NOTE:

These materials were prepared by subcontractors for consideration by the *Committee on Geographic Variation in Health Care Spending and Promotion of High-value Care*. These analyses were commissioned and overseen by the Committee. However, the findings and views expressed in the subcontractor reports do not necessarily reflect those of the NRC/IOM or the Committee. Neither the methodology nor the subcontractor reports have been subject to formal institutional review for the Interim Report. As the committee continues to review the findings from the analyses contained herein, we invite you to provide feedback on the content of these reports. Please note that any comments will be entered into the project's public access file, and will be available for public review.

Provide Feedback

IOM Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Care

A Modeling of Policy Recommendations

David Auerbach, Peter Huckfeldt, Peter Hussey, Abby Alpert, Hangsheng Liu, Victoria Shier, Christopher Lau, Mark Totten, Ateev Mehrotra

RAND Health

WR-970-IOM
January 2013
Prepared for the Institute of Medicine

Preface

The Center for Medicare and Medicaid Services asked the Institute of Medicine to investigate geographic variation in health care spending for Medicare beneficiaries and to make recommendations on payment changes that both address this geographic variation and promote high-value care. The resulting Institute of Medicine Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Care was chaired by Joseph P. Newhouse and Alan M. Garber.

To inform their recommendations, this Committee asked RAND to model three policies that can potentially reduce geographic variation and improve value: (1) pay for performance, (2) bundled payment, and (3) accountable care organizations.

This work was sponsored by the Institute of Medicine for which Robin Graham served as senior project officer. The research was conducted in RAND Health, a division of the RAND Corporation. A profile of RAND Health, abstracts of its publications, and ordering information can be found at www.rand.org/health

Abstract

Per beneficiary Medicare spending varies widely across the United States. This geographic variation in spending has raised concern because it may reflect unnecessary variation in practice that does not improve health. Therefore, there has been interest in what policies might decrease geographic variation in spending. We modeled the potential impact of three policies (pay for performance, bundled payment, and accountable care organizations) on variation in spending across Hospital Referral Regions. We compared 2008 Medicare spending for each Hospital Referral Region under the baseline case and various scenarios for each of the three policies. Under our models, we find that neither pay for performance nor accountable care organizations has substantial impact on geographic variation in Medicare spending. In contrast, bundled payment substantially decreases geographic variation in spending for the care included in the bundles.

Table of Contents

Preface ii
Abstract iii
Figures v
Tables vi
Summary viii
Acknowledgements x
Abbreviations xi
Chapter 1 Introduction and Overlying Principles 1
Chapter 2 Pay for Performance 4
Introduction 4
Brief Methods 4
Scenarios and Assumptions 5
Results 7
Limitations 12
Summary 12
Chapter 3 Bundled Payment 13
Introduction 13
Brief Methods 13
Scenarios and Assumptions 14
Results 16
Limitations 22
Summary 22
Chapter 4 ACOs 24
Introduction 24
Brief Methods 24
Scenarios and Assumptions 25
Results 27
Limitations 31
Summary 31
Chapter 5 Conclusions 33
References 35
Appendices 37
Appendix 1 – Pay-for-Performance (P4P) Methods 37
Appendix 2 – Bundled Payment Methods 48
Appendix 3 – Accountable Care Organizations (ACO) Methods 58

Figures

Figure 2.1. Impact of The Hospital P4P Programs on HRR per Beneficiary Spending on Inpatient Care Under Scenario 3 10
Figure 3.1. Impact of Bundled Payment on HRR Per Beneficiary Spending for Bundled Services under Scenario 2 21
Figure 4.1. Proportion of Medicare FFS Beneficiaries in HRRs Enrolled in Medicare or Private ACOs 28
Figure 4.2. Relationship Between HRR per-Beneficiary Medicare Costs and ACO Penetration 29
Figure 4.3 Impact of ACOs on HRR per Beneficiary Spending under Scenario 4 30

Tables

Table 2.1. P4P Scenarios Modeled 6
Table 2.2. Distribution of Percentage Change in Provider Payments Under Scenarios 1 and 3 7
Table 2.3. Percentage of Providers Receiving an Increase or No Change in Payment under P4P for Scenario 3 (Linear 15%) 8
Table 2.4. HRRs With Smallest And Largest Reductions In Per-Beneficiary Medicare Expenditures Under Scenario 3 9
Table 2.5. Impact of P4P Programs on Total per Beneficiary Medicare Spending and Geographic Variation in Spending 10
Table 2.6. Correlation Between the Performance on Quality Measures Used in the Inpatient P4P Program and Inpatient Spending per Beneficiary by HRR 11
Table 3.1. Clinical Conditions and Related MS-DRGs Included in the Hypothetical Medicare Bundled Payment Program 14
Table 3.2. Bundled Payment Scenarios Modeled 15
Table 3.3. Distribution Of Provider-Level Payment Changes Under Bundled Payment Scenario 2.....17
Table 3.4. In Six Selected HRRs, the Percentage of Providers Receiving Either an Increase or No Change in Medicare Payment Under Bundled Payment Scenarios 2 and 5 18
Table 3.5. Spending by Category of Services That Would Be Included in Bundles for Selected MS-DRGs, by Quartile of Spending, 2008 19
Table 3.6. Impact of Bundled Payment on per Beneficiary Spending and Geographic Variation in Spending for Bundled Services 20
Table 3.7. Impact of Bundled Payment on Total per Beneficiary Medicare Spending and Geographic Variation in Spending..... 22
Table 4.1. ACO Scenarios Modeled 26
Table 4.2. ACO Enrollment Groups and Associated Number of Beneficiaries, as of October 2012.....27
Table 4.3. Impact of ACOs on Total per Beneficiary Medicare Spending and Geographic Variation in Spending 31
Table 5.1. Summary of Three Modeled Policies on Spending and Geographic Variation..... 33
Table A.1.1. Steps in P4P Analysis Plan 37
Table A.1.2: Quality Measures Used in Home Health Value-Based Purchasing Program 41
Table A.1.3: Quality Domains Used in Nursing Home Quality-Based Purchasing Demonstration.....43
Table A.1.4: Quality Measures Used in Nursing Home Quality-Based Purchasing Demonstration by Domain.....45

Table A.2.1. Clinical Conditions And Related MS-DRGs Included In The Hypothetical Medicare Bundled Payment Program 49
Table A.2.2. Carrier and Outpatient Bundle Components and Associated Adjustment Ratio 54
Table A.3.1. Independent Variables In The ACO Penetration Prediction Model 61

Summary

The Institute of Medicine (IOM) Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Care is investigating the underlying causes of geographic variation in Medicare spending. The Committee asked RAND to analyze three policies that can potentially reduce this variation and improve the value of care provided to Medicare beneficiaries: (1) pay for performance (P4P), (2) bundled payment, and (3) accountable care organizations (ACOs). We modeled the impact of these policies on overall Medicare spending and variation in spending across Hospital Referral Regions (HRRs).

P4P was modeled as a budget-neutral program in which a fixed percentage of all Medicare reimbursement to providers was redistributed toward hospitals, nursing homes, and home health agencies with superior performance on relevant quality measures. If providers with high performance on those measures also had lower-than-average spending, geographic variation would be reduced. In our five scenarios, we altered key program design parameters such as how much money was tied to incentive payments, who received the incentive payments, and how providers responded to those incentives. In our models, we found that many providers had large increases or reductions in their reimbursement under the pay-for-performance program. However, there was no systematic relationship between performance on the quality measures and spending per beneficiary in the HRR. Geographic variation was therefore unaffected.

We also modeled a bundled payment policy in which a provider received a fixed payment for all care occurring from the point of hospital admission through 30 days after hospital discharge. The fixed payment was based on condition-specific historical Medicare spending for bundled services. Therefore, providers with higher spending on such bundles would receive reduced payment and those with lower spending would receive an increased payment. We implemented the bundle for ten conditions, encompassing 15 percent of total Medicare spending. If providers who received reductions in payment clustered in HRRs with higher overall Medicare spending, then bundled payment would decrease variation in Medicare spending. We investigated a range of scenarios that varied factors such as what services were included in the bundle and which providers received a national bundled payment rate. Overall reductions in Medicare spending ranged from 0% to 4% and were largely determined by how the bundled payment was set. We found significant reductions in HRR-level variation in per-beneficiary Medicare spending for the services included in the bundle, with more modest reductions in HRR-level variation for total Medicare spending per beneficiary. The more modest reduction in variation in total Medicare spending is driven by the small fraction of total Medicare spending included in the bundles.

Finally, we modeled an ACO policy. ACOs were introduced into Medicare under the Affordable Care Act. Groups of providers who become an ACO collectively ‘share savings’ with Medicare if they meet certain quality benchmarks and reduce total spending for their enrolled

fee-for-service Medicare beneficiaries below a target level. If ACOs are more likely to develop in high-cost areas (or achieve relatively greater savings in high-cost areas), then geographic variation in Medicare spending would be reduced. We found that areas of the country with characteristics of integrated care systems were more likely to have ACOs. Yet, we observed little clustering of current ACOs in high-cost areas, and therefore geographic variation in spending was essentially unchanged.

Generally, the three polices we modeled aim to improve the value of Medicare by encouraging low-cost and/or high-quality medical practice on the part of providers. While they may indeed improve the value of Medicare, only the bundled payment policy, under our models, significantly reduced geographic variation in Medicare spending.

Acknowledgements

We thank David Adamson, Carol Cheng, Diane Bronowicz, and Benjamin Galick for their valuable contributions to this work. We appreciate the thoughtful input of the IOM Committee members, Thomas J. Hoerger, Benjamin Yarnoff, Rosalie Pacula, Barbara Wynn, Cheryl Damberg, and Neeraj Sood on our models and results.

Abbreviations

ACA	Affordable Care Act
ACE	Angiotensin-Converting Enzyme
ACOs	Accountable Care Organizations
AHA	American Hospital Association
AMI	Acute Myocardial Infarction
BPCI	Bundled Payment For Care Improvement Initiative
CAHs	Critical Access Hospitals
CMI	Centers for Medicare and Medicaid Innovation
CMS	Centers For Medicare And Medicaid
CNA	Certified Nursing Assistant
CV	Coefficient of Variation
DME	Durable Medical Equipment
DRG	Diagnostic-Related Group
DSH	Disproportionate Share Hospitals
ESRD	End-Stage Renal Disease
FFS	Fee-For-Service
HHA	Home Health Agency
HRRs	Hospital Referral Regions
IME	Indirect Medical Education
IOM	Institute Of Medicine
MedPAR	Medicare Provider Analysis and Review
MS-DRGs	Medicare Severity Diagnosis Related Groups
P4P	Pay For Performance
PCPs	Primary Care Physicians
RN	Registered Nurse
SAF	Standard Analytic File
SSP	Shared Savings Program
VBP	Value-Based Purchasing Program

Chapter 1 Introduction and Overlying Principles

Geographic variation in health care spending has been a persistent concern for US policymakers. Medicare spending per fee-for-service (FFS) beneficiary in 2008 ranged from about $6,000 in Rapid City, South Dakota to over $18,000 in Miami, Florida. Some of this geographic variation is driven by differing health status and needs of the residents of these regions. However, evidence suggests that a significant component of this geographic variation results from variation in medical practice (Congressional Budget Office, 2008b). That observation has led to the natural question: could the high-cost regions spend less and have the same outcomes?

As part of its work investigating the underlying causes of geographic variation in health spending, the Institute of Medicine (IOM) Committee on Geographic Variation in Health Care Spending and Promotion of High-Value Care asked RAND to investigate three policies that could potentially reduce geographic variation in Medicare spending and improve the value of care provided: (1) pay for performance (P4P), (2) bundled payment, and (3) accountable care organizations (ACOs). We estimated the impact of these policies on overall Medicare spending and variation in spending across Hospital Referral Regions (HRRs). HRRs represent geographically-contiguous distinct health care markets as defined by where patients receive tertiary hospital care. There are 306 HRRs in the United States.

The three policy interventions selected by the Committee share several characteristics. The policies have received significant interest from Medicare and private payers and are being tested in many ongoing initiatives. More fundamentally, they seek to address the problems inherent with the current fee-for-service (FFS) payment system. Providers have wide discretion in deciding on a patient's treatment – because of considerable uncertainty about the best course of treatment given a patient's history and symptoms. Yet FFS pays providers 'piecemeal' for individual procedures, office visits, diagnoses, or hospitalizations, with adjustments for local wages and input costs, and patient complexity, but no adjustment for quality of care provided or penalty for provision of excess or inappropriate care. If an underlying driver of geographic variation is inefficient practice patterns, implementation of these three policies could reduce geographic variation in Medicare spending.

Current Medicare bundled payment policy seeks to address the problems of FFS by paying a single price for a bundle of related services that begins with a hospitalization. It can be seen as a natural extension of the concept of the Diagnosis-Related Group that is currently used to pay for inpatient care. Depending on where the bundled payment is set, providers face an incentive to decrease the cost of services provided within the bundle. Accountable Care Organizations embody the same concept, but on a larger scale. All care provided to a given Medicare beneficiary is considered a 'bundle' in which providers receive bonus payments (shared savings)

if that total cost falls below a given target, and they simultaneously meet certain standards for quality of care. Finally, P4P seeks to better reward providers for choosing services that yield higher quality and outcomes of care thus potentially decreasing the incentive to provide inefficient care.

It is beyond the scope of this paper to directly evaluate whether these policies would, in fact, reduce distortions and improve the efficiency (that is, quality and outcomes of care provided given the cost) of Medicare. Instead we seek to answer, under a relatively constrained set of assumptions and circumstances, whether they would reduce geographic variation in Medicare spending. Based on feedback from the Committee, we employed several underlying principles in simulating the impact of each policy. First, our focus was on the medium-term impact of these policies. Our estimates do not include transition periods for any change in policy and therefore our results should be interpreted as the impact of a steady-state, mature program that one might expect three to five years after the policy is implemented. Second, the Committee asked us to consider realistic and robust policy interventions. To ensure that the policies are realistic, we based our analysis on policies that have already been implemented in Medicare (either as pilots or full programs). To ensure that they are robust -- large enough to have a measurable impact and to help illustrate the potential impacts of an expanded, large-scale program -- we modified the existing programs in many key ways. For example, instead of a voluntary program, we modeled a mandatory bundled payment program.

Third, we use an accounting-style approach to 'model' the impact of these policies. For example, based on input from the Committee, we included a relatively constrained set of behavioral effects (i.e. provider responses to the modeled payment policies) – which were supported by established literature. Given the constraints of the project, we were unable to consider dynamic effects of the policies (aspects that are state-dependent or would change over time). We recognize that the longer- and shorter-term effects of these policies may differ from those as presented. Also, in estimating the impact of the policies, we compared observed Medicare spending in 2008 for each HRR with modeled spending under each of the three new payment policies as if they were implemented in that year. We chose 2008 because it was the most recent year for which we had full baseline data available. Although geographic variation in health spending (along with other features of the Medicare population such as their health or propensity to seek care) may have changed between 2008 and the present, we assume that the 2008 baseline data allow for a reasonable approximation of the steady-state effect of the three policy options in any given year.

For each of the three policies, we present the results across different scenarios. These scenarios reflect different options on how the policies could be implemented as well as uncertainty in their impact. The scenarios provide a range by which to gauge the impact of these policies on Medicare spending. Finally, while each of the three policy options required a different modeling approach, for each policy scenario we provide two common outputs - impact on total Medicare spending and on geographic variation in spending. Total Medicare spending is

measured as per Medicare beneficiary spending; geographic variation in spending is captured by the coefficient of variation across HRRs in per beneficiary spending, and the ratio of the 75th to 25th percentiles in per beneficiary spending across HRRs. The common outputs allow the reader to easily compare the impact of the policies.

Our analysis focused on Medicare spending. Though we recognize that other outcomes are important, we chose not to study, for example, the impact of these policies on quality of care, on provider profitability, or spillover effects on patients who do not have Medicare.

In the main body of this report, we describe each of the three policies in a separate chapter which takes the following form: we introduce the policy then provide an overview of our methods, the scenarios modeled, the assumptions behind our methods, the key results, and a brief discussion. A concluding chapter follows the three results chapters. Details of our methods are provided in the appendices.

Chapter 2 Pay for Performance

Introduction

Under a P4P program, providers—such as hospitals, medical groups, or nursing homes—receive greater reimbursement if they attain a high level of performance on quality measures or improve their performance on quality measures by a sufficient degree. Medicare has implemented full or pilot pay-for-performance programs for numerous provider types, including hospitals, nursing homes, dialysis units, and home health agencies (Centers for Medicare and Medicaid Services, undated). We modeled the potential impact of implementing national P4P programs for hospitals, nursing homes, and home health agencies on geographic variation. Our goal was to assess whether these programs would attenuate or increase differences in spending between high- and low-cost regions (but not to make a statement about the effectiveness of the programs in improving quality of care or reducing costs). The potential impact of P4P on geographic variation in spending depends on the underlying relationship between provider performance on quality and spending in HRRs. If high-quality providers are disproportionately located in low-spending HRRs and (conversely) low-quality providers are in high-spending HRRs, then P4P will decrease geographic variation. If the opposite is true, then P4P could increase geographic variation. If there is no relationship between quality and HRR-level spending, then P4P may have no effect on geographic variation. The magnitude of the impact on geographic variation will depend on how much money is allocated as incentive dollars and how that money is allocated.

Brief Methods

We modeled the effects of implementing three Medicare P4P programs that target (1) hospitals, (2) nursing homes, and (3) home-health agencies. The modeled P4P policies were based on existing or pilot Medicare P4P initiatives—specifically, the Hospital Value-Based Purchasing Program, the Nursing Home Quality-Based Purchasing Demonstration and the Home Health Pay-for-Performance Demonstration. However, our model differed from these programs in important ways. For example, two of these Medicare initiatives targeted only a few areas of the country, and in general across the Medicare initiatives the fraction of overall spending devoted to incentives is low. In contrast, we modeled national programs in which a larger fraction of spending was devoted to P4P incentives. We briefly summarize our methods here and details are provided in Appendix 1.

First, we calculated total Medicare payments to each hospital, nursing home, and home health agency provider in 2008 from Centers for Medicare and Medicaid (CMS) claims data. Then, we used publicly available data from Medicare on performance on the quality measures

used in the P4P programs to generate quality scores for each provider. Reflecting existing and prior P4P programs, we measured quality based on both achievement and improvement. To calculate the achievement score, each provider's performance on the quality measures was compared to a benchmark (e.g. more than 95% of hospitalized patients with heart failure were prescribed an ACE inhibitor). To calculate the improvement score, each provider's performance was compared to their own performance from two years prior. Thus, even if a provider performs poorly relative to their peers, the provider may still be eligible for incentive dollars if they exhibit improvement.

We used these quality scores to compute incentive payments that would be paid to that provider under a P4P program. Incentive payments were computed to be budget neutral. To achieve budget neutrality, we subtracted 2% or 15% (conservative or robust scenarios respectively) from total Medicare payments to all providers to create a pool of incentive dollars. These incentive dollars were allocated to higher-quality providers based on either a tournament-style approach or linear exchange curve (explained below). The incentive bonuses or penalties were then applied to baseline Medicare payments in 2008.

We then aggregated providers' baseline Medicare payments and payments under P4P to the HRR level to estimate the effects of the modeled policies on overall, inpatient, nursing home, and home health spending by HRR.

Scenarios and Assumptions

We modeled five scenarios that vary based on several key program design features (Table 2.1). The first is the amount of money allocated to incentive payments as a fraction of overall Medicare reimbursement for each provider type. We modeled two levels of incentives, 2% and 15%. The conservative program (2%) is more consistent with current P4P programs. The robust program (15%) reflects the Committee's desire for a more aggressive P4P program.

The second design feature we varied is how incentive dollars were allocated to providers. We modeled two mechanisms: tournament and linear-exchange curve. In the tournament mechanism, we identified the top 20 percent of providers based on achievement and improvement. Only these top providers received incentive payments. In the linear exchange curve method, we allocated incentive dollars on a continuum. The provider with the worst performance received no incentive payments, and as performance increased, providers received larger and larger payments (allocated as a percentage increase relative to their baseline Medicare reimbursement).

An explicit goal of P4P is to create financial incentives that lead providers to improve quality. Despite widespread interest in P4P, research on CMS demonstrations and private-sector P4P programs has found that P4P programs have had, at most, a modest impact on overall quality (Mehrotra et al., 2009; Rosenthal and Frank, 2006; Lindenauer et al., 2007). However, it is possible that larger payment incentives and permanent programs, rather than demonstrations, will lead to larger effects. To incorporate behavioral effects of the policy (i.e. changes in

quality), under Scenario 4, we assumed there would be an improvement in quality by, on average, 10% across all providers. We assume there would be larger relative improvement among providers with lower initial quality scores and less relative improvement among providers with higher initial quality scores. Last, we varied how we measured provider payments. In Scenarios 1 through 4, we measured each provider's payments, including adjustments for area-level wages and input prices, indirect medical education (IME) and disproportionate share hospitals (DSH). In Scenario 5, we excluded these adjustments to standardize payments across HRRs. More details on how these scenarios were implemented are provided in Appendix 1.

Table 2.1. P4P Scenarios Modeled

	Scenarios				
	1	**2**	**3**	**4**	**5**
Fraction of spending in incentives					
Conservative program – 2%	X	X			
Robust program – 15%			X	X	X
How incentives are allocated					
Only top performers (tournament style)	X				
Linear exchange function		X	X	X	X
Behavioral response					
No	X	X	X		X
Yes				X	
Payment adjustments					
Includes all adjustments	X	X	X	X	
DSH/IME and geographic adjustments not included					X

Beyond quality improvement, P4P could drive other potential behavioral effects. For example, providers could selectively try to care for patients they believe lead to higher quality scores and therefore more incentive dollars (e.g. those with higher socioeconomic status). Also, providers could shift the services they provide. For example, instead of trying to improve quality performance, a hospital might close a heart failure unit if its performance on heart failure quality measures is poor. P4P could also impact spending beyond reimbursement. For example, improved nursing home quality could reduce Medicare spending through fewer readmissions. Such behavioral effects and downstream consequences could influence P4P's impact on total Medicare spending and geographic variation.

While these potential changes are important, we chose not to include these behavioral effects and downstream consequences in our models. The prior literature does not provide sufficient evidence to inform plausible estimates of the size of these behavioral effects or downstream consequences. This lack of a behavioral response in the literature could simply reflect the limited nature of the literature itself, but we felt it was important to ground our assumptions on what is known. Therefore in our models, the impact on Medicare spending was only driven directly by the redistribution of reimbursement via P4P incentives.

To date, P4P programs have generally only included quality measures as a basis for incentives. However, there is growing interest in adding resource or cost measures into P4P programs. We chose not to include such measures in our models because none of the Medicare

initiatives which we use as a basis for our models do so. Also, the use of resource use measures as a basis for incentives has significant overlap with the other two policies (bundled payment and ACOs) we modeled.

Results

Impact of P4P on Providers

The P4P programs resulted in large shifts in payments for individual providers. This was most evident in scenarios where 15% of total payments were tied to incentives. In Table 2.2, we present the distribution of the percentage change in payments for providers under two illustrative scenarios. In Scenario 1 (2% total payments tied to incentives, tournament style), most hospitals received a reduction in payments close to 2%, while the hospitals who received incentives saw a nearly 10% increase in payments. In Scenario 3 (15% total payments tied to incentives, linear exchange curve), the change in payments for hospitals, relative to the baseline payment, ranges from -5.6% (5^{th} percentile) to 6.5% (95^{th} percentile). Because there was more variation in quality scores for home health agencies and nursing homes than for hospitals, there was wider distribution of percent changes in payments for these two types of providers.

Table 2.2. Distribution of Percentage Change in Provider Payments Under Scenarios 1 and 3

A: Scenario 1 --Tournament 2%

Provider Type	5th	25th	50th	75th	95th
Hospital	-2.0%	-1.8%	-1.7%	-1.2%	9.9%
Home Health	-2.0%	-2.0%	-2.0%	-1.1%	10.3%
Nursing Home	-2.0%	-2.0%	-2.0%	-2.0%	8.4%

Note: The percentile columns indicate the percentage of providers facing the given payment change. For example, 5 percent of hospitals receive an increase in payments that is larger than 9.9%.

B: Scenario 3 -- Linear 15%

Provider Type	5^{th}	25^{th}	50^{th}	75^{th}	95^{th}
Hospital	-5.6%	-2.2%	0.0%	2.6%	6.5%
Home Health	-11.5%	-5.4%	-0.4%	4.6%	14.6%
Nursing Home	-11.2%	-6.4%	-1.5%	5.7%	17.8%

Note: The percentile columns indicate the percentage of providers facing the given payment change. For example, 5 percent of hospitals receive an increase in payments that is larger than 6.5%.

In Table 2.3, we estimate changes in payments for the three types of providers in six selected HRRs with a large number of Medicare beneficiaries. We chose three HRRs within California to contrast the impact of P4P on different markets that are similar in their regulatory and competitive environments, but vary in their baseline spending. The other three HRRs were chosen for geographic diversity. Across the HRRs and provider types, the percentage of

providers with an increase in payments varies from 35% to 79%, with a great deal of diversity even within the three California HRRs.

Table 2.3. Percentage of Providers Receiving an Increase or No Change in Payment under P4P for Scenario 3 (Linear 15%)

Type of Provider	HRR	Percent of providers with no change or an increase in payments	Number of Providers in HRR[1]
Hospital	CA - Los Angeles	42%	76
	CA – Sacramento	61%	18
	CA - San Francisco	47%	15
	GA – Atlanta	58%	48
	MO - St. Louis	53%	38
	PA - Philadelphia	49%	37
Home Health	CA - Los Angeles	53%	367
	CA – Sacramento	79%	24
	CA - San Francisco	79%	14
	GA – Atlanta	36%	44
	MO - St. Louis	44%	81
	PA - Philadelphia	61%	54
Nursing Home	CA - Los Angeles	35%	249
	CA – Sacramento	53%	55
	CA - San Francisco	59%	29
	GA – Atlanta	42%	92
	MO - St. Louis	41%	209
	PA - Philadelphia	47%	158
Total (IP, HH, NH)	CA - Los Angeles	45%	692
	CA – Sacramento	61%	97
	CA - San Francisco	60%	58
	GA – Atlanta	45%	184
	MO - St. Louis	43%	328
	PA - Philadelphia	51%	249

[1] Providers are included in Table 2.3 if they participate in P4P and are eligible for an incentive payment. We assigned providers to HRRs based on the modal patients' residence.

Impact of P4P on HRRs

Despite the large shift in payments for individual providers, we saw a much smaller impact of P4P on total spending within HRRs. Table 2.4 displays the HRRs with the smallest and largest reductions in total per-beneficiary Medicare expenditures under one of the more robust scenarios, Scenario 3 (15% total payments, linear exchange curve). Under this scenario the HRRs with the five largest reductions in total per beneficiary Medicare payments received a 1.2 to 2.0% reduction and the HRRs with the largest increases in payments experienced an increase ranging from 1.4 to 1.8%.

Table 2.4. HRRs With Smallest And Largest Reductions In Per-Beneficiary Medicare Expenditures Under Scenario 3

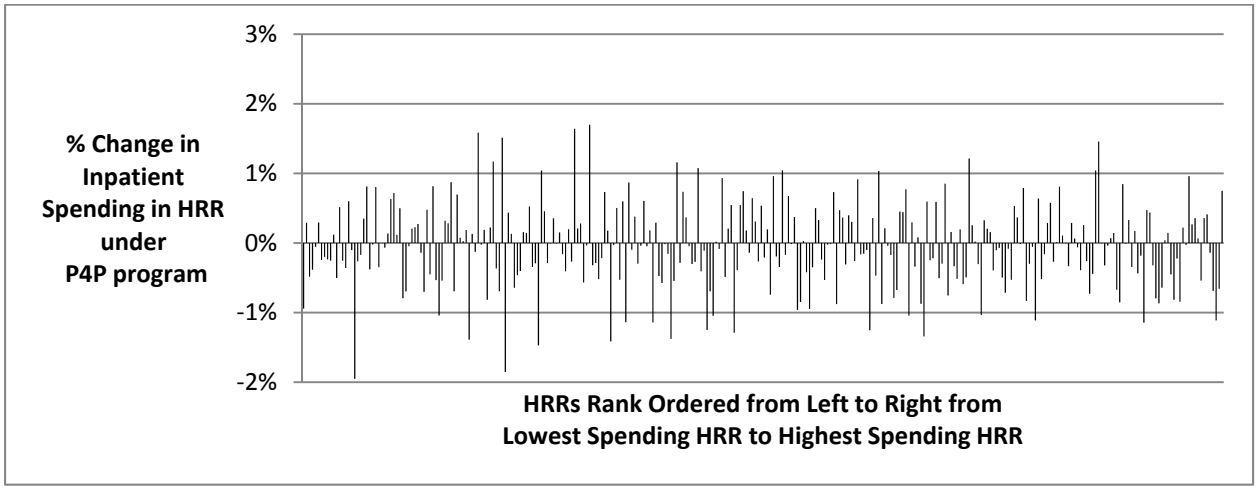
HRRs with largest payment reductions		HRRs with largest payment increases	
HRR	**% change**	**HRR**	**% change**
Bend, OR	-2.0	Hudson, FL	1.8
Grand Forks, ND	-1.8	Traverse City, MI	1.6
Bakersfield, CA	-1.4	Portland, ME	1.5
Bronx, NY	-1.4	Greenville, SC	1.4
Lansing, MI	-1.2	Bangor, ME	1.4

The smaller impact of P4P on HRRs compared to the impact on individual providers can be attributed to several factors. First, within any given HRR there are providers who gain and others who lose under the P4P program, thereby attenuating the overall percent change in HRR spending. Second, we found little correlation between changes in payment in HRRs across the P4P programs for the three provider types. In other words, an HRR that had a large increase in payments under the hospital P4P program was not more likely than any other HRR to have an increase in payments under the home health or nursing home P4P programs. Last, there is a substantial fraction of Medicare spending (e.g., physician services) that is not impacted by the P4P programs.

Impact of P4P on Spending and Geographic Variation in Spending

We use the hospital P4P program to illustrate the impact of P4P on geographic variation in spending. In Figure 2.1 HRRs are ranked from lowest to highest in terms of baseline per beneficiary spending on inpatient care from left to right across the graph. As the figure shows, HRRs with increases or decreases in Medicare spending are roughly evenly distributed from left to right, indicating lack of a strong relationship between the impact of P4P and baseline per beneficiary spending. In the nursing home and home health P4P programs, we also see no relationship between P4P bonus payments and baseline per beneficiary spending.

Figure 2.1. Impact of The Hospital P4P Programs on HRR per Beneficiary Spending on Inpatient Care Under Scenario 3



In Table 2.5, we show variation across HRRs in total per beneficiary Medicare spending at baseline (current reimbursement policies) and under the five P4P scenarios. Because we modeled a budget-neutral program, the P4P program had no impact on per Medicare beneficiary spending. With respect to geographic variation in Medicare spending, we found that there is essentially no impact of P4P as measured by either the coefficient of variation or the ratio of the spending for the HRR at the 75th percentile over the HRR at the 25th percentile.

Table 2.5. Impact of P4P Programs on Total per Beneficiary Medicare Spending and Geographic Variation in Spending

	Actual Medicare Spending					Standardized Medicare Spending	
	Baseline	**Scenario 1 (Tourn, 2%)**	**Scenario 2 (Linear, 2%)**	**Scenario 3 (Linear, 15%)**	**Scenario 4 (Behav, Linear, 15%)**	**Baseline**	**Scenario 5 (Linear 15%)**
Per Medicare Beneficiary Spending (Mean $2008)	9,036.9	9,036.9	9,036.9	9,036.9	9,036.9	8,660.3	8,660.3
Coefficient of Variation (Across HRRs)	0.171	0.172	0.171	0.171	0.171	0.159	0.160
75th/25th (Across HRRs)	1.229	1.227	1.228	1.226	1.226	1.224	1.221

The lack of an impact on geographic variation in spending is driven, in part, by the relatively small portion of total spending in individual HRRs affected by P4P. Perhaps more importantly, there is little underlying relationship between baseline per beneficiary spending in a HRR and the percent gain or loss in spending under P4P for the providers in that HRR.

Association between Performance on Quality Measures and Spending

The impact of P4P on geographic variation in spending depends on whether high performing providers are clustered in high- or low- spending HRRs. In Table 2.6, we show the correlation between performance on inpatient quality measures and inpatient Medicare spending at the HRR level. Across the measures, we found no consistent relationship between quality and spending. We found a similar lack of consistent relationship between quality and spending for the nursing home and home health P4P programs.

In our models we used the quality measures Medicare selected for its P4P programs. There are other quality measures that *could* be used for P4P programs. If measures were selected where there was a stronger and consistent relationship between quality and spending then P4P could potentially have more of an impact on geographic variation.

Table 2.6. Correlation Between the Performance on Quality Measures Used in the Inpatient P4P Program and Inpatient Spending per Beneficiary by HRR

Selected Quality Measures Used in Hospital P4P Program	Correlation Coefficient
Heart failure patients with discharge instructions	0.16
Surgery patients with recommended venous thromboembolism prophylaxis ordered	0.07
Pneumonia patients with appropriate initial antibiotic selection	0.06
Surgery patients with prophylactic antibiotic received prior to surgery incision	0.04
Pneumonia patients with blood cultures in Emergency Department	0.00
Cardiac surgery patients with controlled 6 A.M. postoperative blood glucose	-0.01
Surgery patients with appropriate prophylactic antibiotic selection	-0.10
Surgery patients with prophylactic antibiotics discontinued appropriately	-0.10
Heart attack patients with PCI within 90 minutes of hospital arrival	-0.13
Heart attack patients with fibrinolytic received within 30 minutes of hospital arrival	-0.17

Note: HRRs are the unit of analysis for this correlation. Performance on quality measures is a payment-weighted average of all providers within the HRR.

Impact of Behavioral Response and Standardized Medical Spending

In Scenario 4, we modeled a situation where the P4P program leads to quality improvement. In this scenario, we assumed that hospitals with low-quality at baseline experienced larger increases in quality relative to high-quality hospitals. This assumption meant that there was compression in the variation in performance across the providers. Under this scenario, the magnitude of payment changes under P4P at the HRR level was smaller. However, we do not see an appreciable impact of the behavioral response on geographic variation. In Scenario 5, we used

standardized spending per beneficiary instead of actual spending per beneficiary. In this scenario, the baseline coefficient of variation declines from 0.171 to 0.159. However, under this scenario, we also found that P4P had little impact on geographic variation.

Limitations

Our modeling of the impact of P4P on geographic variation in Medicare spending faces important limitations. For example, we modeled a limited set of quality measures, focusing on quality measures used in current P4P programs. If P4P programs adopted other quality measures that were highly correlated with spending, then P4P could potentially impact geographic variation in spending. Also, we modeled programs that only include incentives that focus on quality of care. This is consistent with the majority of existing P4P programs. If resource use measures were included, the impact of P4P on geographic variation in spending could be very different. In addition, we focus on P4P programs for inpatient, SNF, and home-health care and therefore do not address other important parts of Medicare spending. For example, we did not model the upcoming Medicare P4P program for physicians (i.e. value-based payment modifier) because we lacked data on physician-level quality. It is unclear whether a P4P program for physicians would have a different impact on Medicare spending.

Another limitation is that there are differences in the years of data from which we drew our baseline spending data (2008) and quality performance (we used the most recent quality data available from 2008-2011). Also, we incorporated limited provider behavior in our models. As noted above, P4P could drive improved quality (e.g. decreased readmissions) and spending. This is another avenue whereby P4P could potentially reduce geographic variation. Yet we chose not to include such behavior in our models because of a lack of plausible estimates.

Summary

We found that implementing P4P nationwide for hospitals, nursing homes, and home health agencies would have little impact on geographic variation in total Medicare spending. Under the most robust P4P program that we modeled (i.e., those with 15% of incentives tied to quality performance), many individual providers experienced large increases or decreases in their payments. However, this did not translate into an impact on geographic variation. The primary explanation for this result is the lack of a consistent relationship between performance on the quality measures used in the P4P programs and spending per beneficiary in the HRR.

Chapter 3 Bundled Payment

Introduction

Bundled payment is a method of paying health care providers. Under this method, providers receive a single payment for "bundles" of related health care services across a given episode of care. In contrast, under FFS payment providers receive payments for individual services. Under bundled payment, providers have a financial incentive to provide services for the entire bundle at the lowest cost. They also have more flexibility in the allocation of resources to treat patients most efficiently and effectively. For example, they can use bundled payment revenues to provide services that may not be reimbursed under FFS, such as care coordination services. Bundled payment is being piloted by Medicare in the Bundled Payment for Care Improvement Initiative (BPCI) and by many private health plans. A systematic review of bundled payment programs found weak but consistent evidence that they reduce costs without adversely impacting quality, although previous programs have mostly been more limited in scope than BPCI and other recent bundled payment initiatives (Hussey, et al, 2012).

The effects of bundled payment policies on total Medicare spending depends on the payment rate chosen for the bundles; the policies could be designed to reduce spending, be budget neutral, or increase spending. Bundled payment would reduce geographic variation in spending across HRR if the providers receiving decreases in Medicare payments are clustered in HRRs with high total Medicare spending. On the other hand, if there were no systematic geographic pattern to the changes in payment rates under bundled payment relative to the status quo, no substantial change in geographic variation in health spending would be expected.

Brief Methods

We modeled the effects of a hypothetical *mandatory* Medicare bundled payment program. The main features of the hypothetical program were fashioned after the original design of the Medicare BPCI, a *voluntary* bundled payment initiative.

Details of our methods are provided in Appendix 2. In brief, we defined bundles of care to include all Medicare Part A and Part B services provided to beneficiaries hospitalized during a time window spanning from hospital admission through 30 days post-discharge. As described below, in one alternative scenario, we did not include physician services.

We created bundles for a limited list of conditions. The conditions were chosen on the basis of high historical Medicare utilization for acute and post-acute care. Bundles were created for

each condition as defined by Medicare Severity Diagnosis Related Groups (MS-DRGs).[2] We selected ten conditions (27 MS-DRGs) with high prevalence in the Medicare population (Table 3.1). [3]Together, the services included in bundles for these 27 MS-DRGs accounted for 15 percent of total Medicare Part A and Part B spending in 2008 (American Hospital Directory, 2009).

Table 3.1. Clinical Conditions and Related MS-DRGs Included in the Hypothetical Medicare Bundled Payment Program

Condition	MS-DRGs
Acute Myocardial Infarction	280, 281, 282
Congestive Heart Failure	291, 292, 293
Chronic Obstructive Pulmonary Disease	190, 191, 192
Gastrointestinal Bleed	377, 378, 379
Hip Fracture	480, 481, 482
Kidney/Urinary Tract Infection	689, 690
Lower Extremity Joint Replacement	469, 470
Pneumonia	193, 194, 195
Septicemia	871, 872
Stroke	064, 065, 066

Scenarios and Assumptions

We modeled seven scenarios that primarily varied the method used to determine the bundled payment rates (Table 3.2). The first bundled payment scenario is a single national base rate. The base rate for each bundle (defined by MS-DRG) was calculated as a 5% discount off the national median Medicare payment amount. This scenario is consistent with a fully implemented Medicare prospective payment system for bundled care.

[2] Bundled payment rates were not adjusted for clinical severity beyond the use of MS-DRG-specific rates. This is consistent with the Medicare BPCI, in which applicants have the option to propose additional risk adjustment methods of their choosing but the default is to use MS-DRGs. MS-DRGs were designed to explain variation in the cost of inpatient facility services, but the extent to which they explain variation in cost of modeled bundles of care is unknown.

[3] Specifically, these ten conditions were among the top twenty in terms of volume of hospital stays in FY 2007 and 2008. The ten conditions include the top five conditions in terms of volume; the other five were chosen because they were high volume users of post-acute care (e.g. stroke, hip fracture) or served as important examples of medical conditions (e.g. acute myocardial infarction).

Table 3.2. Bundled Payment Scenarios Modeled

	Scenarios						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Minimum volume threshold for participation							
No minimum volume threshold	X						X
Provider≥10 bundles for each MS-DRG		X	X	X	X	X	
How bundled payment rate will be determined							
National base rate (5% reduction off historical median)	X	X				X	X
Lower of national base rate or hospital-specific historical rate			X				
Blend of national base rate, hospital-specific				X			
Budget-neutral national rate					X		
National rate, acute + post-acute only						X	
Geographic price adjustments							
Includes all adjustments	X	X	X	X	X	X	
DSH/IME and geographic adjustments not included							X

The second bundled payment scenario is identical to Scenario 1 but includes a minimum volume threshold of 10 bundles per year. Any hospital that provides fewer than 10 of a particular MS-DRG would continue to receive status quo payments for that bundle. This scenario is designed to limit the financial risk of bundled payment to low-volume providers.

The third scenario sets a payment rate that is the minimum of: (a) the national base payment rate from Scenario 2 and (b) the provider-specific baseline payment. That is, if providers' baseline payments are above the national base rate, then they receive the adjusted base rate; otherwise they receive their baseline payment. This approach is similar to payments under the Tax Equity and Fiscal Responsibility Act, in which hospitals were paid based on incurred costs subject to a limit.

The fourth scenario uses a blend of 50% of the adjusted national base price (as in Scenario 2) and 50% of the hospital-specific baseline payment. This scenario is similar to periods of transition to Medicare prospective payment systems, where providers have typically been paid a blended payment, with the proportion of the payment for the national base price increasing over time. Therefore, this approach would represent a plausible near-term scenario where Medicare would be in the midst of a transition to a national bundled payment rate.

The fifth scenario is a budget-neutral version of Scenario 2. It uses a national base rate and a minimum volume threshold, but then applies a global adjustment factor so that total payments for bundled services do not change relative to the status quo. In contrast to the other scenarios, in Scenario 5 we are assuming that Medicare will choose to implement a bundled payment program that does not reduce overall spending.

The sixth scenario includes only the index hospitalization plus post-acute care and additional hospitalizations occurring within 30 days of acute discharge. This scenario reflects a bundled payment program that includes facility-based providers, preserving status quo payment methods for physician professional services. This approach recognizes that a facility-based program may be more feasible than a program including physicians, while achieving many of the benefits, since the majority of spending in bundles is for facility-based services.

Finally, the seventh scenario is equivalent to the first scenario (national rate without a volume threshold), but unlike the other six scenarios, prices are standardized to remove payment amounts related to area input prices, medical education, and disproportionate share hospital status on geographic variation in spending. This scenario allows one to distinguish geographic variation in spending that is due to the volume of bundles versus geographic variation in payment rates per bundle.

Bundled payments would significantly change payments to providers, with the intent of creating incentives for providers to deliver care more efficiently while maintaining quality. However, bundled payments could also have unintended effects on provider behavior. Potential unintended behavioral effects include: selection of patients on the basis of risk factors; maximizing the number of episodes by recommending bundles of services for patients who may be treated by other modalities; "up-coding" the severity of patients to receive a larger payment; and shifting care to settings or time periods not covered under the bundle in order to qualify for additional reimbursement. In addition, intended effects may include changing the provision of services used within the bundle, which will have differential impact on providers (e.g., hospital spending may remain constant, but home health agency spending could drop) and beneficiary out-of-pocket payments but not Medicare payments per bundle.

Such behavioral changes would alter the effect of bundled payment on provider profit margins, total Medicare spending, and geographic variation. However, the prior literature does not provide sufficient evidence to inform plausible estimates of the size of these behavioral effects, and particularly how these effects would vary across geographic areas. Reflecting this, we made the following simplifying assumptions: (1) providers may change the mix and quantity of services inside the bundle, but these changes will not affect the amount Medicare pays; (2) providers will not change the volume or mix of bundles provided; and (3) providers will not change the mix and quantity of services used outside the bundle.

Results

Impact of Bundled Payment on Providers

The "provider" under a bundled-payment model includes the hospital and all other providers (e.g., nursing homes, home health agencies) involved in the patient's care. The provider-level impacts of bundled payment on Medicare payments would vary considerably, because both the

historical volume of bundled episodes and average Medicare spending per-bundle varies widely across providers. For example, for Acute Myocardial Infarction (AMI, MS-DRG 280), Medicare paid 45 percent more per episode ($24,113 vs. $16,583) to the 75th percentile provider compared to the 25th percentile provider (in percentiles of the baseline spending distribution). For lower-extremity joint replacement (LEJR, MS-DRG 470), Medicare paid 34% more per episode ($23,257 vs. $17,376) for the 75th vs. 25th percentile provider[4].

Table 3.3 shows the distribution of provider-level changes in Medicare payments under a hypothetical bundled payment program (Scenario 2). Because the distribution of costs per bundle is skewed towards high-cost bundles, under a hypothetical bundled payment program, providers are more likely to see a large reduction in payments instead of a large increase in payments. For example, under a bundled payment scenario for AMI, one fourth of providers receive a payment reduction of 10.4% or greater, while half of providers receive no change or an increase in payment. For LEJR under the bundled payment scenario, one quarter of providers receive a reduction of 12.9% or greater, while another one quarter of providers receive a payment *increase* of 3.1% or greater.

Aggregating across all 10 conditions in the hypothetical bundled payment program (again under Scenario 2), we estimate that the median provider would receive a 9.7% decrease in payments. In Scenario 2, the payment rate was set at a 5% discount off the national *median* Medicare payment amount and therefore the majority of providers would receive a decrease in payments. Thus, in this scenario, even if a provider receives increased Medicare payments for a particular MS-DRG, the overall change in payment for all MS-DRGs included in the bundled payment program is likely to be negative. This result is consistent with limited within-provider correlation in spending across the 10 conditions (results not shown). Nevertheless, some providers *would* receive increased reimbursement relative to historical payments under bundled payment.

Table 3.3. Distribution Of Provider-Level Payment Changes Under Bundled Payment Scenario 2

	Percentile of provider-level payment change nationally				
MS-DRG	**5th**	**25th**	**50th**	**75th**	**95th**
280 (AMI)	-29.5%	-10.4%	0.0%	0.0%	7.9%
470 (LEJR)	-27.6%	-12.9%	-1.8%	3.1%	24.7%
All MS-DRGs	-26.1%	-16.1%	-9.7%	-2.2%	3.3%

[4] MS-DRG 280: Acute Myocardial Infarction (AMI) discharged alive with MCC and MS-DRG 470: Major joint replacement or reattachment of lower extremity w/o MCC. We display these MS-DRGs because they represent the highest volume MS-DRGs within high volume medical (AMI) and surgical (joint replacement) hospitalizations.

In Table 3.4, we estimate changes in payments for bundled episodes for providers in six HRRs under Scenarios 2 (base rate set at 5% reduction off of national median) and 5 (budget neutral base rate) relative to historical payments. Under Scenario 2, a large majority of providers receive a reduction in payments in Los Angeles, St. Louis, and Philadelphia. However, more than 1/3 of providers in Sacramento and San Francisco receive an increase or no change in payments. In contrast, under Scenario 5, many more providers receive an increase or no change in payments, including the majority of providers in Sacramento, San Francisco, Atlanta, and St. Louis. This example highlights that the fiscal impact of bundled payment on providers is sensitive to how payment rates are set.

Table 3.4. In Six Selected HRRs, the Percentage of Providers Receiving Either an Increase or No Change in Medicare Payment Under Bundled Payment Scenarios 2 and 5

HRR	Number of hospitals in the HRR[5]	Percent of providers with no change or an increase in payments	
		Scenario 2: National base rate (5% reduction off of national median)	Scenario 5: National base rate (Budget neutral)
CA - Los Angeles	90	14%	41%
CA - Sacramento	20	35%	90%
CA - San Francisco	16	38%	69%
GA – Atlanta	59	22%	83%
MO - St. Louis	60	10%	57%
PA - Philadelphia	39	3%	31%

In response to these changes in payments, we would expect providers to change the provision of care to reduce the cost of producing bundles.[3] For example, providers could make efforts to substitute lower-cost post-acute care alternatives or reduce the number of specialist physician consultations provided during bundles. If providers were able to reduce the costs of producing bundles by an amount greater than any reduction in Medicare payments for the bundles, they would be able to increase profit margins. Under current policy, in which separate Medicare prospective payment systems are used for each provider type, there is little incentive for efficient use of services within the bundle.

To illustrate the potential for providers to reduce the costs of producing bundles by changing the number and type of services provided in bundles, we compared utilization patterns between high-cost and low-cost providers (Table 3.5). These calculations show the potential effects of bundled payment if providers with high-baseline payments for bundles changed the provision of

[5] While a variety of entities could theoretically be the recipient of a bundled payment (e.g. acute care hospital, physician group, post-acute provider, etc.), we assume that the hospital from which the patient is first discharged (i.e., the provider of the index hospital stay) receives the bundled payment.

services to match the patterns of providers with historically low-baseline payments under bundled payment.

Table 3.5 compares the amount of spending on the index hospital stay, post-acute care, and hospital readmissions between providers in the highest and lowest quartile of historical spending for services that would be bundled.[6] For AMI (Panel A), total episode spending was almost $8,000 lower in the low spending quartile than in the high spending quartile. Medicare spending on readmissions comprised 70% of the total difference, compared to 4% for the index hospital stay and 26% for post-acute care. In the top quartile, 40% of patients are readmitted compared to only 20% in the lowest quartile. Thus, reducing readmissions may be a potential way for acute care hospitals with high historical spending per episode to reduce costs and increase margins under a bundled payment program. For lower extremity joint replacement (Panel B), Medicare spending per episode was over $5,600 higher for the highest spending quartile than the lowest quartile. In contrast to AMI, this difference was driven primarily by spending on post-acute care, which averaged $8,813 in the high cost quartile and only $3,690 in the low cost quartile. As a result, more efficient use of post-acute care may preserve the profit margins of acute care hospitals in the highest cost quartile under bundled payment.[7]

Table 3.5. Spending by Category of Services That Would Be Included in Bundles for Selected MS-DRGs, by Quartile of Spending, 2008

	Highest spending quartile	Lowest spending quartile	Difference (% of total difference)
A. Acute Myocardial Infarction (AMI) (MS-DRG 280)			
Total	$21,863	$14,012	$7,851 (100%)
Index acute care stay	$10,011	$9,713	$298 (4%)
Post-acute care	$4,591	$2,537	$2,054 (26%)
Readmissions	$7,261	$1,761	$5,500 (70%)
B. Lower extremity joint replacement (LEJR) (MS-DRG 470)			
Total	$20,584	$14,941	$5,643 (100%)
Index acute care stay	$10,840	$10,815	$25 (0%)
Post-acute care	$8,813	$3,690	$5,123 (91%)
Readmissions	$931	$436	$495 (9%)

[6] We only include providers with volume of 10 or more bundled episodes. In order to focus on the effects of utilization patterns, not Medicare payment rate differences, we standardized prices using the method outlined in Gottlieb et al (2010).

[7] These differences could be driven in part by differential patient severity *within* MS-DRG across providers.

Impact of Bundled Payment on Spending and Geographic Variation in Spending

The predicted impact on Medicare spending and geographic variation in Medicare spending for *bundled services* varies across the seven scenarios modeled (Table 3.6). Under six of the seven scenarios, predicted Medicare spending would decrease relative to the base case, by an amount ranging from 6% (Scenario 4, the blended rate) to 23% (Scenario 7, a price-standardized national rate). This is by design, because the bundled payment rates were set at a lower level relative to average baseline payments. The exception is Scenario 5, which is budget neutral by design.

Variation across HRRs in spending on bundled services would decrease under every scenario relative to the base case. In Scenarios 1-5, the coefficient of variation ranges from 0.124 (Scenario 2, national base rate with volume threshold) to 0.139 (Scenario 4, blended rate), compared with 0.158 in the base case. Scenarios 6 and 7 also decrease variation to a similar degree.

Table 3.6. Impact of Bundled Payment on per Beneficiary Spending and Geographic Variation in Spending for Bundled Services

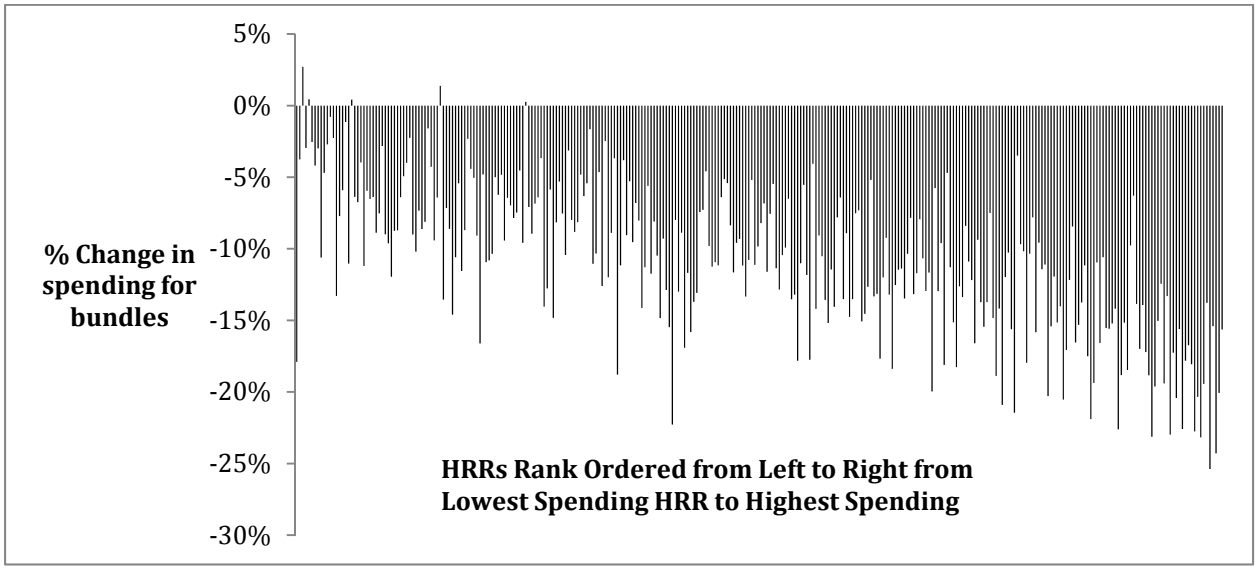
	Scenarios									
	B1-5	**S1**	**S2**	**S3**	**S4**	**S5**	**B6**	**S6**	**B7**	**S7**
Spending on bundled services per beneficiary (% decrease)	$1,369	$1,172 (14%)	$1,193 (13%)	$1,174 (14%)	$1,281 (6%)	$1,369 (0%)	$1,180	$1,013 (14%)	$1,241	$958 (23%)[8]
Coefficient of variation (across HRRs)	0.158	0.126	0.124	0.128	0.139	0.131	0.162	0.127	0.149	0.122
75th/25th (Across HRRs)	1.211	1.155	1.157	1.160	1.178	1.165	1.221	1.161	1.222	1.182

B1-5 Base case for Scenarios 1-5
S1 National rate, no volume threshold
S2 National rate, volume ≥10
S3 Lower of national and historical rate, volume ≥10
S4 Blended rate, volume ≥10
S5 National rate, budget neutral, volume ≥10
B6 Base case for Scenario 6
S6 National rate, acute and post-acute only, volume ≥10
B7 Base case for Scenario 7
S7 Price standardized national rate, no volume threshold

[8] The payment reduction is larger in S7 than the non-standardized scenarios. In both the S7 and S1, we set a base national rate at 0.95 of the median of historical payments. However, in S1 we then apply price adjustments to the base national rate that are, on average, greater than 1. Thus in practice, the national rate is greater than 0.95 of the median of historical payments, while in the standardized scenario the national rate is exactly 0.95 of the median of historical payments.

The coefficient of variation in spending for bundled services in Table 3.5 decreases under the bundled payment scenarios because bundled payment would result in larger payment decreases in HRRs with higher base case spending levels (Figure 3.1). In the figure, each bar represents an HRR, and the bars are ranked from left to right in order of increasing base case spending for bundled services per beneficiary. Under the assumptions of Scenario 2, bundled payment would result in decreased spending for bundled services in most HRRs. The percentage decrease tends to be larger at the right side of Figure 3.1, among HRRs with higher base case spending.

Figure 3.1. Impact of Bundled Payment on HRR Per Beneficiary Spending for Bundled Services under Scenario 2



The bundled payment scenario would also be expected to decrease *total* Medicare spending per beneficiary, and the variation in total spending across HRRs (Table 3.7). Because the bundled payment scenario only affects 15 percent of total spending, the degree of change for total Medicare spending is smaller than the degree of change in spending for bundled services only. Total Medicare spending and spending for bundled services were highly correlated (data not shown). As a result, the direction of the effects is similar for total Medicare spending and spending for bundled services.

Table 3.7. Impact of Bundled Payment on Total per Beneficiary Medicare Spending and Geographic Variation in Spending

	Scenarios									
	B1-5	**S1**	**S2**	**S3**	**S4**	**S5**	**B6**	**S6**	**B7**	**S7**
Spending on bundled services per beneficiary (% decrease)	$9,037	$8,844 (2%)	$8,865 (2%)	$8,845 (2%)	$8,951 (1%)	$9,037 (0%)	$9,037	$8,874 (2%)	$8,660	$8,383 (3%)
Coefficient of variation (across HRRs)	0.171	0.167	0.166	0.168	0.169	0.165	0.171	0.167	0.159	0.156
75th/25th (Across HRRs)	1.229	1.213	1.213	1.218	1.222	1.214	1.229	1.214	1.224	1.215

B1-5 Base case for Scenarios 1-5
S1 National rate, no volume threshold
S2 National rate, volume ≥10
S3 Lower of national and historical rate, volume ≥10
S4 Blended rate, volume ≥10
S5 National rate, budget neutral, volume ≥10
B6 Base case for Scenario 6
S6 National rate, acute and post-acute only, volume ≥10
B7 Base case for Scenario 7
S7 Price standardized national rate, no volume threshold

Limitations

These estimates are subject to several key limitations. First, the scenarios modeled include a strong set of behavioral assumptions, in particular that providers will not change the volume or mix of bundles and that providers will not change the mix and quantity of services used outside the bundle. Past experience with Medicare payment reform shows that providers respond adeptly to financial incentives, and thus it is likely that the volume and mix of bundles and services used outside the bundle will change, with resulting effects on total Medicare spending and geographic variation in Medicare spending that are not captured in our analysis. However, past experience does not provide enough information to support estimates of the size of these effects across geographic areas. Second, the changes in Medicare spending and geographic variation we estimated are sensitive to the bundled payment scenarios that we imposed, including the clinical conditions subject to bundled payment, the providers included in the bundled payment program, the method for determining payment rates, etc. It is likely that an implemented bundled payment system would differ from what we modeled in the analyses, leading to different effects on spending and variation.

Summary

Our results demonstrate that bundled payment has the potential to reduce geographic variation in Medicare spending. Because bundled payment, by design, reduces variation in the

amount paid for services in the bundle, the size of the effect is largely determined by how many services are included in the bundle and how bundled payment rates are determined. Some geographic variation in spending would persist, because of variation in the number of bundled episodes per beneficiary. Geographic variation in the number of bundled episodes per beneficiary could reflect either geographic variation in level of illness or variation in treatment patterns. In our non-price standardized scenarios (Scenarios 1-5), some of the residual geographic variation under bundled payments reflects Medicare payment rates adjustments for area-input prices.

Our findings also suggest that there is considerable variation across HRRs and across providers in treatment patterns within conditions. An important remaining question is how use of services within the bundle will change after bundled payment, and specifically whether bundle recipients will converge on more efficient and effective approaches for acute and post-discharge care.

Chapter 4 ACOs

Introduction

Medicare created ACOs as part of the Affordable Care Act. ACOs use the concept of shared savings to encourage providers to improve the value of care. In brief, provider organizations apply to become ACOs. This designation means that the organizations assume responsibility for the total costs of care for assigned Medicare FFS beneficiaries. If Medicare costs for assigned beneficiaries are below a target, Medicare would pay the provider organizations bonus payments (and in some cases, the organization would pay Medicare if costs exceeded the target). These bonus payments are a fraction of the amount by which total costs fall below the target (hence, the reduction is 'shared' between Medicare and the organization). The bonus payments are reduced if certain quality standards are not met.

Medicare has allowed for several variations within the ACO program – generally along the lines of organizations' bearing greater risk as they gain more experience with the program. ACOs under the Pioneer program face both 'upside' and 'downside' risk – that is, they would receive bonus payments if spending is below targets (upside risk), but they must pay CMS if spending exceeds the target (downside risk). ACOs under the Medicare Shared Savings Program (SSP) face only upside risk initially but are eventually expected to take on downside risk after a few years of experience with the program. Currently there are 32 Pioneer ACOs and 116 SSP ACOs.

In our models, the impact of ACOs on overall Medicare spending is dependent on our assumptions about how providers respond to the ACO program (described below), and that some providers in areas without observed ACOs thus far would be capable of overcoming barriers to form ACOs. The impact of ACOs on geographic variation in spending depends primarily on where ACOs form. If ACOs do indeed achieve cost savings (which is assumed), and are more likely to form in higher-cost areas, then geographic variation in Medicare spending could be reduced.

Brief Methods

Our models of the impact of ACOs on geographic variation required (1) estimating the number of Medicare FFS beneficiaries associated with each ACO, (2) assigning beneficiaries associated with each ACO to an HRR, (3) projecting where future ACOs might form (in those scenarios where we assumed expansion of ACO program), and (4) estimating Medicare savings associated with each assigned beneficiary.

The methods are described in more detail in Appendix 3. In brief, to determine the number of Medicare FFS beneficiaries with each ACO, we either obtained actual numbers from Medicare, allocated beneficiaries in proportion to the number of primary care physicians associated with the ACO, or assigned an average number of beneficiaries per ACO. Each ACO had a minimum of 5,000 beneficiaries. Primary care physicians were identified using publicly available data.

The second step involved assigning Medicare FFS enrollees to an HRR. To do so, we used a variety of methods including using data directly from CMS or assigning beneficiaries based on where the primary care physicians in the ACO were located. In the third step, we predicted future participation in ACOs for each HRR. Using current ACO participation, we identified factors in the HRR that were associated with greater ACO participation. Based on the regression coefficients, a 'likelihood' of ACO penetration was predicted for each HRR. Those predictions were used to project additional ACO penetration in HRRs until 20% of all FFS beneficiaries, nationally, were represented.

In the fourth step, across the scenarios, we used different estimates of the savings per beneficiary. These estimates were based on estimates in the literature and varied between Pioneer and SSP ACOs because they bear different amounts of risk.

Scenarios and Assumptions

We modeled the impact of ACOs on Medicare spending and geographic variation in spending under five scenarios (Table 4.1). Across the scenarios, we vary our estimates of how much ACOs will reduce Medicare spending and the participation in ACOs.

Current enrollment in ACOs (as of August, 2012) represents approximately 7% of FFS beneficiaries. In Scenarios 1 and 2 we assume that ultimate participation *declines* from that level to settle at 5% of FFS beneficiaries - based on the idea that a fraction of ACOs may fail because of difficulties with the model and lack of enthusiasm among providers. To implement this 'failure' rate, we subtracted a random portion of observed ACO enrollment to reach the 5% target participation rate. Under the high participation scenarios (Scenarios 3, 4, and 5) we assume future *growth* in existing or new ACOs to the extent that the ACO program will ultimately include 20% of Medicare FFS beneficiaries. To reach that level, we used a two-fold approach. First, we added existing private-sector ACO-like entities that we were able to identify. Their addition increased Medicare FFS participation to roughly 10%. We attributed Medicare beneficiaries to these ACOs using the same methods as for Medicare ACOs. Second, we projected additional participation using the regression-based approach described above to reach our goal participation level of 20% of Medicare FFS beneficiaries. In this step we assume that the characteristics of an HRR that are associated with the formation of a current ACO will also predict where future ACOs will form.

The scenarios also varied in the assumed savings experienced by Medicare as a result of ACOs. Both the Congressional Budget Office and CMS estimate that ACOs would save

Medicare through changes in provider behavior that reduce expenditures on the part of enrolled beneficiaries that outweigh, on net, the shared savings bonus payments (Congressional Budget Office, 2008a; Centers for Medicare and Medicaid Services, 2011). Those savings amount to roughly 1% per enrolled beneficiary (which implies a substantially larger reduction in overall expenditures because Medicare retains only a minority of the overall savings while the organization retains the remainder). This level of savings is consistent with other situations when providers face risk of total costs (Gaynor, Rebitzer and Taylor, 2004; Song et al., 2012).

Across the scenarios, we used low and high-savings scenarios that are based on this 1% savings estimate. We varied the savings rate, because of the uncertainty surrounding these savings estimates. For the SSP ACOs we assumed a 0.5% savings in the low scenario and 3% savings in the high scenario. For the Pioneer ACOs we assumed a 1% savings in the low scenario and 5% savings in the high scenario. We assumed a higher savings estimate for the Pioneer ACOs due to the higher risk faced by those organizations (though we remain agnostic about whether the higher savings are a direct result of the greater risk, or whether organizations with a greater capacity to achieve savings self-select into the Pioneer program).

Table 4.1. ACO Scenarios Modeled

	Scenarios				
	1	**2**	**3**	**4**	**5**
% of Medicare FFS beneficiaries enrolled					
Low: 5%	X	X			
High: 20%			X	X	X
Medicare savings per enrolled Beneficiary					
Low: 0.5% (Shared Savings); 1%, (Pioneer)	X				
High: 3% (Shared Savings); 5% (Pioneer)		X	X		
Scaled: High x (HRR per beneficiary spending relative to mean)				X	X
Geographic price adjustments					
Includes all adjustments	X	X	X	X	
DSH/IME and geographic adjustments not included					X

Finally, we employed one additional important variation in the expected reduction in Medicare spending by ACOs. The ‘scaled savings’ scenarios (Scenarios 4 and 5) employed the assumption that a *higher* degree of savings is expected among ACOs in higher-cost areas. This is a distinct possibility if high-cost areas have a relatively high proportion of beneficiaries with

poor health status or inefficient practices, affording more opportunities for reductions in spending.[9] Note that this assumption, by itself, would result in a reduction in the geographic variation in Medicare spending even if ACO formation were evenly distributed across the US; the higher-cost areas would face a bigger spending reduction than the lower-cost areas. In these scenarios, the default savings rate was multiplied by the ratio of per beneficiary spending in the HRR of the ACO to overall mean spending across the US.

Finally, in Scenario 5 we estimated savings relative to standardized baseline spending rather than unadjusted baseline spending.

Results

How Many Enrollees Are in ACOs?

Table 4.2 summarizes our estimates of the number of Medicare FFS in each grouping of ACOs. In sum, they comprised 3.4 million FFS beneficiaries, or roughly 10% of Medicare FFS beneficiaries. The Pioneer ACOs have a higher number of beneficiaries per ACO – which is a plausible finding if larger, more established and integrated organizations with greater experience handling risk have disproportionately enrolled in the Pioneer ACO program.

Table 4.2. ACO Enrollment Groups and Associated Number of Beneficiaries, as of October 2012

Type	Number of ACOs	Total assigned enrollees	% of Medicare FFS beneficiaries*	Average number of enrollees per ACO
Pioneer ACOs	32	755,878	2.3%	23,621
SSP ACOs	116	1,575,000	4.7%	13,578
Private ACOs**	77	1,047,112	3.1%	13,598
Total	225	3,377,990	10.1%	15,013

* As in all scenarios, results are reported relative to the 2008 baseline.
**Private ACOs were organizations from the following groups (excluding some organizations such as insurer-based groups that were very dissimilar to current Medicare ACOs): Dartmouth-Brookings ACO learning network, Council of Affordable Physician Practices, CMS physician group demonstration participants, Premier's Partnership for Care Transformation Implementation Initiative.

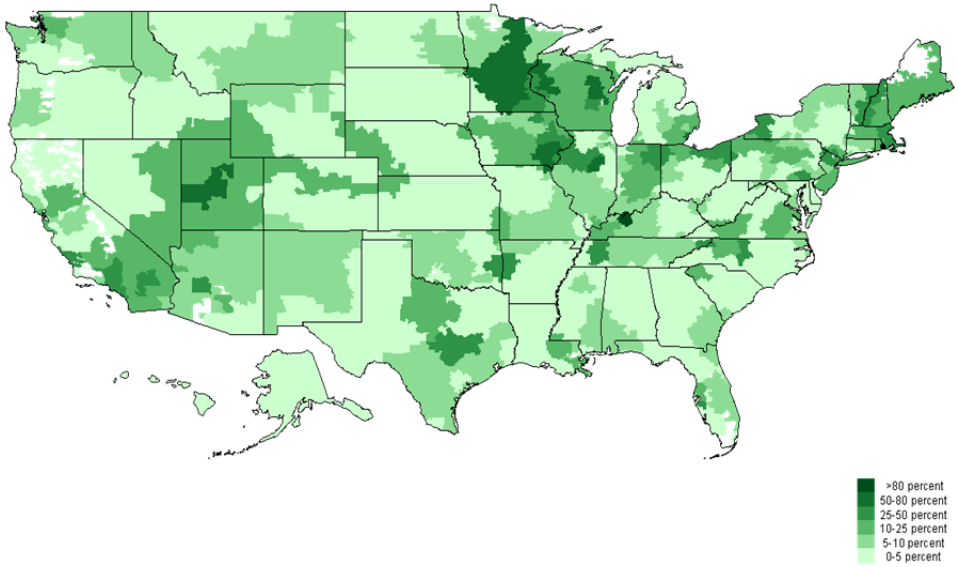
Geographic Distribution of Medicare FFS Beneficiaries in an ACO

Figure 4.1 shows the resulting distribution of the above enrollees by HRR. Penetration in ACOs is distributed very unevenly by HRR – 8 of 306 HRRs have more than 50% of their

[9] For example, much of the savings observed in the Alternative Quality Contract, a private ACO-like entity in Massachusetts, was achieved via referral to lower-cost providers and use of less expensive services (Song et al., 2012).

Medicare FFS beneficiaries enrolled in an ACO[10]; 31 HRRs have between 25 and 50% enrolled; 119 have between 5 and 25% enrolled and 148, roughly half of the HRRs, have less than 5% enrolled.

Figure 4.1. Proportion of Medicare FFS Beneficiaries in HRRs Enrolled in Medicare or Private ACOs



Characteristics of HRRs with an ACO

We used the HRR-level participation in ACOs as the dependent variable in a regression model to understand what characteristics of the HRR predicted formation of an ACO. Details of the regression are described in Appendix 3. In brief, we found that the following HRR characteristics were associated with ACO formation: more hospitals affiliated with a health system, more physicians in large groups, higher population density, higher income, more physicians accepting Medicare patients, a higher Medicare Advantage penetration rate, and a higher percentage of hospital revenue from capitation or risk-sharing contracts. HRRs with more physicians per beneficiary were less likely to have an ACO. These measures, taken together, explained roughly 20% of the observed variation in ACO penetration by geographic area – thus, there may well be other factors which we do not observe that also predict where ACOs have formed.

[10] While some of the very-high participation HRRs may be due to our assignment algorithm, which was a mere approximation of beneficiary locations in some cases with limited information (see Appendix 3), the overall findings are robust to alternative assumptions about these cases.

Impact of ACOs on Geographic Variation and Medicare Spending

For ACOs to reduce geographic variation in Medicare spending, they must be preferentially located in high-cost areas of the US. Figure 4.2 displays the relationship between ACO penetration in HRRs and the mean per-beneficiary Medicare cost in the HRR in 2008. There is slightly higher ACO penetration in higher-cost HRRs but the relationship is weak (correlation coefficient is .05 (p=.34)). Because of that relatively weak relationship between ACO formation and high spending areas, ACOs are unlikely to have a large impact on geographic variation in Medicare spending.

Figure 4.2. Relationship Between HRR per-Beneficiary Medicare Costs and ACO Penetration

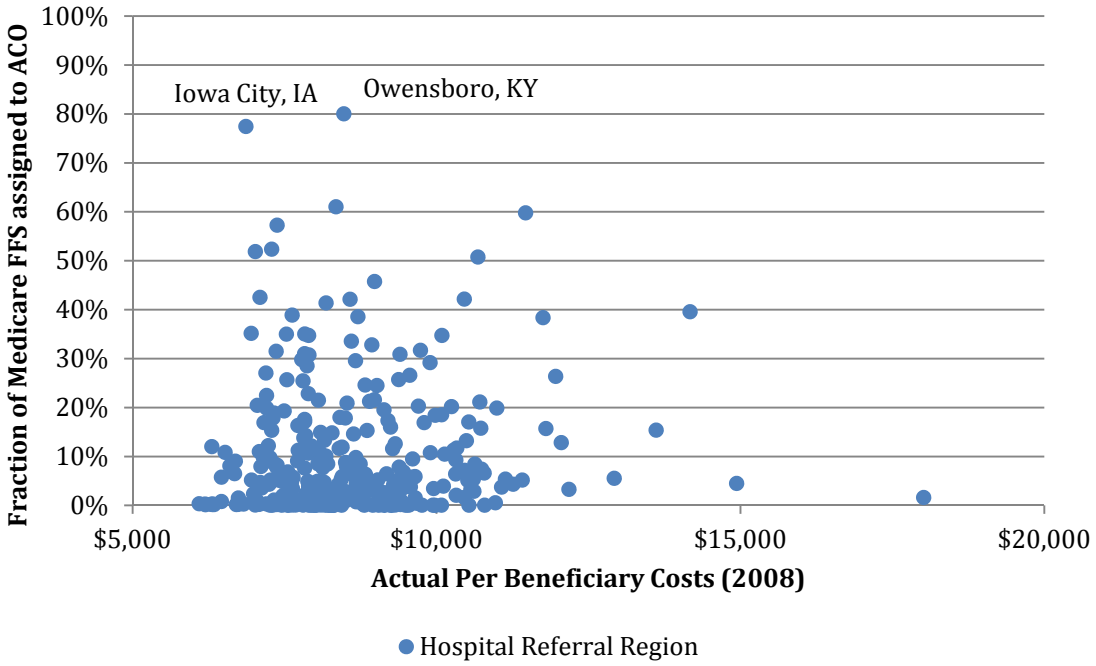
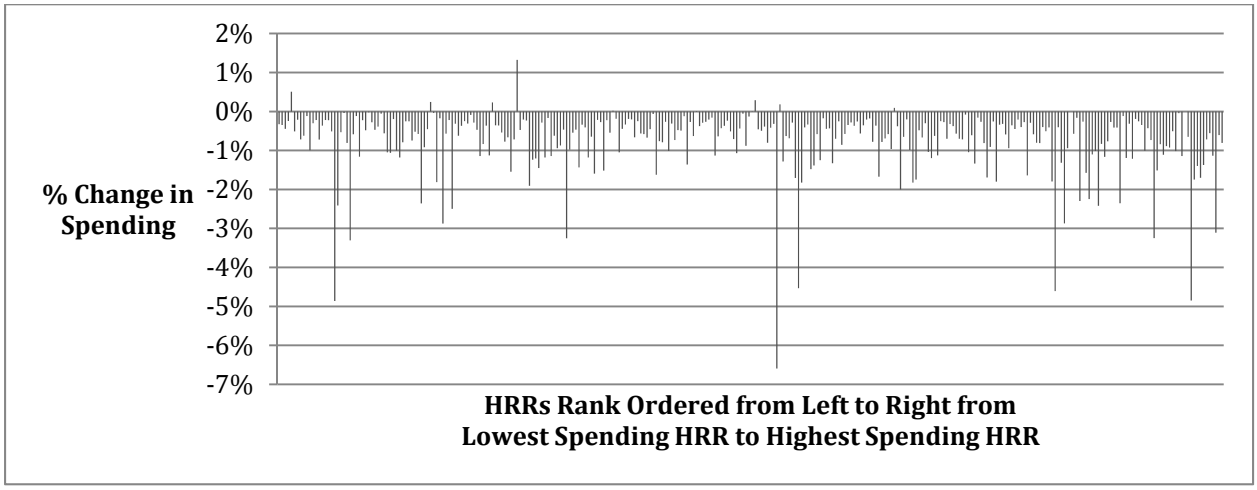


Figure 4.3 displays the relationship between Medicare per beneficiary costs (where HRRs are lined up from left to right in order of increasing costs) and expected reductions in Medicare spending for each HRR under Scenario 4 (20% FFS participation and scaled savings). The bars show the percentage change in Medicare spending for each HRR relative to the 2008 baseline – which range from an increase in Medicare spending of 1% to a decrease of nearly 7%. The magnitude of savings is mainly a function of ACO penetration in the HRR though they are also affected by local trends in Medicare spending in prior years (see Appendix 3 for more details).

Though we did find a set of factors associated with higher ACO penetration (noted above), there was no obvious pattern or similarity across the highest-penetration HRRs. Critically for analysis of geographic variation, even in Scenario 4 (shown in Figure 4.3) which is the most

likely to reduce geographic variation because of the scaled savings, the relationship between Medicare savings in an HRR and baseline per beneficiary costs is not strong.

Figure 4.3 Impact of ACOs on HRR per Beneficiary Spending under Scenario 4[11]



In Table 4.3, we present the impact of ACOs on total Medicare spending and geographic variation in spending. Total Medicare costs decrease in all scenarios, unsurprisingly, as savings were assumed with ACO penetration. Overall savings to the program range from negligible to as high as approximately .8% in the high-penetration scenarios. The effect on geographic spending is very small. The Coefficient of Variation in HRR-level Medicare spending is reduced from .1714 to .1705 under the Scenario 3 (high savings rate, high penetration) and further to .1699 under Scenario 4 (high savings rate, higher penetration, scaled savings). Those reductions are mirrored in the ratio of the 75th to the 25th percentile.

The scenario using standardized spending as the 'baseline' rather than actual per beneficiary costs results in a smaller reduction in geographic variation compared its otherwise equivalent scenario (Scenario 4) that uses the same assumptions. That result suggests that some of the small degree of association between ACO locations and high-cost areas was due to association with the aspects of those areas that are high cost, including high prices, teaching costs, input costs or other adjustments that are removed in the standardization process.

[11] A few HRRs show an increase in costs due to the ACO policy – in those rare cases, the ACOs in the region are expected to experience extremely low cost growth trends such that they end up spending far below their targets (without deliberately changing anything about their care delivery). Some of the distance between their actual spending and their target is due to real reductions in expenditures due to behavioral effects, yet some is due to the secular spending trend in their area beyond any behavior on the part of the ACO. On net, in these few cases, Medicare shared savings payments exceed the true reduction in Medicare expenditures due to behavior and the ACO ends up costing Medicare.

Table 4.3. Impact of ACOs on Total per Beneficiary Medicare Spending and Geographic Variation in Spending

	Scenarios						
Description	**Baseline**	**1**	**2**	**3**	**4**	**Baseline**	**5**
	(Actual costs)	(low pen, low savings)	(low pen, high savings)	(high pen, high savings)	(scaled savings)	(Standardized costs)	(scaled savings, standardized costs)
Total Spending per beneficiary ($, % Decrease)	$9,037	$9,031 -0.07%	$9,018 -0.21%	$8,965 -0.80%	$8,964 -0.81%	$8,660	$8,593 -0.77%
Coefficient of Variation	0.1714	0.1711	0.1711	0.1705	0.1699	0.1594	0.1592
75th / 25th Percentile	1.2289	1.2277	1.2246	1.2287	1.2263	1.2242	1.2222

Limitations

There are several limitations to our modeling of the effect of the current and a potentially expanded Medicare ACO program on Medicare spending and geographic variation in spending. Our process for estimating where ACO enrollees reside is far from perfect because we did not have direct data on these locations for all ACOs. Second, we projected future enrollment in an expanded program based partly on approximations of locations of private-sector ACOs – reasoning that the same factors leading to their formation could also lead to formation of Medicare ACOs in the future. Yet these entities are not necessarily the same types of entities as Medicare ACOs. Our projections were also based on assumptions that the area characteristics associated with ACO formation predict where future ACOs will form – there could well be other factors at play such as provider-payer relationships or market concentration which we were not able to observe or measure in our model. Finally, we made several assumptions about provider behavior in response to the incentives embodied in the ACO structure – assumptions which may prove inaccurate.

Summary

We estimate that 7% of Medicare FFS beneficiaries are in an ACO currently and that several characteristics of HRRs (e.g., percentage of physicians in large groups and population density) are predictive of ACO formation. However, we find little relationship between per-beneficiary Medicare costs in a HRR and ACO formation. Therefore we see no substantive impact of ACOs on geographic variation in spending.

There are several potential explanations for this lack of relationship. Possibly, competing factors are at play. For example, participation might be expected to be greater in *low*-cost areas if

factors such as infrastructure favorable to formation of ACOs (such as integrated networks, less-fragmented care and experience with risk-based contracts) are also associated with low-health care costs. That would exacerbate geographic variation since we expect ACOs to save Medicare money, on net. On the other hand, participation might be greater in high-cost areas if organizations perceive that there is more 'room' to cut costs and achieve shared savings in high-cost areas. In any case, because of the relatively low participation in ACOs combined with the relatively small expected impact on Medicare spending, the ultimate effect on geographic variation is likely to remain small.

Chapter 5 Conclusions

In Table 5.1, we summarize the impact of the three modeled policies on overall Medicare spending and on geographic variation.

Table 5.1. Summary of Three Modeled Policies on Spending and Geographic Variation

Policy	P4P	Bundled Payment	ACO
Impact on Total Medicare Spending per beneficiary	0%	1-3%	0.04-0.8%
Impact on Geographic Variation	None	Modest	Miniscule

The impact of the three polices on total Medicare spending was largely driven by the assumptions we made in our modeling. P4P was modeled deliberately as a revenue-neutral program. We assumed a set savings rate per Medicare beneficiary in an ACO. Therefore, the magnitude of savings depended on our estimates of the number of beneficiaries participating in the ACO program. In our models of bundled payment we saw higher savings across most scenarios. These were dependent on where we set the bundled prices in our models and how many conditions were included in the bundled payment program. While the savings in the ACO and bundled payment policies are small on a percentage basis, we note that a 1% reduction in Medicare program expenditures amounts to roughly $6 Billion in 2012.

In contrast, the effects of the policies on geographic variation in Medicare spending were not driven by our assumptions.

- P4P payment redistributions had no impact on geographic variation largely because there was no relationship between performance on the P4P quality measures and Medicare spending. The lack of a relationship between cost and quality could reflect heterogeneous providers (with respect to quality) within HRRs.
- The bundled payment policy, in contrast, is more likely to reduce geographic variation in Medicare spending. By paying at a national rate, providers with higher spending received a larger reduction in their payments. Therefore, as long as high-spending providers were in areas with high overall Medicare spending, geographic variation would be reduced. That was, in fact, the case.
- ACOs, with their assumed savings, would reduce geographic variation in spending if they were disproportionately located in high-cost regions of the US, or if savings were differentially larger in high-cost areas. Even though we assumed the latter in some scenarios, there is little relationship between where ACOs have formed and baseline spending and therefore there is only minimal effect on geographic variation in spending.

Our models have several key limitations, many of which are discussed above. First, the modeling exercises were designed to represent realistic versions of policies that could be implemented in the near future. We therefore modeled programs that closely resembled current

Medicare pilots or programs. However, we acknowledge different implementations of these policies could have a different impact on spending. Second, our results are limited by imperfect data. We made simplifying assumptions throughout to allow for modeling of program impacts against a 2008 baseline, but it is possible that some of these changed the nature of the results, or that the Medicare population in the future could be different. Our scenarios and sensitivity analyses attempted to diminish this concern. Finally, we made several assumptions about provider behavior. For example, in the case of bundled payments, we assumed that providers would react to the payment change by either reducing or accepting reduced margins – but that they do not change the number of bundles provided or costs of services outside of the bundle. In general, we only made assumptions for behavioral change that we felt had some plausible basis in the literature. However, we acknowledge that if we could estimate potential behavioral changes, the impact of these policies on geographic variation in spending could be different.

We investigated the impact of three policies that are on the front-burner for increasing the value of Medicare on an outcome that they are not primarily designed to influence - geographic variation in spending. The fact that we found no impact on geographic variation in spending in two of the three does not mean the policies are ineffective. In contrast to policies that directly adjust payments to an entire geographic locale, these policies might reduce geographic variation in spending as a side-effect. Whether they impact geographic variation in spending depends on a number of key factors we analyzed, such as correlation between cost and quality measures (P4P) or the geographic distribution of ACOs. The policies *could* be designed to have a larger impact on geographic variation than we found. For example, a set of P4P measures used could be deliberately refined to include only those where low-cost areas tended to have high scores. Yet under the reasonable sets of choices we made in modeling the policies, we found that only bundled payment significantly reduces geographic variation in spending.

While our results do not speak directly to whether the modeled policies would improve the value of Medicare, we feel that they do highlight some potential ways to improve value. For example, the fact that areas with high costs for the bundled services we identified tend to be high cost overall suggests similar drivers of high costs both within and outside the bundles. Generally, these results should prove useful to policymakers seeking solutions to both the problems of unwarranted variation in spending and value in Medicare.

References

American Hospital Directory, “Early Look at Coding Trends Under MS-DRGs,” Healthcare Financial Management, Vol. 63, No. 3, March 2009, pp. 116-117.

Centers for Medicare and Medicaid Services, "Roadmap for Implementing Value Driven Healthcare in the Traditional Medicare Fee-for-Service Program," undated. As of October 24, 2012: http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/VBPRoadmap_OEA_1-16_508.pdf

Centers for Medicare and Medicaid Services, "Final Waivers in Connection with the Shared Savings Program," *Federal Register,* Vol. 76, No. 212, November 2, 2011, 2011. As of October 24, 2012: http://www.gpo.gov/fdsys/pkg/FR-2011-11-02/pdf/2011-27460.pdf

Congressional Budget Office, "Budget Options Volume I: Health Care," 2008a. As of October 24, 2012: http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/99xx/doc9925/12-18-healthoptions.pdf

Congressional Budget Office, "Geographic Variation in Healthcare Spending," 2008b. As of January 18, 2013: http://www.cbo.gov/ftpdocs/89xx/doc8972/02-15-GeogHealth.pdf

Gaynor, M., J. B. Rebitzer, and L. J. Taylor, "Physician incentives in health maintenance organizations," *Journal of Political Economy,* Vol. 112, No. 4, Aug, 2004, pp. 915-931.

Gottlieb, D., Weiping Zhou, Yunji Song, K.G. Andrews, J. Skinner, and J. Sutherland, *A Standardized Method for Adjusting Medicare Expenditures for Regional Differences in Prices*, Lebanon, N. Hamp.: Technical Report- The Dartmouth Insitute for Health Policy and Clinical Practice, 2010.

Hussey PS, Mulcahy AW, Schnyer C, et al. *Closing the Quality Gap: Revisiting the State of the Science,* Vol. 1: *Bundled Payment: Effects on Health Care Spending and Quality*, Rockville, MD: Agency for Healthcare Research and Quality, Evidence Reports/Technology Assessments, No. 208.1, 2012.

Lindenauer, P. K., D. Remus, S. Roman, M. B. Rothberg, E. M. Benjamin, A. Ma, and D. W. Bratzler, "Public reporting and pay for performance in hospital quality improvement," *N Engl J Med,* Vol. 356, No. 5, Feb 1, 2007, pp. 486-496.

McWilliams, J. Michael, and Zirui Song, "Implications for ACOs of variations in spending growth," *The New England Journal of Medicine,* Vol. 366, No. 19, 2012, pp. e29-e29.

Mehrotra, A., C. L. Damberg, M. E. Sorbero, and S. S. Teleki, "Pay for performance in the hospital setting: what is the state of the evidence?," *Am J Med Qual,* Vol. 24, No. 1, Jan-Feb, 2009, pp. 19-28.

Papke, Leslie E., and Jeffrey M. Wooldridge, "Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates," *Journal of Applied Econometrics,* Vol. 11, No. 6, 1996, pp. 619–632.

Rosenthal, M. B., and R. G. Frank, "What is the empirical basis for paying for quality in health care?," *Med Care Res Rev,* Vol. 63, No. 2, Apr, 2006, pp. 135-157.

Song, Zirui, Dana Gelb Safran, Bruce E. Landon, Mary Beth Landrum, Yulei He, Robert E. Mechanic, Matthew P. Day, and Michael E. Chernew, "The 'Alternative Quality Contract,' based on a global budget, lowered medical spending and improved quality," *Health Affairs (Project Hope),* Vol. 31, No. 8, 2012, pp. 1885-1894.

Vanlare, Jordan, "Personal Communication," Centers for Medicare and Medicaid, July 12, 2012, 2012.

Appendices

Appendix 1 – Pay-for-Performance (P4P) Methods

This Appendix describes in more detail the methods and data used to model the impact of P4P on Medicare spending and geographic variation in Medicare spending. We modeled the effect of three national Medicare P4P initiatives on geographic variation in Medicare spending. Our goal was to assess whether these programs will attenuate or increase differences in spending between high and low cost regions.

The modeled P4P policies are based on existing or pilot Medicare P4P (also known as value-based purchasing) initiatives that target (1) hospitals, (2) nursing homes, and (3) home-health agencies. In our models, we have changed several key aspects of some of the existing or pilot Medicare P4P programs and these changes are noted below. These changes were primarily made for consistency and to ensure a common structure (e.g., how much money is tied to incentives) across the three P4P programs. Some changes were made because we lacked certain data elements. Lastly, the Committee's charge to us was to model a more robust P4P program and for that reason, we increased the fraction of money devoted to incentives in some scenarios.

The key steps in the analysis are outlined in Table A.1.1 below. For each hospital, nursing home, and home health agency, we first calculated total Medicare payments to that provider in 2008. Then using publicly available quality data on the quality measures used in the P4P program, we estimated incentive bonuses or penalties that would be paid to each provider under the P4P program. Based on the enrollee's residence, we assigned each service by a hospital, nursing home, or home-health agency to an HRR. We then estimated HRR-level changes in payment due to P4P bonuses and penalties for overall spending, and separately for nursing home,, inpatient, and home health spending by HRR.

Table A.1.1. Steps in P4P Analysis Plan

1.	Calculate baseline total payments to each provider in the nation
2.	Subtract a percentage of all payments (2% or 15%) to provider to create a pool of incentive payments
3.	For each provider calculate quality score (incorporating behavioral change or assuming no behavioral change)
4.	Based on quality score determine a percentage change in payment for each provider.
5.	Calculate quality-adjusted total payments to each provider in nation
6.	Aggregate payments to providers by HRR
7.	Characterize changes in geographic variation and total Medicare spending

To estimate the range of the impact, we modeled five scenarios in which we varied several key program design parameters in the P4P programs. The modeled P4P programs only included incentives for quality of care. None of these programs, as currently designed, tie incentives to

costs or resource use/efficiency of care. By design, the P4P programs we modeled did not impact overall Medicare spending. This is consistent with budget-neutral CMS value-based purchasing programs that have been implemented in demonstration or pilot form.

Data

Data on inpatient, nursing home, and home health spending

The data for the analysis consisted of the following: 100% MedPAR FFS claims for acute hospitals and skilled nursing facilities100% home health standard analytic files (SAF). We also used the 100% Medicare denominator file, provider of services files.

Data on total Medicare spending by HRR

We obtained data on total Medicare spending by HRR using published data by the IOM entitled "HRR Level Demographic, Cost, Utilization, and Quality Data."[12] The underlying data comes from the CMS Chronic Conditions Warehouse[13] which contains all Medicare claims for beneficiaries who are enrolled in the fee-for-service (FFS) program as well as enrollment and eligibility data. The analyses were conducted by the Center for Medicare and Medicaid Services.

Data on quality of hospitals, nursing homes, and home health agencies

We aggregated quality data from a variety of sources. These quality data were then used as inputs in computing composite quality scores for each provider that determined their payment under the P4P programs. In general, our goal was to obtain data from the most recent year available and historical data from two years prior to the most recent year available. This was consistent with the lag period for the inpatient hospital VBP program developed by CMS.

Much of our data came from publicly available files on quality measures from the Home Health Compare[14] and Nursing Home Compare[15] websites. We used the most recent data available. The quality data for home health agencies covered the reporting periods of 2011 and 2009. For nursing homes, we obtained staffing data from 2012, resident outcome data from 2010 and 2008, and deficiency data from 2011 and 2009. These quality data were supplemented by data from Medicare's Online Survey, Certification and Reporting (OSCAR) and Minimum Data Set (MDS) files.

For hospitals, we obtained the proposed Value-Based Purchasing Program (VBP) Adjustment Factors for FY 2103 directly from CMS (Vanlare, 2012). These adjustment factors

---

[12] iom.edu/Activities/HealthServices/GeographicVariation/Data-Resources.aspx

[13] ccwdata.org/index.php

[14] www.medicare.gov/homehealthcompare/

[15] www.medicare.gov/nhcompare/

were reported for each participating hospital and allowed us to directly compute incentive payments, without needing to compute each hospital's quality scores.

In some analyses, we used aggregate quality scores at the HRR-level. We obtained aggregate scores for hospitals from the IOM HRR-level spreadsheet and generated analogous payment-weighted average quality scores from our claims data for home health agencies and nursing homes.

Data on provider characteristics

In supplemental analyses (not included in the body of the report), we examined the association between gains/losses under P4P and key provider characteristics such as ownership status, staffing, size, safety net status, and region. For hospitals, we obtained data on provider characteristics from the 2008 American Hospital Association (AHA) database. For nursing homes and home health agencies, we obtained data from the CMS Provider of Services File and Medicare Cost Reports.

Enrollee population studied (inclusion and exclusion criteria)

The study population for all analyses included Medicare fee-for-service beneficiaries aged 65 and older who were enrolled in Parts A and B for the entire year or who were enrolled in Parts A and B until their death.

Beneficiaries excluded from our analysis were those who: enrolled at any time in a Medicare Advantage plan, became eligible after January 1st, 2008, had only Part A or Part B benefits, were disabled or had end stage renal disease, and lived outside the United States (e.g., Puerto Rico). In total, our study population includes approximately 54% of the total Medicare population for 2008.

We excluded Maryland providers from the P4P programs. Maryland hospitals are paid using an all-payer system and were exempted from the FY 2013 Medicare Hospital VBP program. To maintain consistency across all P4P programs, we excluded Maryland providers from the analyses.

Assigning Medicare beneficiaries to a Hospital Referral Region (HRR)

Based on each beneficiary's zip code in the Denominator file, we assigned the beneficiary to an HRR. We used a cross-walk that has been made publicly available by researchers at Dartmouth.[16] Beneficiaries with invalid zip codes (e.g. 00000) were excluded from the analysis.

Addressing area-level and other adjustments to Medicare payments

Medicare payments were price-standardized to omit DSH/IME payments and account for area-level wage and price adjustments in Scenario 5. In this scenario, differences in spending

[16] http://www.dartmouthatlas.org/tools/downloads.aspx

across regions can be attributed to differences in utilization without being confounded by other factors such as having a disproportionate share of teaching hospitals. To standardize payments for each HRR, we applied the ratio of standardized-to-actual spending for each provider type as reported in the IOM HRR-level spreadsheet for 2008. In all other scenarios, total Medicare payments are used.

How P4P programs were modeled

Our analyses were modeled after actual Medicare VBP programs and demonstrations for hospitals, home health and nursing homes. Whenever possible, we used the same methods outlined by CMS to compute quality scores from existing quality measures. Quality scores were then mapped to incentive payments. While we used the actual quality measures in each of the three programs, our models differed from the programs in two ways: (1) what fraction of spending is allocated to incentive payments and (2) how those incentive dollars are allocated.

The amount of money allocated to incentives was measured as a fraction of overall Medicare reimbursements for each provider type (e.g. total payments to home health agencies). We modeled two levels of incentives, 2% and 15%. The conservative program (2%) is consistent with current P4P programs. The robust program (15%) reflects the Committee's desire for a more aggressive P4P program. To remain budget-neutral, we generated a pool of incentive payments by decreasing all providers' payments by an equal amount. For example, all hospital inpatient payments were reduced ("withheld") by 2% or 15%. The money in this incentive pool was allocated to providers qualifying for incentive payments.

We modeled two mechanisms to allocate the incentive dollars: tournament and linear-exchange curve. The tournament mechanism is based on the CMS nursing home P4P program while the linear-exchange curve is based on the CMS hospital P4P program. We chose to model both mechanisms of allocating dollars across *all three* P4P programs because the mechanism of distributing incentives could have an impact on geographic variation in spending and therefore it was important to be consistent across the three programs.

In the sections below, we outline: 1) how quality scores are computed in each P4P program and 2) how incentive dollars are allocated under each of the five scenarios.

How quality scores were computed

Hospitals

Our hospital model was based on the Medicare Hospital VBP program that was implemented nationally in FY 2013.[17] In prior work, CMS computed quality scores for each hospital and used those scores to generate their likely incentive payments. These incentive payments were

[17] http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html

distributed as proposed "Adjustment Factors" for FY 2013. Instead of repeating this process, we used the proposed Adjustment Factors for FY 2103 from CMS to compute incentive payments for participating hospitals. The adjustment factors (which range from 0.992 to 1.009) were multiplied by base Medicare payments to determine total payments under P4P. We normalized these adjustment factors so that, when applied to base payments, total national payments equaled the 2008 value rather than Medicare payments in 2013. This maintains budget neutrality. We also modified the adjustment factors to generate a 2% or 15% incentive pool rather than the 1% pool that was implemented by CMS for FY2013. For future years, the Hospital VBP program will be using 2%.

Home Health

Our home health model was based on the CMS Home Health Pay-For-Performance Demonstration, which was implemented in seven states between January 2008-December 2009.[18] This demonstration used a tournament-style approach to allocate incentive payments to the top 20% of providers in achievement and improvement for each of seven quality measures. The total incentive pool was allocated to each of the measures (separately for achievement and improvement) based on the percentages shown in Appendix Table 2 below. Providers could earn incentive payments for achievement on some measures and improvement on other measures.

Table A.1.2: Quality Measures Used in Home Health Value-Based Purchasing Program

Quality Measure	Achievement Pool	Improvement Pool	Total
Incidence of Acute Care Hospitalization	22.5 %	7.5 %	30 %
Incidence of Any Emergent Care	15 %	5 %	20 %
Improvement in Ambulation / Locomotion	7.5 %	2.5 %	10 %
Improvement in Bathing	7.5 %	2.5 %	10 %
Improvement in Management of Oral Medications	7.5 %	2.5 %	10 %
Improvement in Status of Surgical Wounds	7.5 %	2.5 %	10 %
Improvement in Transferring	7.5 %	2.5 %	10 %
Total	75 %	25 %	100 %

To maintain consistency with the hospital and nursing home programs, which map a single composite quality score to a single incentive payment, we modified the design of the home health program to mimic the hospital VBP program. We preserved the use of the seven quality measures and the relative weights of these measures from the home health demonstration, but followed the hospital VBP approach to allocating payments. Specifically, the approach for computing quality scores, based on the hospital program, proceeded in three steps.

[18] http://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/Medicare-Demonstrations-Items/CMS1189406.html

First, we computed "achievement" and "improvement" scores for each quality measure. We used the seven[19] process and outcome quality scores that were identified in the home health demonstration.

- $q_{im}^{Achieve}$ is the achievement score for provider i for measure m. This was calculated according to the equation: $q_{im}^{Achieve} = 9 \cdot \left(\frac{s_{im} - k_m^{Achieve}}{k_m^{Bench} - k_m^{Achieve}}\right) + 0.5$ and rounded to the nearest whole number. s_{im} is the reported value of the quality measure in the current year, k_m^{Bench} is the mean of the top decile of the distribution for measure m in the baseline year, and $k_m^{Achieve}$ is the median of the distribution for measure m in the baseline year.
- $q_{im}^{Improve}$ is the improvement score for provider i for measure m. This was calculated according to the equation: $q_{im}^{Improve} = 10 \cdot \left(\frac{s_{im} - s_{im}^{Baseline}}{k_m^{Bench} - s_{im}^{Baseline}}\right) - 0.5$ and rounded to the nearest whole number. s_{im} is the reported value of the quality measure in the current year, $s_{im}^{Baseline}$ is the reported value of the quality measure in the baseline year, and k_m^{Bench} is the mean of the top decile of the distribution for measure m in the baseline year.

Second, we computed composite quality scores for subsets of similar quality measures. In the hospital VBP program, there were separate scores for "Patient Experience of Care" measures and "Clinical Process of Care" measures. In adapting this to the home health case, we computed separate scores for the two "Clinical Outcome" measures (i.e. Incidence of Acute Care Hospitalization and Incidence of Any Emergent Care) and the remaining five "Clinical Process of Care" measures. To compute the "Clinical Outcome" score, we first took the maximum of the achievement or improvement score for each of the five measures (each on a 0-10 point scale). We then divided the sum of these maximum scores by the total number of points possible to generate a score from 0-10.[20] The "Clinical Process of Care" score was computed analogously.

Third, we computed the overall composite quality score: Total Quality Score = 0.5* Clinical Process of Care + 0.5* Clinical Outcome. We used weights equal to 0.5 to mimic the weights used for the process (0.3 + 0.2) and outcome (0.1 * 5) measures in the home health demonstration as shown in Appendix Table 1.2.

The key limitations in applying the hospital algorithm to compute home health scores were (1) achievement and improvement are weighted 3:1 in the home health demonstration, but get equal weight in the hospital VBP program (i.e. the maximum of achievement and improvement is selected); (2) the measures "Incidence of Acute Care Hospitalization" and "Incidence of Any

[19] We do not include the measure "Incidence of Any Emergent Care" because it is missing from the Medicare Compare data in the most recent year.

[20] If the provider does not report a score for a given measure, the total number of points possible will be reduced to reflect that omission. If more than half of the measures are missing for a provider, they will not receive an incentive payment.

Emergent Care" measures are weighted 3:2 in the home health demo, but get equal weight in the hospital VBP program; 3) providers can receive incentive payments for some quality measures and not others in the home health demonstration, but receive incentive payments based on only a composite quality score in the hospital demonstration.

Nursing Homes

Our nursing home P4P program is based on the CMS Nursing Home Quality-Based Purchasing Demonstration[21] which was implemented in three states (Arizona, New York, and Wisconsin) and 171 nursing homes between July 2008 – June 2012. A detailed description of the Demonstration can be found elsewhere.[22]

The demonstration was designed to ensure budget neutrality, but used a different design from other Medicare P4P programs. The incentive pool is created by the money Medicare saves via avoidable hospitalizations. The assumption is that nursing homes that improve quality will drive decreases in avoidable hospitalizations. An incentive pool was created for each year for each State in the demonstration. The incentive pool was the savings in excess of 2.3% of total Medicare expenditures.

In the Demonstration, nursing homes' performance was assessed on 4 domains: (1) staffing, (2) appropriate hospitalizations, (3) outcome measures, and (4) survey deficiencies (Table A.1.3).

Table A.1.3: Quality Domains Used in Nursing Home Quality-Based Purchasing Demonstration

Quality Domain	Total
Staffing	30 points
Hospitalization	30 points
Resident Outcomes	20 points
Survey Deficiencies	20 points
Total	100 points

Each domain had specific quality measures and unique point allocations (Table A.1.4). For example, for staffing, there were three measures, registered nurse hours per resident day (10 points), licensed nursing staff staffing (5 points), total nursing hours per day (divided into licensed nursing and certified nurse assistant, 5 points each), and nursing staff turnover (10 points). The nursing homes were ranked to generate a point total using the following formula: # of points = percentile rank*.10. Nursing homes obtained an extra 0.1 points if they were in the 99th percentile. In contrast, for the resident outcome measures, a nursing home received zero

[21] The program is also called the Nursing Home Value-Based Purchasing Demonstration

[22] www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/NHP4P_FinalReport.pdf, key design refinements, http://www.cms.gov/Medicare/Demonstration-Projects/DemoProjectsEvalRpts/downloads/NHP4P_Refinements_Report.pdf

points if they were at or below the 5th percentile and the maximum points at or above the 95th percentile. Within the 5th and 95th percentile they used the following formula, points earned = maximum number of points for measure * (nursing home value for measure-5th percentile value of measure)/(95th percentile value of measure -5th percentile value of measure).

Separate measures were included for residents with short versus long stays. Some nursing homes only care for short-stay or long-stay residents. There were also minimum requirements for how many measures on which a nursing home must be rated upon for the nursing home to be eligible for P4P.

An overall achievement and improvement score were generated from the total points obtained from each measure. Nursing homes with an overall performance that was in the top 20 percent of *achievement* received incentive payments. Nursing homes in the top 20 percent of *improvement* would qualify for a performance payment in recognition of their improved performance, as long as their performance level was at least as high as the 40th percentile. Among nursing homes in the top 20 percent, nursing homes in the top decile received more payments than those in the second decile.

To maintain consistency with the hospital and home health programs, we made several key changes to the nursing home program. First, several of the quality measures were not available to us (Table A.1.4). For example, the staff turnover measure was based on direct payroll data from the nursing homes which is not publicly available for all nursing homes. We re-weighted each of three domains for which we had quality measures, 43% staffing, 29% resident outcomes, and 29% survey deficiencies.

Second, we mimicked the inpatient P4P program's design when generating achievement and improvement scores (details are available elsewhere and outlined in the hospital section above23). In short, for each quality measure, we generated an achievement and benchmark thresholds. The achievement threshold was the median performance in the base year. The benchmark was the mean performance among the top decile of performers in the base year. The achievement score for a nursing home for given measure was 0 to 10 points. If the nursing home scored below the achievement threshold they received 0 points. If the nursing home scored above the benchmark threshold they received 10 points. Nursing homes that scored between the achievement and benchmark thresholds received 1 to 9 points. Each nursing home received an improvement score = (9*(current year's core – base year's score)/(benchmark threshold – base year's score))-0.5. All scores were rounded to nearest whole number. For each measure, the nursing home was assigned the greater of the achievement and improvement score. If there were no base-year data, the nursing home could only receive an achievement score. For staffing

[23] [23] http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/index.html

measures, there were no base year data. We calculated both the achievement and benchmark thresholds using the most recent data.

Table A.1.4: Quality Measures Used in Nursing Home Quality-Based Purchasing Demonstration by Domain

Quality Measures	Included in RAND Modeling
Domain: Nurse Staffing	
Nursing hours per resident day	X
Total nursing hours per resident day	X
Total nursing staff turnover percentage	
Domain: Hospitalizations	
Percent of residents w/ a potentially avoidable hospitalization	
Percent of residents w/ a potentially avoidable hospitalization	
Domain: Residency Outcomes	
Percent of residents whose need for help with daily activities has increased	X
Percent of residents whose ability to move about in and around their room worsened	X
Percent of high risk residents who have pressure sores	X
Percent of residents who had a catheter inserted and left in their bladder	X
of residents who were physically restrained	X
Percent of residents w/ improving level of ADL functioning	x[24]
Percent of residents who improve status on mid-loss ADL functioning (transfer, locomotion) or remain completely independent in mid-loss ADLs	x
Percent of residents with failure to improve bladder incontinence	X
Domain: Survey Deficiencies	
Survey performance score	X[25]

Finally, we calculated a total score for each domain and then based on the domain weights noted above, calculated a total quality score for each provider.

How the behavioral response was modeled

Modeling the behavioral effects of P4P on quality presented a challenge. Despite widespread interest in P4P, research on CMS demonstrations and private-sector P4P programs has found that P4P programs have had, at most, a modest impact on overall quality (Mehrotra et al., 2009; Rosenthal and Frank, 2006; Lindenauer et al., 2007). However, it is possible that larger payment incentives and permanent programs, rather than demonstrations, will lead to larger effects. To

[24] The ADL scores were generated using the MDS database by RAND

[25] These were generated using the OSCAR database by RAND.

incorporate behavioral effects of the policy (i.e. changes in quality), under Scenario 4, we assumed there will be an overall improvement in quality by on average 10% across all providers. We also assumed that the improvement would be largest among those who started at a low baseline and smallest among those who had the highest quality at baseline, thus narrowing the quality gap. To do this we reduced the gap between a provider's quality score at baseline and the maximum possible score by 5%, i.e.

$$Quality\ Score' = Quality\ Score + (Max\ Possible\ Score - Quality\ Score) * 0.05.$$

This approach generated an average increase in quality scores of nearly 10% for the three provider types. Many of the quality metrics are scaled from 0 to 100% and the top providers are already at 100%. This "ceiling effect" constrains any quality improvement among these top providers. Together these modeled changes compressed the distribution of quality across providers and therefore led to less re-distribution of money across HRRs.

How incentive dollars were allocated to providers

In our next step, we computed incentive payments as a function of each provider's composite quality score. Providers did not receive incentive payments if they were ineligible for the P4P program or did not report quality measures. For example, the hospital P4P program excludes hospitals with fewer than30 hospitalizations in a year for the conditions studied and those hospitals designated by Medicare as critical access hospitals (CAHs).

Incentive payments were computed as a percentage of baseline Medicare payments for each provider. The hospital VBP program uses the base operating diagnostic-related group (DRG) payment amount as the baseline, and we follow this approach. Using the 100% Medicare claims files, we calculated providers' total payments from Medicare in 2008 subtracting DSH, IME, and outlier payments. We then subtracted 2% or 15%, (conservative or robust programs, respectively) of all payments to all providers in the nation. This created a pool of incentive dollars. The incentive pool dollars were then allocated across providers by either the tournament or linear exchange curve mechanisms.

In the tournament mechanism, we identified the top 20 percent of providers in terms of their composite quality scores. The top 20 percent received payments from the incentive pool, while the remaining providers each experienced a loss of 2% or 15%. Each provider's share of the incentive dollars was proportional to the share of total Medicare reimbursement received by the provider in that year.

In the linear exchange curve method, we allocated incentive dollars on a continuum. Providers with the worst performance receive no incentives and as performance increases providers get larger and larger incentives. Incentives were allocated to providers as a percentage increase proportional to their composite quality score. The slope of the linear exchange function was set to achieve budget neutrality based on total Medicare reimbursement. Again, total Medicare reimbursement was based on reimbursement calculated from the base rate, excluding DSH, IME, and outliers.

Finally, we summed the incentive payments for each provider with their base payment and their adjustment payments (IME, DSH, and outliers) to compute total Medicare payments under each P4P scenario. The baseline is simply the total payments for each provider from the 100% claims data. For providers who were ineligible for P4P (e.g., missing quality measures), their total Medicare payments stayed constant under P4P.

Aggregating incentive payments to the HRR level

After determining each provider's total payments, both under the baseline and each of the P4P scenarios, we allocated those payments to HRRs based on the zip code of residence of the beneficiary. We then summed total payments within each HRR for hospitals, nursing homes, and home health providers. To compute total Medicare payments for HRRs, we used the 2008 IOM HRR-level spreadsheet. We applied the incentives as a percentage change in base payments to the IOM data for hospitals, nursing homes, and home health and then added all other Medicare payments (e.g. physician services, prescription drugs, durable medical equipment, etc.) to these totals by HRR. Given small differences in study samples between the claims data and the IOM data, we applied a scalar adjustment factor to normalize our claims data to sum to the total payments in the IOM data for hospitals, nursing homes, and home health.

The HRR total Medicare payments and component payments were used to calculate measures of regional variation in spending (e.g. coefficient of variation, ratio of 75^{th} to 25^{th} percentile costs) at baseline and under the modeled policies. These values were also used to describe the characteristics of HRRs and providers experiencing large and small changes in Medicare payments.

Summary of Scenarios in models

Table 2.1in the body of the report summarizes the five scenarios that were modeled. Scenarios 1 and 2 both assumed that 2% of total payments are allocated as incentive payments, but they varied the method by which incentives are allocated. In Scenarios 3 and 4 we assumed that 15% of payments are allocated as incentives. We did not model a tournament style allocation of incentives under the robust program because such a large shift of dollars to a small set of providers does not seem realistic. We only modeled a behavioral response under the more robust program (15% of payments tied to incentives) where such a response is more likely. Scenario 5 is similar to Scenario 3, but we excluded area-level adjustments, medical education, and DSH from provider payments.

Appendix 2 – Bundled Payment Methods

Appendix 2 describes in more detail the methods and data used to model the impact of bundled payment on Medicare spending and geographic variation in Medicare spending.

Constructing Historical Bundled Episodes

The bundled payment baseline and modeled scenarios are all based on historical spending patterns. Our first step was to calculate bundle costs using 2007 to 2008 Medicare claims data.

Data Sources

The base data for the analysis consisted of the following: 100% MedPAR FFS claims for acute hospitals, skilled nursing facilities, inpatient rehabilitation facilities, and long term care hospitals; 100% home health standard analytic files (SAF), 5% Carrier SAF claims, 5% Outpatient SAF claims, and 5% Inpatient SAF claims. We also used the 100% Medicare denominator file, provider of services files, and Acute and Inpatient Rehabilitation Facility Impact files.

Episode definition

Bundled episodes were identified by acute inpatient discharges and included the index hospitalization and other services provided during a 30-day post-discharge period. We examined inpatient discharges occurring from December 2, 2007 through December 1, 2008 (with episode end dates spanning from January 1 through December 31, 2008). We included acute hospitalizations for all of the 27 MS-DRGs associated with the ten study conditions, which were chosen on the basis of high historical Medicare utilization of acute and post-acute care (Appendix Table A.2.1). Specifically, these ten conditions were among the top twenty in terms of volume of hospital stays in FY 2007 and 2008 (American Hospital Directory, 2009). The ten conditions include the top five conditions in terms of volume and five other conditions that were either high volume users of post-acute care (e.g. stroke, hip fracture) or served as important examples of medical conditions (e.g. acute myocardial infarction). The MS-DRG system was implemented in FY 2008 (starting in October 2007). There were no substantive changes in the definitions for FY 2009 (starting in October 2008) that affected the conditions in our sample.

Table A.2.1. Clinical Conditions And Related MS-DRGs Included In The Hypothetical Medicare Bundled Payment Program

Condition	MS-DRGs
Acute Myocardial Infarction	280, 281, 282
Congestive Heart Failure	291, 292, 293
Chronic Obstructive Pulmonary Disease	190, 191, 192
Gastrointestinal Bleed	377, 378, 379
Hip Fracture	480, 481, 482
Kidney/Urinary Tract Infection	689, 690
Lower Extremity Joint Replacement	469, 470
Pneumonia	193, 194, 195
Septicemia	871, 872
Stroke	064, 065, 066

Exclusions

We made the following exclusions:

- Episodes triggered by index stays (in the first month) that were themselves 30-day readmissions for prior index stays prior to our sample period.
- Index stays in IPPS exempt acute hospitals (i.e. critical access hospitals), although we counted readmissions in such hospitals. We use the last four digits of the Medicare provider number to identify provider types: we categorize IPPS facilities as code ranges 0001-1199 and 1300-1399.
- Index stays in Maryland hospitals or beneficiaries from Maryland.
- Index stays outside of 50 states +DC
- Index stays that were transfers
- Patients who died in the hospital
- Beneficiaries without continuous Part A and B during 2007 or 2008 (based whether the index hospital discharge occurred in 2007 or 2008). If patients died, we required continuous Part A and Part B up through the month when they died.
- Aged into Medicare during 2007 or 2008 (depending on year of index hospital discharge)
- Any months of MA enrollment in 2007 or 2008 (depending on year of index hospital discharge)
- Medicare eligible because disabled or ESRD in 2007 or 2008 (depending on year of index hospital discharge)

Identifying bundled services: 100% data

For the services covered in the 100% data (listed below), we used study-assigned beneficiary identifiers to identify utilization in the 30-day period following the discharge from the index hospitalization. Again, we identified provider types using the last four digits of Medicare provider numbers. Specifically, we identified the following services in the 100% data:

- Acute inpatient facility (readmissions): acute hospital admissions for a beneficiary with dates of admission <30 days from the discharge date of the index discharge (we include psych hospitals, cancer hospitals, and critical access hospitals in this definition, although they form a small proportion of total readmissions).
- Long term care hospital (provider number 2000-2299) admissions <30 days from discharge date
- Inpatient rehabilitation facility (provider number 3025-2099) and admissions<30 days from discharge date
- Skilled nursing facility (provider numbers 5000-6499) and admissions<30 days from discharge date
- Home health: admissions<30 days from discharge date of index discharge
- If post-acute episodes straddle the 30-day post-discharge period, then we only include the prorated amount occurring within the 30-day post-discharge period in the historical episode spending.

Identifying Bundled Services: 5% Data

We had only had access to 5% Carrier and Outpatient claims. As a result, we calculated geographic cell-level carrier and outpatient measures for each MS-DRG, which we then merged into the bundle-level file. The specific services we obtain from the 5% claims include outpatient, ambulatory surgery center, evaluation and management, procedures, imaging, lab use, other tests, Part B drugs, and other Part B use. We more precisely describe the identification of specific services in the adjustment methodology section below.

The construction of cell-level averages from the 5% claims data occurred as follows. First, we identified eligible acute discharges in the 5% Inpatient SAF claims data. Then, using study-assigned beneficiary-IDs we linked the discharges to service use in the Carrier and Outpatient 5% claims. Then, we take geographic means from the 5% data. Due to small sample sizes, we use a hierarchical approach and use the smallest geographic area that has at least 20 episodes for the MS-DRG: HRR, state, census division. In HRRs where a smaller geographic area has at least 10 episodes and the next larger geographic area has at least 20 episodes, we used the average of the two geographic area means.

Modeling Alternative Bundled Payment Rates

We modeled seven alternative bundled payment rate scenarios:

S1. National base rate with payment adjustments
S2. National base rate with payment adjustments and a volume threshold≥10
S3. Lower of national and historical provider payment, volume threshold ≥10
S4. Blend of national and historical provider payment, volume threshold ≥10
S5. National base rate with payment adjustments, volume threshold ≥10, budget neutral
S6. National base rate with payment adjustments, volume threshold ≥10, post-acute only
S7. Price standardized national rate, no volume threshold

These seven scenarios fall into three general categories: (1) a national rate, (2) the lower of national and historical provider payment, and (3) a blend of the national rate and historical provider payment. In this section, we discuss the construction of each general case, and specific details of individual scenarios.

National Rate (S1, S2, S5, S6, S7)

Scenario 1:

Step 1: Calculate total episode payments for each historical bundled episode (acute inpatient payments plus skilled nursing, inpatient rehabilitation, home health, long-term care hospital, outpatient, and carrier payments)
Step 2: Find the median episode payment, reduce by 5% to create unadjusted national base rate
Step 3: Find the average proportion of total episode payments in each care setting (e.g. fraction acute inpatient, SNF, IRF, etc.)
Step 4: Multiply fraction for each care setting by unadjusted national base rate to produce base payment components by care setting.
Step 5: Adjust each base payment component by HRR-level adjustment factor (described in next section) to create HRR-specific adjusted base payment components.
Step 6: Aggregate across care-settings to create adjusted national base rate

How Scenario 2 Differs from Scenario 1

Acute providers with fewer than 10 cases are exempt from policy; we assumed that each episode receives historical payment

How Scenario 5 Differs from Scenario 1

Acute providers with fewer than 10 cases are exempt from policy; we assumed that for each episode receives historical payment. It is budget neutral, so we used the mean episode payment in Step 2. In addition, after Step 6 we applied a global adjustment factor to enforce that mean episode payment under policy is equal to pre-policy mean.

How Scenario 6 Differs from Scenario 1

Acute providers with fewer than 10 cases are exempt from policy; we assumed that each episode receives historical payment. Only the index hospitalization, readmissions, and post-acute care were included in the bundle.

How Scenario 7 Differs from Scenario 1

It is price standardized, so we used the unadjusted national base rate as the new rate. Baseline scenario is standardized by a weighted average of the inverse of the HRR-level adjustment factors for each component (described in next section).

Lower of adjusted national base rate and hospital-specific historical spending (S3)

Step 1: Find the average historical episode payment for each discharging acute provider.
Step 2: Find the lower of the average provider historical episode payment and the adjusted national base rate, and this will be the payment rate.
Step 3: If a discharging acute provider has a volume less than 10, then each episode for that provider will receive the historical episode payment.

Blend of adjusted national base rate and hospital-specific historical spending (S4)

Step 1: Payment rate is 50% adjusted national base rate and 50% provider-level historical episode spending.
Step 2: If a discharging acute provider has a volume less than 10, then each episode for that provider will receive the historical episode payment.

Strategy to Incorporate CMS Payment Adjustments

Adjustments Overview

In each rate-setting approach, we adjusted the national base rate to reflect geographic variation in wages, input prices, the presence of medical training programs (IME payments), the proportion of low-income patients (DSH payments), and other adjustment factors used by Medicare. CMS is likely to incorporate such adjustments in the implementation of a bundled payment policy, resulting in some degree of geographic variation in episode spending even with a national rate. This section outlines our approach for making such adjustments.

Approach

IOM has published standardized and unstandardized payments by type of service (e.g. acute, skilled nursing, inpatient rehabilitation, etc.) by HRR in 2008. For each HRR, we calculated "adjustment ratios" for each type of service that reflect all the adjustments made to Medicare payments. For example, if the national base payment rate for the skilled nursing component of a bundled payment was $200 and the adjustment ratio for skilled nursing in a HRR was 0.9, the bundled payment amount for the skilled nursing component of bundles for patients in that HRR would be $200*0.9 = $180. The steps for creating the adjusted national base rates for each HRR are outlined below.

Step 1: Find the unadjusted national base rate.

We calculated the unadjusted national base rate as a 5% reduction on the national median of historical Medicare payments across episodes. Specifically, we included acute inpatient payments plus skilled nursing, inpatient rehabilitation, home health, long-term care hospital, outpatient, and carrier payments occurring within 30 days of acute discharge.

Step 2. Distribute the national base rate into separate components for each care setting.

Next, we found the average fraction of total episode payments corresponding to each care setting (acute, skilled nursing, etc.) across episodes for the entire nation. We applied these fractions to the unadjusted national base rate to create separate payment components for each care setting.

Step 3. Create HRR-level actual-to-standardized spending ratios for each care setting

The CMS-constructed database, "Table 1: HRR Level Demographic, Cost, Utilization, and Quality Data" [26] includes HRR-level actual and standardized spending for each component of the national base rate (inpatient, SNF, IRF, LTCH, outpatient, and carrier payments), where the standardization adjusts for differences in local wages or other input prices, extra payments for providers serving low-income populations or training doctors, and other adjustments.[27] We used these measures to create actual-to-standardized spending ratios (the "adjustment ratios") for each care setting. These ratios reflect price and other adjustments for a given care setting in each HRR relative to the national standard.

Step 4. Adjust each payment component by the actual-to-standardized ratio to create HRR-specific adjusted base rates by care setting.

We multiplied the unadjusted base payments for each care setting by the actual-to-standardized ratios to create HRR-specific adjusted payments for each care setting. While the mapping from components of the national rate to adjustment ratios is straightforward for the facility payments (e.g. inpatient hospital, IRF, SNF etc.), it is more complicated for the Outpatient and Carrier claims. The table below describes the mapping of outpatient and carrier claims and the associated ratio constructed from the IOM database.

[26]http://www.iom.edu/Activities/HealthServices/GeographicVariation/~/media/Files/Activity%20Files/HealthServices/GeographicVariation/CMSfiles/Table1.xlsx

[27] As outlined in the IOM Technical supplement for standardizing Medicare spending here: http://www.iom.edu/Activities/HealthServices/GeographicVariation/~/media/Files/Activity%20Files/HealthServices/GeographicVariation/CMSfiles/TechSupplement.docx

Table A.2.2. Carrier and Outpatient Bundle Components and Associated Adjustment Ratio

Component	Spending measures	Adjustment ratio definition constructed from IOM database
Outpatient	All claims in Outpatient file	Hospital outpatient (OP) actual Medicare costs (facility payments for outpatient visits paid under the outpatient prospective payment system and outpatient visits at CAHs)
ASC	Carrier claims with LINE_PLACE_OF_SRVC_CD = '24' for "Ambulatory Surgical Center" and LINE_CMS_TYPE_SRVC_CD = 'F' for "ASC facility usage for surgical services" and PRVDR_SPCLTY = '49' for "Ambulatory Surgical Center" (This category should have hierarchical precedence over other carrier service types, i.e. services that meet ASC criteria and also have BETOS codes listed below should be assigned to ASC only)	Ambulatory Surgery Center (ASC) actual Medicare costs
E&M	Carrier claims with BETOS = M**	Evaluation and Management (E&M) actual Medicare costs (physician claims for office visits, hospital visits, ER visits, nursing home/home visits, specialist visits, and consultations)
Procedures	Carrier claims with BETOS = P** except Anesthesia	Procedure (PROC) actual Medicare costs (physician claims for major other procedures, major cardio procedures, major orthopedic procedures, eye procedures, ambulatory procedures, minor procedures, oncology procedures, endoscopy, and dialysis)
Imaging	Carrier claims with BETOS = I**	Imaging (IMG) actual Medicare costs (advanced imaging, standard imaging, echography, and imaging procedures)
Lab	Carrier claims with BETOS = T1*	Lab Tests (LABTST) actual Medicare cost
Other Tests	Carrier claims with BETOS = T2*	Other Tests (OTHTST) actual Medicare cost
Part B drugs	Carrier claims with BETOS: 01D chemotherapy 01E other drugs	Part B Drug (DRUG) actual Medicare cost (drugs paid for under Medicare Part B; does not include payments made to plans under Medicare Part D)
Part B other	All other carrier claims (except DME) including: P01 Anesthesia 01B chiropractic 01C enteral and parenteral	Part B Other (OTHER) actual Medicare cost (anesthesia, ambulance services, chiropractic services, enteral and parenteral nutrition services, vision/hearing/speech services, and services provided in outpatient rehab facilities, comprehensive outpatient rehab facilities, and community mental health centers)

BETOS = Berenson-Eggers Type of Service codes. This is a system used to categorize claims.[28]

[28] Description of BETOS system is available at https://www.cms.gov/apps/ama/license.asp?file=/MedHCPCSGenInfo/downloads/betpuf08.zip

Step 5. Aggregate the adjusted care components within HRR to create HRR-specific adjusted national base rates.

Finally, we aggregated the adjusted payment components across care settings included in the episode to create an HRR-specific adjusted national payment rate.

An alternative approach to Steps 1-5 would involve (1) calculating a 5% reduction off of median payments for each care setting (or some discount off of median payments in each care setting), (2) adjusting each component using the HRR-specific adjustment factors, and (3) summing adjusted components across care settings to create the HRR-specific adjusted national payment rate. However, the sum of median payments in each care setting does not necessarily equal the median of overall bundled episode spending. We thought it was conceptually clearer to base the bundled rate as a percentage reduction off the median of historical bundled episode payments rather than the sum of median component payments. The results are unlikely to vary significantly between the two approaches.

Strengths and weaknesses of standardization approach

This approach is straightforward to implement and the standardization is consistent with that in the baseline CMS database. In reality, these adjustments would be made at a provider rather than HRR-level, but this approach preserves variation in spending at the HRR-level, which is the outcome of interest for the study.

Our results for Scenario 3 (lower of the national base rate and a hospital-specific historical payment) may differ from an approach that makes adjustments at the provider level. However, we do not anticipate this to dramatically impact estimated changes in geographical variation in spending.

Calculation of HRR-Level Per Beneficiary Medicare Spending on Bundled Services and HRR-Level Total Medicare Spending Under Baseline and Alternative Scenarios

Our main results are HRR-level per beneficiary Medicare spending on bundled services and total per beneficiary Medicare spending at baseline (i.e. under historical episode spending) and under each alternative payment rate scenario. In each case, we start with episode-level data with the historical payment and each alternative rate. Then, we aggregate the episode-level data to the HRR-level.

For results examining per-beneficiary Medicare spending on bundled services, we divided the aggregate spending by the number of Medicare beneficiaries in the HRR (from the IOM spreadsheet described in the prior section). Then, we calculated average per-beneficiary spending across HRRs (weighted by the number of beneficiaries) and variation in per-beneficiary spending across HRRs as measured by across-HRR coefficient of variation (using HRR-level standard deviation divided by the unweighted across-HRR mean) and the 75^{th} to 25^{th} percentile (in the HRR-level distribution).

For results examining per-beneficiary total Medicare spending, we calculated *changes* in aggregate HRR-level payments under each payment rate scenario relative to the baseline case. Then, we applied these changes to the total Medicare spending from the IOM spreadsheet to find the new level of total Medicare spending, and divide by the number of beneficiaries. Finally, we found average across HRR total Medicare spending per beneficiary (weighting by number of beneficiaries), the HRR-level coefficient of variation, and 75th to 25th payment ratio (in the HRR-level distribution).

Standardization Strategy in Provider-Level Episode Comparisons

Finally, we compared standardized measures of Medicare spending for each component of the bundle at the provider level. In order to ensure that the provider-level results are not driven by low-volume providers, we enforced a volume threshold of 10 to be included in the analysis. Because these comparisons are at the provider rather than HRR-level, the HRR-level adjustment factors are less appropriate. Instead, we used an approach developed by Gottleib et al. (2010) to standardize each component of the bundle. Due to missing information needed to construct standardized payments, we drop 6% of episodes compared to the analyses with non-price standardized data.

Wage Index

First we generated a wage index measure. We used the CBSA-specific Medicare wage index for 2008. Then we found the following adjusted wage index.

$$WI_m = 0.25 + 0.75 * (\text{CBSA wage index}) \qquad (1)$$

Initial Acute Payments

To generate the index hospitalization standardized payments, we calculated (2) below:

Acute hospital expenditures (HE) for condition *k* for individual *i* and region *j*

$$HE_{ikj} = P * DRG_k + OUTLIER_{ik}/WI_j \qquad (2)$$

P= national DRG price in FY 2008 or 2009
DRG_k= the weight for procedure k
$OUTLIER$ is the outlier payment on claim
WI index factor calculated in (1)

Readmission Acute Payments

We used the same formula as in the initial acute, using the relevant MS-DRG weight if it is a different than the original, but prorating the amount if the episode straddled the 30-day post-discharge period.

Skilled Nursing Payments

For skilled nursing we normalized actual payments by the wage index in 1. Again, we prorated if the skilled nursing stay straddles the 30-day post-discharge period.

$$\text{Standardized payments} = \text{Actual payments} / \text{wage index}$$

Long-Term Care Hospital Payments

For long-term care hospitals, similar to skilled nursing, we normalized the actual payments by the wage index in (1), and then prorate.

$$\text{Standardized payments} = \text{Actual payments (including outliers)} / \text{wage index}$$

Home Health Payments

Gottlieb does not propose a standardization approach for home health care.(Schoen et al., 2012) We used the same wage-index methodology because labor adjustments are applied to 77% of home health payment close to the 75% weighting in the formula above.

$$\text{Standardized payments} = \text{Actual payments} / \text{wage index}$$

Inpatient Rehabilitation Facilities

For IRFs, we took total payments, subtracted IME and DSH adjustments from the claim, and deflated any "rural" adjustment to the payment. Then, we adjusted by the wage index as above. Finally, we prorated if the IRF stay straddled the 30-day post-discharge episode.

Appendix 3 – Accountable Care Organizations (ACO) Methods

This Appendix describes in more detail the methods and data used to model the impact of ACOs on Medicare spending and geographic variation in Medicare spending. We describe our methods for the analysis of the geographic distribution of beneficiaries assigned to ACOs and the savings to the Medicare program that could result, at the level of the HRR. We describe separately the methods used for assigning beneficiaries to Medicare ACOs, private ACO-like entities, and 'projected' or forecast ACOs. To be consistent with the other two policy interventions, we excluded HRRs located in Maryland from our analysis.

Geographic Distribution of Beneficiaries Covered by Medicare ACOs

We first estimated the geographic distribution of the beneficiaries covered by ACO's participating in Medicare's existing ACO programs. For the Pioneer ACO program, data on the geographic distribution of beneficiaries were obtained directly from representatives for the Centers for Medicare and Medicaid Innovation (CMI).

However, similar data on ACOs participating in the Medicare Shared Savings (SSP) program were not available and consequently had to be approximated using several sources of data as described in the following section.

CMS selected ACOs to participate in the SSP program in two rounds. The first round was completed in April 2012 and covered 27 ACOs. In their announcement, the agency provided information on the number of beneficiaries expected to be managed by each of the participating provider organizations. However, information on the locations of each ACO's beneficiaries was not released. In order to approximate the HRR locations of the beneficiaries for the April ACOs, we started by using each ACO's website and physician network search engines to estimate the number and location of each ACO's primary care providers (PCPs), specifically those physicians specializing in internal medicine, family medicine, and geriatric medicine. The beneficiaries covered by each ACO were then allocated to HRRs in proportion to the geographic location of the ACO's PCPs. When information on PCP locations was not available, we allocated beneficiaries in proportion to distribution of provider clinics or hospitals by HRR, where we placed an equal weight on each clinic or hospital.

In July, a second round of 89 ACOs was selected to participate in the SSP program and projected by CMI to cover 1.2 million Medicare beneficiaries. However, as of this writing, the agency has not released information on how many beneficiaries each individual ACO is expected to manage, thus presenting the dual challenge of approximating both the number and location of each ACO's beneficiaries. We approximated the number of beneficiaries in each ACO as being proportional to the number of PCPs that we identified as being associated with the ACO, with the additional caveat that each ACO had a minimum of 5,000 beneficiaries. PCPs were identified using the same approach described for the April ACOs. To approximate the geographic location

of beneficiaries, we again used publicly available information on the location of each ACO's PCPs, clinics, and/or hospitals.

Unlike the April SSP ACOs, some of the 89 July SSP ACOs did not publish sufficient PCP, clinic, or hospital information on their websites to approximate the number of beneficiaries in each ACO and their location. In those cases, we attributed an average number of beneficiaries to the ACO based on the location of the ACO headquarters – that is, the geographic distribution of the beneficiaries was approximated by allocating 50 percent to the HRR where the ACO's headquarters was located and the remaining 50 percent evenly across adjacent HRRs.

Geographic Distribution of Beneficiaries Covered by Private ACO-like entities

In some of our scenarios, we incorporated information on private ACOs. These ACOs were identified using data on integrated provider organizations and provider groups that have expressed interest in developing ACOs through the following ACO collaborative and learning networks:

- Dartmouth-Brookings ACO learning network
- Council of Affordable Physician Practices
- CMS physician group demonstration participants29
- Premier's Partnership for Care Transformation Implementation Initiative

We initially identified a total of 81 prospective ACOs, after ruling some out that were insurer or pharmaceutical-based organizations and did not include physician or hospital providers. We excluded 4 additional organizations because of insufficient information on their provider networks leaving us with a final total of 77 prospective ACOs. The geographic distribution of the beneficiaries covered by these prospective ACOs was estimated using a method similar to the one employed for the July wave of the SSP program. First, for each prospective ACO, the PCPs and their respective location were identified using the provider organization's website physician network search feature. This geographic distribution was then converted to an approximation of the geographic distribution of covered beneficiaries using the HRR specific average Medicare Fee-For-Service patient panel size. The total number of beneficiaries covered by each prospective ACO was then scaled down by 50 percent to account for limited participation among each organization's total network of PCPs and to be consistent with the average size of the Medicare ACOs.

Once we developed a baseline approximation for the geographic distribution of Medicare beneficiaries participating in existing ACOs and potential ACOs, we employed several steps to convert that to a percentage participation in ACOs among an HRR's FFS beneficiaries. First, our ACO enrollee data were for 2012, but all of the modeling work is based on the 2008 baseline data. Thus, we had to approximate Medicare FFS enrollment in 2012 (which is not currently

[29] This program is not private and is run by CMS. It has considerable overlap with the ACO program in design.

available) by HRR to provide a comparable denominator. Toward this end, we extrapolated FFS enrollment based on the trend observed from 2007 to 2010.

A second step was required because our enrollment numerator and FFS enrollment in the denominator were not based on consistent universes of beneficiaries. Specifically, the 2008 baseline data we used comprised only elderly FFS beneficiaries while our ACO enrollment figures include some disabled and ESRD beneficiaries who are under 65. We further adjusted our denominator to account for these enrollees.

In HRRs where the number of allocated beneficiaries (the numerator) exceeded the recorded total number of beneficiaries (the denominator), we capped the total allocation at 80 percent of recorded number of beneficiaries. This allocation ceiling was applied only to the Owensboro, KY HRR.

Projecting the Geographic Distribution of Beneficiaries Covered by future ACOs ('projected' ACOs)

In order to examine the effects of an expanded Medicare ACO program (Scenarios 3-5), we needed to project the geographic distribution of beneficiaries covered by ACOs for ACOs that do not yet exist. Based on feedback from the Committee, we chose to target a national participation rate among Medicare FFS beneficiaries of 20% - double the participation rate resulting from the allocation of beneficiaries from the Medicare ACO programs and the Private ACOs (10.1%). We assumed that the ACOs with the additional 9.9% of FFS beneficiaries would be formed in HRRs that were similar to HRRs currently with ACOs. Specifically, we used Medicare and private participation, at the HRR level as the dependent variable in a regression with HRR-level factors on the right hand side, and imputed additional FFS ACO participation in proportion to the predicted values from the regression.

We adopted a fractional logit model (Papke and Woolridge, 1996) rather than a linear probability model for two reasons. First, since the dependent variable is in percentage points, applying a linear probability model would result in various issues such as heteroscedasticity, a non-normal error term, and nonlinearity in the effects of predictors. Second, predicted penetration rates could either be less than zero or greater than one. We used a generalized linear model with a logit link function and a binomial family. The model was estimated using the quasi-likelihood method. It takes the following form:

$$g^{-1}(P_i) = \beta_0 + X_i\beta_1$$

Where g is a logit link function, the family is binomial, P_i is the ACO penetration rate in HRR i, X_i is a vector of HRR level characteristics of HRR i, βs are parameters to be estimated. Robust standard errors were used in the model. HRR level characteristics included population, income, physician supply, Medicare Advantage penetration, hospital market characteristics, Medicare spending and its growth rate. Specific measures and data sources are shown below.

Table A.3.1. Independent Variables In The ACO Penetration Prediction Model

Independent variables	Year	Data source
Logged population	2010	Census Bureau[30]
Logged population density (# per square mile)	2010	Census Bureau[31]
Logged median household income	2008	Dartmouth[32]
PCP per 100,000 residents	2006	Dartmouth[33]
Proportion of physicians in a group with 10 or more physicians	2010	SK&A[34]
Proportion of PCPs accepting Medicare patients	2010	SK&A
MA penetration rate	2008	Medicare Claims
Hospital Herfindahl Index	2010	AHA
Proportion of hospitals affiliated with a system	2010	AHA
Proportion of hospital revenue from capitation or risk-sharing contracts	2010	AHA
Proportion of hospitals that have a joint venture with physicians or physician groups	2010	AHA
Medicare spending per beneficiary (actual costs)	2010	IOM[35]
Growth rate in Medicare spending per beneficiary	2007-2010	IOM

Marginal effects were calculated based on one standard deviation change in an independent variable of interest while all other variables were set at the mean level.

Following the regression, we scaled our baseline estimates of beneficiary participation in ACOs according to the predicted participation rates under an expanded Medicare ACO program until total participation reached 20%

Expected Medicare ACO Program Savings Rate

The savings rates assumed for ACOs and their justification are described in the scenario description in the main body of the report. We made several additional assumptions in the implementation of those savings rates. First, in the case of private and projected ACOs, we had to allocate their participants to either the SSP-level savings rate or the Pioneer level. Those

[30] Available at: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml. Accessed on 08/09/2012.
[31] Population data are from the US Census 2010, available at: http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml. Accessed on 08/09/2012. Land area data are from Dartmouth, Available at: http://www.dartmouthatlas.org/tools/downloads.aspx?tab=37. Accessed on 09/25/2012.
[32] Available at: http://www.dartmouthatlas.org/tools/downloads.aspx?tab=37. Accessed on 09/25/2012.
[33] Available at: www.dartmouthatlas.org/downloads/tables/2006_hosp_phys_hrr.xls. Accessed on 08/07/2012.
[34] SK&A is a private vendor that maintains provider lists. RAND has purchased their office-based physician file, which contains several key characteristics of those physicians such as those noted in the table.
[35] Available at: /iom.edu/Activities/HealthServices/GeographicVariation/~/media/Files/Activity%20Files/HealthServices/GeographicVariation/CMSfiles/Table1.xlsx. Accessed on 09/25/2012.

allocations were made in proportion to existing ratios of SSP and Pioneer participation within the HRR.

A second adjustment made to savings estimates was based on the fact that the ACO spending targets are updated based partly (in the case of the Pioneers) or fully (in the case of the SSP ACOs) on national average increases in Medicare spending. Thus, ACOs in areas of the country that face systematically high growth in spending would have to achieve even greater reductions in spending estimates (i.e., they begin at a relative disadvantage) to achieve shared savings and, likewise, are more likely to end up paying Medicare for shared losses (McWilliams and Song, 2012).

To account for this effect, we modeled adjustments in Medicare savings attributable to ACOs accounting for each ACO's ability to reduce its own expected spending relative to its target (based on the HRR where its enrollees were located). These adjustments were applied at the HRR level and calculated in three steps. In the first step, we approximated the counterfactual per beneficiary spending in an HRR absent any ACO intervention – assuming prior local growth trends continue. This was estimated by inflating the baseline estimates of HRR per beneficiary costs by their average spending growth rates for the prior three years as estimated by McWilliams (McWilliams and Song, 2012). The second step was to apply the assumed behavioral savings rates described in the main report above. The last step was to calculate the effective per beneficiary savings by multiplying the difference between the baseline estimates of HRR per beneficiary costs and the Medicare per beneficiary spending calculated in the second step by 70 percent, the rate at which savings relative to the Medicare benchmark are remitted to ACOs. The sum of the per-beneficiary savings remitted by Medicare and the realized Medicare per beneficiary spending represents the total per beneficiary spending by Medicare. The ratio of this sum to the counterfactual per beneficiary spending calculated in the first step represents the total spending as a percent of spending absent intervention. The final effective HRR savings rate is represented by one minus this ratio.

Estimating Effective Savings by HRR

For a given geographic distribution of beneficiaries covered by ACOs participating in the Pioneer and SSP program, the effective savings rate for an HRR was calculated by first multiplying the expected savings rate for each program (as defined by the scenario) by the proportion of beneficiaries within an HRR that was covered by an ACO participating in the corresponding program. The sum of the products was taken to represent the effective savings rate.