.leapfroggroup.org/compendium/>
- Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, Bratzler DW. (2007) Public Reporting and Pay for Performance in Hospital Quality Improvement. *New England Journal of Medicine* 356(5):486–496.
- Loewenstein G, Prelec D. (1992) Anomalies in Intertemporal Choice: Evidence and an Interpretation. *The Quarterly Journal of Economics* 573–597.
- Lowenstein R. (2001 Feb 11) Exuberance Is Rational. *New York Times Magazine*.
- McClellan MB. (2006 Feb 7) Presentation given at National Pay for Performance Summit, Los Angeles, CA.
- McGlynn EA, Asch SM, Adams J, Keeseey J, Hicks J, DeChristofaro A, Kerr EA. (2003) The Quality of Health Care Delivered to Adults in the United States,” *New England Journal of Medicine* 348(26):2635–2645.

- McNeil BJ, Pauker SG, Sox HC, Tversky A. (1982) On the Elicitation of Preferences for Alternative Therapies. *New England Journal of Medicine* 306(21):1259–1262.
- Medicare Payment Advisory Commission (MedPAC). (March 2005) *Report to the Congress: Medicare Payment Policy*. Washington, DC: MedPAC.
- Med-Vantage. (2006) *Provider Pay-for-Performance Incentive Programs: 2005 National Study Results*. San Francisco, CA: Med-Vantage, Inc.
- Mehrotra A, Pearson SD, Coltin KL, Kleinman KP, Singer JA, Rabson B, Schneider EC. (2007) The Response of Physician Groups to P4P Incentives. *American Journal of Managed Care* 13(5):249–255.
- Meyerowitz BE, Chaiken S. (1987) The Effect of Message Framing on Breast Self-Examination Attitudes, Intentions, and Behavior. *Journal of Personality and Social Psychology* 52(3):500–510.
- Nahra TA, Reiter KL, Hirth RA, Shermer JE, Wheeler JRC. (2006) Cost-Effectiveness of Hospital Pay-for-Performance Incentives. *Medical Care Research and Review* 63(1 Suppl):49S–72S.
- Peterson ED, Roe MT, Mulgund J, et al. (2006) Association Between Hospital Process Performance and Outcomes Among Patients with Acute Coronary Syndromes. *Journal of the American Medical Association* 295(16):1912–1920.
- Pham, HH, Coughlan, O'Malley AS. (2006). The Impact of Quality-Reporting Programs on Hospital Operations. *Health Affairs*. 25(5): 1412-1422.
- Premier, Inc. (2006) *Centers for Medicare and Medicaid Services (CMS)/Premier Hospital Quality Incentive Demonstration Project: Project Overview and Findings from Year One*. Charlotte, NC: Author.
- Reiter KL, Nahra TA, Wheeler JRC. (2006) Hospital Responses to Pay-for-Performance Incentives. *Health Services Management Research* 19(2):123–134.

- Rosenthal MB, Frank RG, Li Z, Epstein AM. (2005) Early Experience with Pay-for-Performance: From Concept to Practice. *Journal of the American Medical Association* 294(14):1788–1793.
- Rothe H. (1970). Output Rates Among Welders: Productivity and Consistency Following Removal of a Financial Incentive System. *Journal of Applied Psychology* 54:549–551.
- Sauter KM, Bokhour BG, White B, Young G, Burgess JF, Berlowitz D, Wheeler JRC. (2007). Early Experiences of a Hospital-based Pay-for-Performance Program. *Journal of Healthcare Management* 52(2):95–108.
- Schuster MA, McGlynn EA, Brook RH. (1998) How Good Is the Quality of Health Care in the United States? *Milbank Quarterly* 76(4):517–563.
- Shekelle P. (2007 Apr 4) Medicare’s Hospital Compare Performance Measures and Mortality Rates. *Journal of the American Medical Association* 297(13):1430–1431; author reply 1431.
- Skinner J, Chandra A, Staiger D, Lee J, McClellan M. (2005) Mortality After Acute Myocardial Infarction in Hospitals That Disproportionately Treat Black Patients. *Circulation* 112:2634–2641.
- Sorbero ME, Damberg CL, Shaw R, et al. (2006) *Assessment of Pay-for-Performance Options for Medicare Physician Services: Final Report*. RAND Working Paper prepared for the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. Santa Monica, CA: RAND.
- Thaler R. (1985) Mental Accounting and Consumer Choice. *Marketing Science* 4(3):199–214.
- Thompson RE. (2005) Is Pay for Performance Ethical? *Physician Executive* 31(6):60–62.
- Titmuss RM. (1970) *The Gift Relationship: From Human Blood to Social Policy*. New York, NY: Allen & Unwin.

- Ubel PA, Hirth RA, Chernew ME, Fendrick AM. (2003) What Is the Price of Life and Why Doesn't It Increase at the Rate of Inflation? *Archives of Internal Medicine* 163(14):1637–1641.
- Wenger NS, Solomon DH, Roth CP, MacLean CH, Saliba D, et al. (2003) The Quality of Medical Care Provided to Vulnerable Community-Dwelling Older Patients. *Annals of Internal Medicine* 139(9):740–747.
- Werner RM, Bradlow ET. (2006) Relationship Between Medicare's Hospital Compare Performance Measures and Mortality Rates. *Journal of the American Medical Association* 296(22):2694–2702.
- Williams SC, Schmaltz SP, Morton DJ, Koss RG, Loeb JM. (2005) Quality of Care in U.S. Hospitals as Reflected by Standardized Measures, 2002–2004. *New England Journal of Medicine* 353(3):255–264.