1. Introduction

Value-based purchasing (VBP) refers to a broad set of performance-based payment strategies that link financial incentives to providers' performance on a set of defined measures. Both public and private payers are using VBP strategies in an effort to drive improvements in quality and to slow the growth in health care spending. Nearly 10 years ago, the Department of Health and Human Services (HHS) and the Centers for Medicare and Medicaid Services (CMS) began testing VBP models with their hospital pay-for-performance (P4P) demonstrations, known as the Premier Hospital Quality Incentive Demonstration (HQID) and the Physician Group Practice (PGP) Demonstration, which provided financial incentives to physician groups that performed well on quality and cost metrics. The use of financial incentives as a strategy to drive improvements in care dates back even further among private payers[2,9] and Medicaid programs, which began to experiment with P4P in the mid-1990s and early 2000s. These early private payer P4P programs generally focused on holding providers accountable for their quality performance and targeted physician groups, individual physicians, and hospitals.[19, 20, 52]

Although the published evidence from P4P programs implemented by private-sector payers between 2000 and 2010 showed mostly modest results in improving performance,[3–10] public and private payers have continued to experiment with the use of financial incentives as a policy lever to drive improvements in care. Many of the early P4P program designs have evolved over time to include a larger and broader set of measures, including resource use and cost metrics, in an effort to reward providers for delivering value,* and many are deploying a wider range of incentives. Additionally, other VBP models have since emerged and are currently being tested, including accountable care organizations (ACOs) and bundled payment programs that include both quality and cost design features.

Policy Context and Study Purpose

The Medicare program has gradually been moving toward implementing VBP across various care settings starting with pay-for-reporting programs (e.g., the Hospital Inpatient Quality Reporting program and the Physician Quality Reporting Initiative) and P4P demonstrations to gain experience. The 2010 Patient Protection and Affordable Care Act[11] significantly expands VBP by requiring the Medicare program to implement, develop plans for, and test in the context

*Value is defined as the outcomes (outputs) achieved divided by the cost or resources used (inputs) to generate those outcomes.

of demonstrations the use of VBP across a broad set of providers and settings of care (i.e., physicians, skilled nursing homes, home health agencies, ambulatory surgery centers, long-term hospitals, rehabilitation hospitals, cancer hospitals, psychiatric hospitals, and hospice facilities), as shown in Table 1.1. For example, the Patient Protection and Affordable Care Act required HHS to submit plans for implementing VBP in ambulatory surgery centers, home health, and skilled nursing homes to Congress in 2011. The Hospital Value-Based Purchasing Program, which makes payment adjustments (both bonuses and penalties) to hospitals based on performance, began implementation in October 2012, and the Physician Value-Based Payment Modifier will start in January 2015. The Patient Protection and Affordable Care Act further links provider payments to cost reductions and quality improvements through the implementation and testing of VBP models such as ACOs and bundled payments. To that end, Medicare has begun the Bundled Payments for Care Improvement demonstrations and the ACO shared savings programs and demonstrations. Moreover, Congress is actively considering ways to revise the physician fee schedule (i.e., the sustainable growth rate or SGR) to incorporate VBP incentives so that payment policy for physicians paid under fee-for-service (FFS) supports the delivery of high quality care and efficient use of resources.

Table 1.1. 2010 Patient Protection and Affordable Care Act Value-Based Purchasing Provisions

Type of Value-Based Purchasing Program and Setting	Timeline
Pay-for–Performance	
Hospital Value-Based Purchasing Program	October 1, 2012 (current program)
Physicians (or groups of physicians) under Physician Value-Based Payment Modifier	January 1, 2015, for a subset of physicians January 1, 2018, for all physicians (program to be implemented)
Inpatient critical access hospitals	No later than 2 years after date of act (May 1, 2010) (demonstration program)
Hospitals excluded from the Hospital Value-Based Purchasing Program due to insufficient numbers of measures and cases	No later than 2 years after date of act (May 1, 2010) (demonstration program)
Long-term care hospitals	No later than January 1, 2016 (pilot program)
Hospice programs	No later than January 1, 2016 (pilot program)
Psychiatric hospitals	No later than January 1, 2016 (pilot program)
Rehabilitation hospitals	No later than January 1, 2016 (pilot program)
PPS-exempt cancer hospitals	No later than January 1, 2016 (pilot program)
Ambulatory surgical centers	Submit plan to Congress no later than January 1, 2011 (plan for program)
Home health agencies	Submit plan to Congress no later than October 1, 2011 (plan for program)
Skilled nursing facilities	Submit plan to Congress no later than October 1, 2011 (plan for program)
Shared Savings	
ACOs	no later than January 1, 2012 (current program)
Bundled Payment	
Hospital/physicians/post-acute care	no later than January 1, 2012(demonstration program)

Despite the widespread enthusiasm for and adoption of P4P programs over the past decade, uncertainty remains about the extent to which these programs have been successful in accomplishing their goals and what design elements work best and under which conditions.[6, 20, 108] Various studies have identified a number of issues, such as uncertainty about the size of incentives required to attain desired changes,[108] whether financial incentives may lead to unintended consequences,[43, 109–111] and measurement issues, including measure reliability and misclassification risk.[112–114]

VBP models are quite new to the health system and represent a work in progress in terms of our understanding of how best to design these programs to achieve desired goals, the optimal conditions that support successful implementation, and provider response to the incentives. Because of the substantial investments that HHS is making to implement and test a variety of VBP models, this is an opportune moment to reflect on what has been learned from the past decade of experimentation that could guide current and future federal policymaking related to VBP program design and implementation. It is also an important moment to consider the type of monitoring and systematic evaluation work that is needed to generate the information that policymakers require to fine-tune VBP program designs and to understand the impact these programs are having related to stated goals.

To that end, the Office of the Assistant Secretary for Planning and Evaluation (ASPE) in HHS asked RAND to review what has been learned about VBP over the past decade that might help inform policymaking. In particular, ASPE asked RAND to address three overarching questions: (1) What is the evidence regarding whether VBP programs have been successful? (2) What are the elements of successful VBP programs? (3) What questions remain unanswered that, if answered, could improve the design and functioning of VBP programs moving forward? This report presents the findings from RAND's review. For recommendations based on the findings, we direct readers to the companion document to this summary report, *Measuring Success in Health Care Value-Based Purchasing Programs: Summary and Recommendations*.

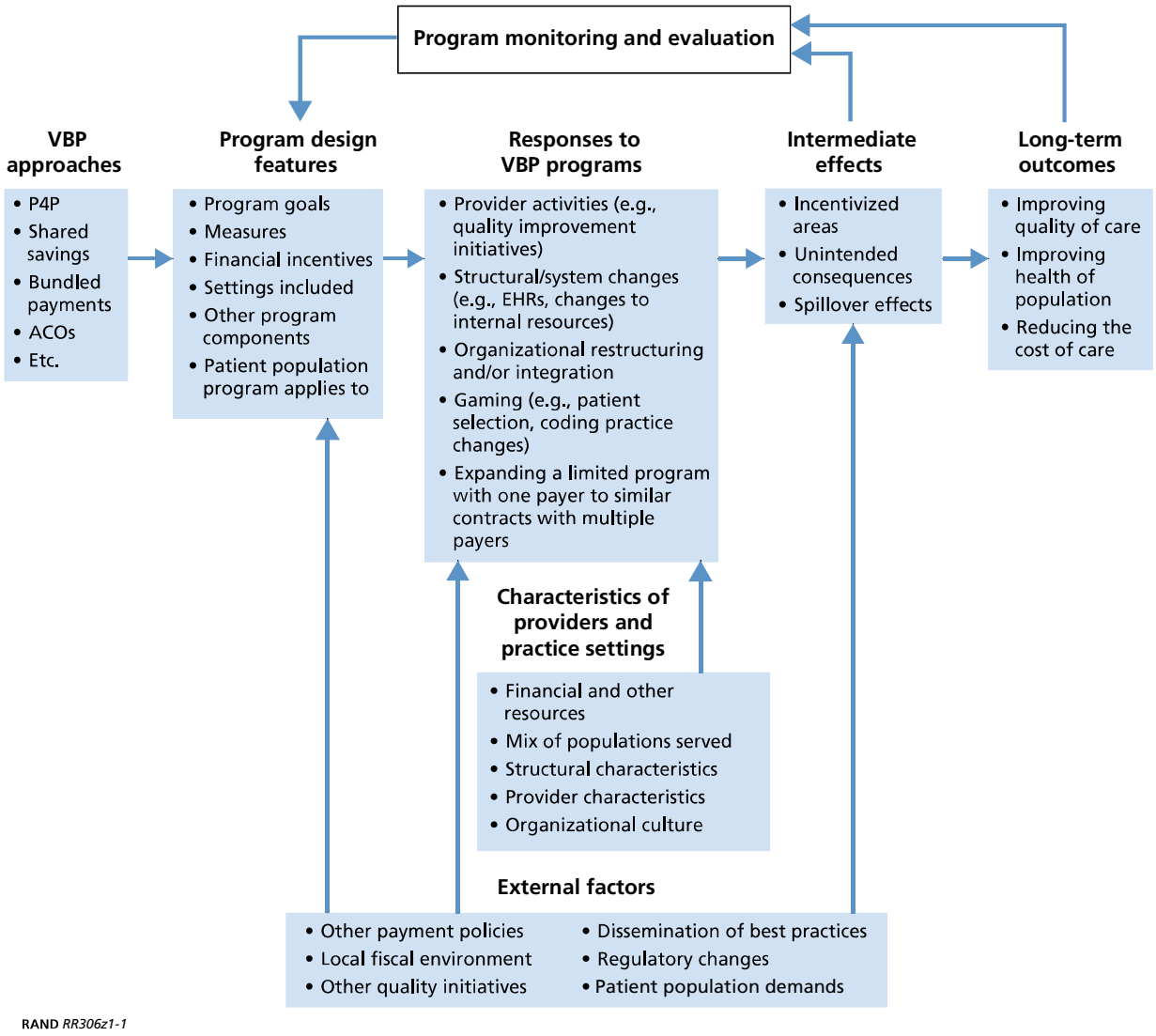
HHS is actively considering the federal government's near- and long-term strategy for how to design and implement VBP programs to achieve the three aims set forth in the National Quality Strategy,[115, 116] which focus on improving the overall quality of care, improving the health of the U.S. population, and making care affordable by reducing the cost of quality health care. In designing their VBP strategy, HHS seeks to apply the best available evidence to guide policymaking regarding the expansion of VBP across a range of Medicare program settings. To inform HHS's strategy development, RAND reviewed the published evidence and consulted with experts on whether VBP programs have been successful in meeting goals, to identify features of successful VBP programs, and to identify knowledge gaps to inform the focus of future evaluation and monitoring efforts.

Conceptual Framework for Assessing the Effects of Value-Based Purchasing Programs

Several useful frameworks have been developed regarding how to evaluate VBP programs. These include a framework by Dudley et al. focused on assessing P4P experiments,[12] one by McHugh and Joshi to guide VBP program evaluation,[117] and another by Fisher et al. specifically focused on evaluating ACOs.[118] The three frameworks have similar elements. For example, the Dudley framework describes three core elements—the incentive, predisposing factors, and enabling factors—that influence or mediate the response to provider incentives for quality improvement, while the framework by McHugh and Joshi defines program design, the participants, and contextual factors as elements associated with impacts and implementation. The Fisher ACO evaluation framework similarly considers program components (i.e., ACO contract characteristics, implementation activities), provider characteristics (ACO structure and capabilities), and external factors (i.e., environmental context) that influence intermediate outcomes and impacts. Because ACOs represent a new model of VBP that is just starting to be tested by both public and private payers, Fisher and colleagues emphasize that both formative and summative research will be important for guiding policymaking. The McHugh and Joshi framework also emphasizes the importance of implementation (i.e., formative evaluation) research.

Figure 1.1 displays the framework we developed for this project, which was adapted from the three existing frameworks. For example, in our VBP framework, program design features, characteristics of providers and practice settings, and external factors correspond to the incentive, predisposing factors, and enabling factors described in the Dudley model. Characteristics of providers and practice settings and external factors are important contextual elements that influence the response of providers to the incentives. Our framework attempts to expand on the previously developed frameworks by detailing some of the specific factors that should be considered within each of the elements of the framework.

Figure 1.1. Value-Based Purchasing Conceptual Framework



Because VBP programs are natural experiments and the associated research is observational in nature, Dudley (2005) underscores that it is critical that evaluators select theory-driven hypotheses about how incentives affect behavior so as to identify potential confounding factors that could explain observed effects.[13] A framework is a useful construct to help develop theory-driven hypotheses.

The conceptual framework offers a foundation for considering the design features of the incentive program, as well as other mediating factors that influence whether and how providers may respond to the incentives and whether programs are successful in reaching stated goals. We refer the reader to Appendix B, which lists a set of factors that our panel of technical experts (discussed below) recommended be systematically collected for all VBP programs to facilitate efforts to evaluate VBP programs and to compare and contrast observed impacts across

programs. The framework also can be used to guide discussions about the design and implementation of existing VBP programs and those in development and to define a structured agenda for monitoring and evaluating VBP programs, with the explicit goal of developing knowledge to improve the functioning of these programs.

Methods and Research Questions

For this study, we defined VBP programs as private or public programs that link financial reimbursement to performance on measures of quality (i.e., structure, process, outcomes, access, and patient experience) and cost or resource use. Three broad categories of VBP models were the focus of our review: (1) P4P, which includes both "pay for quality" and "pay for quality and resource use, efficiency, or costs"; (2) shared savings models that typically, but not exclusively, are being deployed in the context of ACOs; and (3) bundled payments for episodes of care (only when paired with holding providers accountable for performance on quality measures). We excluded pay-for-reporting and demand-side programs (e.g., tiered networks and consumer incentives).

We define each of the three broad types of VBP models as follows:

- **Pay-for-performance** refers to a payment arrangement in which providers are rewarded (bonuses) or penalized (reductions in payments) based on meeting pre-established targets or benchmarks for measures of quality and/or efficiency. These financial incentives are intended to change provider behavior to achieve a set of objectives specified by the payer.
- **Accountable care organization** refers to a health care organization composed of doctors, hospitals, and other health care providers who voluntarily come together to provide coordinated care and agree to be held accountable for the overall costs and quality of care for an assigned population of patients. The ACO payment model ties provider reimbursements to performance on quality measures and reductions in the total cost of care. Under an ACO arrangement, providers in the ACO agree to take financial risk and are eligible for a share of the savings achieved through improved care delivery provided they achieve quality and spending targets negotiated between the ACO and the payer.
- **Bundled payments**[*] are a method in which payments to health care providers are based on the expected costs for a clinically defined episode or bundle of related health care services. The payment arrangement includes financial and quality performance accountability for the episode of care. Episodes can be defined in different ways, cover varying periods of time (e.g., one year for a chronic condition, the period of the hospital

[*] Other common terms used for bundled payment arrangements are *episode-based payment*, *episode payment*, *episode-of-care payment*, *case rate*, *evidence-based case rate*, *global bundled payment*, and *global payment*.

stay and 30 days post-discharge), and include single or multiple health care providers of different types (e.g., hospital only, hospital and ambulatory provider).[1, 104, 119]

Table 1.2 lists the research questions that ASPE asked RAND to address in its review. The questions address three broad areas of inquiry: (1) measuring the performance of VBP programs; (2) the results of performance in VBP programs; and (3) improving the performance of VBP programs.

Table 1.2. Research Questions

Measuring Performance in Value-Based Purchasing Programs
1. What goals should be set? (How should success be defined for VBP programs?)
2. What are the metrics by which VBP programs can and should be evaluated?
3. Which aspects of VBP are measurable and which are not?
4. What is the relationship between health outcomes and what is measured in VBP programs?
Results of Performance in Value-Based Purchasing Programs
5. Based on the metrics used to date, have VBP programs facilitated improvements in quality and value? 5a. What improvements in health outcomes attributable to VBP can we expect, and over what time horizon? 5b. What cost savings attributable to VBP can we expect, and over what time horizon?
6. Does performance on unmeasured aspects of quality of care suffer when providers focus on improving performance on what is being measured ("teaching to the test")? Conversely, are there "spillover effects" whereby quality improvement efforts improve care more broadly?
7. If a provider/institution performs highly on all the VBP metrics but has average performance on everything that is not measured, which proportion of total potential improvement in health will be achieved? In other words, if we imagine that a high-performing health system produces "X" amount more "quality-adjusted life years" than an average-performing system, what fraction of that X would be produced by a health system that was higher-performing on metrics commonly included in VBP programs currently, but was average-performing in unmeasured areas?
8. How likely is it that improvements in our ability to measure what is important will change enough over the next five to ten years to significantly affect the answer to (7)?
9. Are there unexpected effects of VBP programs, including impacts on racial/ethnic and socioeconomic disparities, and access to care?
10. What are the features of the highest-performing providers/institutions and their adaptations to VBP?
11. What are the characteristics of the lowest-performing providers/institutions and their behaviors in response to VBP?
12. How much does it cost a provider/institution to improve on the measured performance areas? 12a. Are the incentive levels of VBP programs sufficient to cover the costs of investing in quality improvement? 12b. How do organizations weight these factors related to VBP and decide on quality improvement investments?
Improving the Performance of Value-Based Purchasing Programs
13. What are the critical gaps in knowledge about VBP, and how can these gaps be addressed?
14. What are the structural and implementation features of the most successful VBP programs?
15. Within VBP programs, how can practices from the highest-performing providers/institutions be disseminated?
16. To what extent can VBP programs that have a positive impact in health care be improved and expanded?

We used three approaches to gather information to address the research questions:

- **Environmental scan of existing value-based purchasing programs:** We reviewed information that was publicly available for both publicly and privately sponsored VBP programs. We extracted information on program characteristics, program effects, and study designs when evaluations were available.

- **A review of the published evaluation literature on value-based purchasing:** We examined the peer-reviewed published literature for studies that specifically evaluated the impact of P4P, ACO, or bundled payment programs. We drew heavily from existing review articles where available.
- **Input from a technical expert panel:** Recognizing that many of the design issues and implementation lessons have not found their way into the published literature and likely never will, we convened a technical expert panel (TEP) to provide input on the study questions. The TEP was composed of VBP program sponsors (i.e., private plans and a regional multi-stakeholder collaborative), providers from health systems who have been the target of VBP programs, health services researchers with expertise in examining the effects of VBP programs, and federal participants who represented public payers. The TEP met twice in person, in May and June of 2013, for an all-day meeting facilitated by RAND staff. We provided the TEP with the findings from the environmental scan of programs and the literature review as background to inform the panelists' discussions.

At the start of the chapters that focus on the environmental scan and literature review, we detail the specific methods we used. For the literature review, we assessed the methodological quality of each study and the strength of the evidence as a whole for each research question that examined the impact or effect of VBP. Our approach followed the methodology outlined in the Agency for Healthcare Research and Quality's (AHRQ's) *Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews*.[120] We did not grade the evidence for those questions with descriptive results (e.g., goals used to assess success).

We rated the methodological quality of each study using three categories of rating (poor, fair, and good). This is a grade of the absolute strength of the evidence presented in each study. **Good** indicates a low risk of bias (i.e., the study has strong methods to guard against bias), **fair** indicates a medium risk of bias, and **poor** indicates a high risk of bias. In the evidence tables, we provide short explanation for each good and poor rating. Factors that contributed to our assessment of methodological quality included study design (randomization, matching), analytic techniques (attempts to control for confounding explanations), intervention characteristics (size of intervention and ability to generalize results more broadly), and conflict of interest/independence of the evaluator. Countries other than the United States have implemented VBP programs, particularly P4P programs, and we have included selected studies from non-U.S. programs in our review; however, the findings from these studies may not readily generalize to the United States because of differences in health system organization and structure, financing, and delivery, as well as substantial differences in the design of the VBP intervention.

We graded the strength of the evidence as a whole for each research question using four grade levels:

- **High—**A high degree of confidence that the evidence reflects the true effect. Additional research is unlikely to change the estimate of the effect.
- **Moderate**—Moderate confidence that the evidence reflects the true effect. Additional research may change the estimate or confidence in the estimate of the effect.

- **Low**—Low confidence that the evidence reflects the true effect. Further evidence is likely to change our confidence in the estimate of effect and is likely to change the estimate. A low rating indicates that there is a high risk of bias and residual confounding.
- **Insufficient**—A lack of evidence to estimate the effect(s).

Key considerations that affect the classification according to this scheme are the risk of bias, consistency across studies, directness, precision, coherence, residual confounding, and strength of association. The three senior researchers on the team (Cheryl Damberg, Melony Sorbero, and Grant Martsolf) independently graded the collective evidence, discussed differences in ratings, and together generated the final rating. Because two of the VBP categories (ACOs and bundled payments) are quite new in their development and implementation, there is currently insufficient evidence on the effects of these VBP models.

Organization of This Report

The remainder of the report addresses the findings from our three data collection approaches. Chapter Two focuses on the results of the VBP environmental scan, while Chapters Three through Five focus on the findings from the literature review for P4P (Chapter Three), ACOs (Chapter Four), and bundled payments (Chapter Five), and Chapter Six summarizes the key points of the TEP's discussion. We give our concluding thoughts in Chapter Seven.