Specialty Payment Model Opportunities and Assessment

Oncology Simulation Report

A product of the **CMS Alliance to Modernize Healthcare (CAMH)**
Federally Funded Research and Development Center

Sponsored by the Centers for Medicare & Medicaid Services (CMS)

Chapin White, Chris Chan, Peter J. Huckfeldt, Aaron Kofner,
Andrew W. Mulcahy, Julia Pollak, Ioana Popescu,
Justin W. Timbie, Peter S. Hussey

Published by the RAND Corporation, Santa Monica, Calif. and The MITRE Corporation, McLean, Va.

For more information on this publication, visit **www.rand.org/t/rr799**

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

The MITRE Corporation is a not-for-profit company that operates multiple federally funded research and development centers (FFRDCs). MITRE provides innovative, practical solutions for some of our nation's most critical challenges in defense and intelligence, aviation, civil systems, homeland security, the judiciary, and healthcare.

www.rand.org
www.mitre.org

Preface

In August 2013, the Centers for Medicare & Medicaid Services (CMS) issued a task order to The MITRE Corporation (MITRE), operator of the CMS Alliance to Modernize Healthcare (CAMH) Federally Funded Research and Development Center (FFRDC). The goal of this task order was to inform the development of alternative payment models for specialty health care services. Claire Schreiber and Mary Kapp have served as the Government Task Lead (GTL) for this work.

This report describes the results of a simulation analysis of a payment model for specialty oncology services. The research addressed in this report was conducted in RAND Health, a division of the RAND Corporation, under a subcontract to MITRE. A profile of RAND Health, abstracts of its publications, and ordering information can be found at http://www.rand.org/health.

Table of Contents

Figures

Tables

Summary

This report describes the results of a simulation analysis of a payment model for specialty oncology services that is being developed for possible testing by the Center for Medicare and Medicaid Innovation at the Centers for Medicare & Medicaid Services (CMS). CMS asked The MITRE Corporation (MITRE), operator of the CMS Alliance to Modernize Healthcare (CAMH) Federally Funded Research and Development Center (FFRDC) and RAND to conduct simulation analyses to preview some of the possible impacts of the payment model and to inform design decisions related to the model.

Data and Methods Used in the Simulation Analysis

The simulation analysis used an episode-level dataset based on Medicare fee-for-service (FFS) claims for historical oncology episodes provided to Medicare FFS beneficiaries in 2010. Spending on those historical episodes was inflated to 2016, which is the first year in which the proposed payment model could be implemented. Under the proposed model, participating practices would continue to receive FFS payments, would also receive per-beneficiary per-month (PBPM) care management payments for episodes lasting up to six months, and would be eligible for performance-based payments based on per-episode spending for attributed episodes relative to a per-episode spending target. For the simulation analysis, the care management payments were set at $160 PBPM; that amount has not yet been finalized by CMS. RAND assumed that only practices providing 50 or more episodes of chemotherapy treatment per year would consider participating and would be eligible to participate in the payment model—these medium- and high-volume oncology practices account for 73 percent of all chemotherapy episodes and 80 percent of total spending on chemotherapy episodes. Among those practices providing 50 or more episodes of treatment per year, 10 percent were assumed to participate in the payment model. For participating practices, the simulation model was used to calculate care management payments, a spending target, and actual spending, taking into account the fact that participating practices have an incentive to reduce spending on attributed episodes. The simulation model was then used to calculate performance-based payments, taking into account the practices' behavioral response.

Conclusion

The simulation described in this report offers several insights into the proposed payment model for oncology:

1. The care management payments used in the simulation analysis—$960 total per six-month episode—represent only 4 percent of projected average total spending per episode (around $27,000 in 2016), but they are large relative to the FFS revenues of participating oncology practices, which are projected to be around $2,000 per oncology episode. By themselves, the care management payments would increase physician practices' Medicare revenues by roughly 50 percent on average. This represents a substantial new outlay for the Medicare program and a substantial new source of revenues for oncology practices.

2. The use of performance-based payments requires the assignment of spending benchmarks, but those benchmarks cannot reflect the desired counterfactual—i.e., spending per episode in the absence of the payment model. Inaccuracies in benchmarks arise due to variation in individuals'

health care spending that cannot be explained by case-mix adjusters. That unexplained variation affects both the historical claims data used to set benchmarks and actual spending in the performance period. As a result, the Medicare program will in some cases pay “noise bonuses”: performance-based payments that are due to inaccurately low benchmarks rather than to providers’ behavioral responses. The likelihood of noise bonuses would be reduced if the payment model were limited to larger practices (e.g., 100 or more episodes).

3. For the Medicare program to break even, participating oncology practices would have to reduce utilization and intensity by roughly 4 percent. (“Breaking even” in this case means that Medicare spending per episode would be the same, with or without the payment model.) Based on research on other similar payment models, a behavioral response of that magnitude may be possible, but it is by no means certain.

4. The break-even point can be reduced—i.e., less of a behavioral response would be required for the Medicare program to achieve savings—if the care management payments are reduced or if the performance-based payments are reduced.

Acknowledgments

We thank Paul B. Ginsburg of the University of Southern California and Sarah Nowak of the RAND Corporation, whose thoughtful reviews and comments greatly improved this report.

We also appreciate feedback on our analysis from Claire Schreiber, Mary Kapp, and other members of the Patient Care Models Group at CMS, as well as Kavita Patel at the Brookings Institution. We are grateful to our colleagues at MITRE, particularly Heidi Giles, Ren Resch, Dan Speece, and Ken Hoffman, for their guidance and counsel. We thank Shawna Beck-Sullivan and Susan Lovejoy at RAND for project support.

Abbreviations

ACE	Acute Care Episode
ACO	accountable care organization
BPCI	Bundled Payment for Care Improvement
CCW	Chronic Conditions Warehouse
CMS	Centers for Medicare & Medicaid Services
DME	durable medical equipment
FFS	fee-for-service
MedPAR	Medicare Provider Analysis and Review
OCM	Oncology Care Model
PBPM	per-beneficiary per-month
PCMH	patient-centered medical home
TIN	Tax Identification Number

1. Background

This report describes research related to the simulation of the possible effects of a payment model for specialty oncology services. That payment model is being developed for possible testing by the Center for Medicare and Medicaid Innovation at the Centers for Medicare & Medicaid Services (CMS). This report builds on the methods and findings in a previous report on the design of an oncology payment model by The MITRE Corporation (MITRE) and RAND (Huckfeldt et al., 2014; referred to hereafter as the “Model Design Report”) that examined the design features of the oncology payment model.

As described in the Model Design Report, episode-based payment, which aims to create incentives for high-quality, low-cost care, has been identified as a promising alternative payment model for oncology care (Bach, Mirkin, and Luke, 2011; McClellan et al., 2013). CMS has already implemented episode-based payment models in several settings. Under the Acute Care Episode (ACE) demonstration, which was first implemented in 2009, physician-hospital organizations receive a single payment for surgical episodes, including physician and facility services (Congressional Budget Office [CBO], 2012). As required by the Affordable Care Act, CMS is implementing a voluntary Bundled Payment for Care Improvement (BPCI) initiative, which expands on the ACE demonstration. Episode-based payments have also been implemented nationally in Medicare for home health services and dialysis. Episode-based payment systems can provide flexibility to health care providers to select among the most effective and efficient treatment alternatives, including activities that are not currently reimbursed under Medicare payment policies (Bach, Mirkin, and Luke, 2011). However, the model design also needs to ensure that high-quality care is delivered and that beneficial treatments are not withheld from patients.

CMS asked MITRE and RAND to conduct analyses to inform design decisions related to an episode-based oncology model for Medicare beneficiaries undergoing chemotherapy treatment for cancer. Based on an analysis by the Moran Company of 2008–2009 Medicare claims data, McClellan et al. (2013) report that chemotherapy and its administration account for approximately 20 percent of Medicare spending on oncology care. The remaining 80 percent of spending includes surgery, radiation therapy, and ongoing management and surveillance; spending in these areas was also studied as part of our analyses.

Description of the Payment Model

CMS is developing an Oncology Care Model (OCM) that is intended to create financial incentives that will promote high-quality care while reducing program expenditures. To be eligible to participate in OCM, oncology practices would need to have certain capabilities, such as 24-hour access to a clinician and electronic health records, and they would have to commit to participating in the model for up to five years. CMS intends for other payers to participate in the payment model as well, so that participating practices will face similar payment incentives for Medicare and non-Medicare patients. Participating practices will be eligible for two additional payments for each episode of oncology care, in addition to standard fee-for-service (FFS) payments:

1. Each practice will receive a per-beneficiary per-month (PBPM) care management fee for each episode of chemotherapy. The care management payments are intended to support provision of enhanced services and could be invested in additional staffing (such as patient navigators) and

infrastructure (such as electronic health records) or in management and care coordination services that are not separately billable under the physician fee schedule.

2. Practices will be eligible for performance-based payments if they have satisfactory quality metrics and if their spending per chemotherapy episode falls below a target. The methodology for assigning those spending targets has not been finalized. Practices' participation in OCM could be terminated by CMS after three years if they fail to demonstrate savings.

OCM will include two options for performance-based payments: "one-sided" and "two-sided." Under both options, each practice will be assigned a per-episode spending target. The one-sided option is "upside only," meaning that practices would receive performance-based payments if they meet quality metrics and their actual spending falls below the target. Under the two-sided option, practices face upside and downside financial risk, and they could receive performance-based payments or could be liable for amounts owed to CMS if spending exceeds the target. For practices fully satisfying quality benchmarks, the performance-based payments would equal 100 percent of the difference between target spending and actual spending. OCM would include a ceiling on the possible performance-based payments (e.g., performance-based payments would be capped at 20 percent of the spending target) and, in the two-sided option, a limit on amounts owed to CMS. There are at least two rationales for including those limits:

1. They reduce the financial risks of the payment model—both for the Medicare program and for participating practices.
2. They somewhat dampen the incentives for practices to either stint on necessary care or select an unusually healthy patient population.

CMS is continuing to refine the payment model, and some design features have not been finalized, such as the method for assigning spending benchmarks to participating practices. To the extent possible, the simulation modeling followed CMS's description of the proposed model, although with several simplifying assumptions. The simulation analysis only included Medicare patients, and so it does not reflect participation by other payers or any possible impacts among non-Medicare patients. For the simulation analysis, the model design features were as follows: Episodes of chemotherapy care began with initiation of chemotherapy and ended six months later, and they included all Medicare-covered services and prescription drugs provided in that window. Practices with Medicare FFS beneficiaries initiating chemotherapy (as defined in the Model Design Report) for cancer were eligible for participation. Although the proposed payment model would not include a minimum volume of oncology episodes per se, for the simulation modeling we assumed that only practices with 50 or more oncology episodes in a year would be eligible to participate. The simulation analysis focuses only on 2016—the first performance year—and does not include any modeling of changes in behavior over time or termination of OCM for practices that fail to demonstrate savings.

The simulation analysis included eight prevalent types of cancer: breast cancer, colorectal cancer, leukemia, lung cancer, lymphoma, ovarian cancer, pancreatic cancer, and prostate cancer. As described in the Model Design Report, episodes of chemotherapy treatment were attributed to physician group practices using a prospective attribution rule that attributed each patient to the practice responsible for the trigger chemotherapy claim (i.e., the claim that is used to identify the initiation of the chemotherapy treatment episode). Each patient in the simulation had one six-month episode of chemotherapy. The services included in the episode included chemotherapy treatment and all other Medicare-covered services provided to the patient during that six-month episode. The episode began with the start of chemotherapy treatment and ended at six months or earlier if the patient died or disenrolled from the FFS program.

The simulated payment model included a $160 monthly PBPM care management fee and one-sided performance-based payments for practices with episode costs falling below a target. The amount of the PBPM payments has not yet been finalized. However, $160 was used to simulate the effects of the PBPM payments for the purposes of this report. In the simulation, we did not include specific quality measures. Based on CMS's description of the proposed model, we used the simplifying assumption that participating practices received 90 percent of the potential performance-based payment, with the remaining 10 percent representing payments withheld due to subpar performance on quality metrics.

In the simulation analysis, each practice was assigned a simulated spending benchmark based on its historical costs per episode and a simple case-mix model that includes the patient's age, state of residence, dual status, and type of cancer. The approach to assigning benchmarks has not been finalized by CMS and may include regional or national data on historical episode costs. In the one-sided option, the spending target equals a benchmark minus 4 percent, while under the two-sided option the target equals the benchmark minus 2.75 percent. Under the one-sided option, practices may receive performance payments equal to 100 percent of the amount by which actual spending falls below the target, while in the two-sided option practices keep or pay 100 percent of the difference between their actual spending and their target. Under both options, performance-based payments (or amounts owed by the practice under the two-sided model) were capped at 20 percent of the target. In addition to the care management fee and the performance-based payments, the simulated payment model assumed that all other FFS Medicare payments for services continued as usual.

Report Overview

Simulation modeling can assist in the development of payment models by assessing the impacts of various design features. The simulation model integrates three types of information:

1. historical data on spending on chemotherapy episodes, including the variation in spending across episodes and physician practices;
2. the model design features, converted into parameters that can be entered into the simulation model; and
3. the range of likely behavioral responses, based on evidence from the literature.

However, because the evidence available to inform potential behavioral responses is limited, the simulation results here focus on a simple set of potential behavioral responses and outcomes.
The simulation focuses on predicting Medicare spending under the payment model scenarios. We do not simulate other potential outcomes, such as changes in patient health outcomes.

This report is organized as follows. In Chapter 2, we describe the construction of an episode-level "baseline" dataset that reflects our projection of spending patterns for chemotherapy episodes in 2016 in the absence of the payment model. In Chapter 3, we describe the methods for simulating provider participation, behavioral responses to the payment model, and Medicare program payments. In Chapter

4, we summarize some of the evidence from the literature on the range of possible behavioral responses to the payment model. In Chapter 5, we summarize the key outputs of the simulation model—simulated total spending on chemotherapy episodes and simulated oncology practice revenues—using different combinations of model design features and assumed behavioral responses. In Chapter 6, we present the results of alternative model designs. In Chapter 7, we provide conclusions.

2. Construction of an Episode-Level "Baseline" Dataset

As the basis for the simulation modeling, we created an episode-level dataset based on all episodes of chemotherapy initiation for Medicare beneficiaries in 2010 for the eight targeted types of cancer, as described in the Model Design Report (Huckfeldt et al., 2014). We used the first chemotherapy drug claim in 2010 (with no prior claims within six months) as the marker of the initiation of an episode of care. For the simulation, episodes terminated six months after initiation or earlier if the beneficiary died. That episode length was chosen in consultation with CMS, taking into account RAND's analysis of the patterns of spending over time relative to initiation of chemotherapy episodes. The Model Design Report describes many other options for terminating episodes and explores some of the implications of using episodes of different lengths. The dataset included information on the beneficiary's demographics and location of residence, the type of cancer (based on International Classification of Diseases, Ninth Revision [ICD-9] primary diagnosis codes reported on claims), actual spending on the episode, and the attributed practice, with spending amounts inflated to 2016. That episode-level dataset is referred to as the "baseline," meaning that it represents our projection of the patterns of treatment and spending on oncology episodes in the absence of the proposed payment model. The baseline dataset did not include some pieces of information that CMS would use to implement the payment model, such as measures of beneficiary health status and comorbidity and measures of oncology practice characteristics and quality of care. The simulated effects of the payment model are calculated by comparing simulated treatment and spending patterns against this baseline.

Data Sources

The primary data source for the baseline dataset was Medicare claims data from the Chronic Conditions Warehouse (CCW) for 100 percent of Medicare FFS beneficiaries in 2009–2012. We used claims data for all types of Medicare covered services, including the Carrier,[1] Medicare Provider Analysis and Review (MedPAR), Outpatient, Durable Medical Equipment (DME), Home Health, Hospice, and Part D files. We also used the Master Beneficiary Summary File as a source of information on beneficiary demographic characteristics and Medicare eligibility.

In practice, all cancer types will be included in the oncology payment model, although the less-prevalent cancers may be excluded from the calculation of performance payments. The simulation model only included episodes assigned to one of eight prevalent types of cancer (see Table 2.2). Each episode was attributed to a physician and a physician practice as described in the Model Design Report. We used Tax Identification Numbers (TINs) to identify physician group practices.

Spending Categories

For the purposes of the simulation model, spending was grouped into 11 categories, shown in Table 2.1. Spending in each of those categories was inflated from 2010 to 2016 using a category-specific inflation factor (see the appendix for details). The spending categories also varied in the share of total spending

[1] The Carrier file comprises claims for services rendered by noninstitutional providers.

that was treated as net revenue to the attributed oncology practice (also shown in Table 2.1). Those simulated physician practice revenues are used as a rough benchmark for comparison with the proposed new revenue sources under the payment model. All of the revenues from administration of chemotherapy were treated as revenues to the physician practice, as were revenues from evaluation and management (E&M) services provided by the attributed practice. For non–Part D chemotherapy, 6 percent of total spending was treated as revenues to the physician practice, which reflects Medicare's 6-percent markup to the average sales price for physician-administered drugs. For the remaining categories, none of the spending was treated as revenue to the physician practice.

Table 2.1. Spending Categories in the Episode-Level Dataset

Spending Category	Inflation Factor (2010 to 2016)	Share of Spending Treated as Revenue to the Attributed Practice	Mean Spending per Episode (2016)	Mean Physician Practice Revenue per Episode (2016)
Chemotherapy, office-based	1.35	6%	$5,054	$303
Administration of office-based chemotherapy	1.06	100%	$483	$483
Chemotherapy, hospital outpatient or DME	1.35	6%	$2,699	$162
Chemotherapy, Part D	1.35	0%	$766	$0
Administration of chemotherapy, hospital outpatient	1.22	100%	$377	$377
Evaluation and management by attributed physician practice	1.06	100%	$450	$450
Professional services provided by non-attributed practices	1.06	0%	$3,088	$0
Emergency department visits (facility fee)	1.22	0%	$204	$0
Hospital inpatient and outpatient services (other than emergency department)	1.15	0%	$8,086	$0
Radiation therapy	1.22	0%	$2,477	$0
All other services (hospice, skilled nursing, home health, etc.)	1.02	0%	$2,957	$0
All	n/a	6.7%	$26,641	$1,776

SOURCE: Authors' analysis. The inflation factors are calculated from Bureau of Labor Statistics Producer Price

Indices, extrapolating 2010–2014 price growth from to 2016. The spending amounts are calculated from the episode-level baseline dataset.

The simulation model has the capability to allow each of these spending categories to be impacted differently by the payment model. In practice, the evidence in the literature is not sufficiently fine-grained to support separate estimated effects for each of those spending categories.

Description of Episodes Used in the Simulation Analysis

The episode-level dataset includes 330,647 oncology episodes, with a mean baseline spending per episode in 2016 of $26,641 (Table 2.2). Prostate cancer and breast cancer are by far the most common, and least expensive, types of cancers. Prostate cancer, in particular, stands out for its low mean spending per episode ($12,968) and its very low mean physician practice revenue per episode ($353).

Table 2.2. Cancer Types Included in the Simulation Modeling

Cancer Type	Number of Episodes	Mean Spending (2016)	Mean Physician Practice Revenue (2016)
Breast cancer	79,700	$17,720	$1,163
Colorectal cancer	30,223	$41,192	$3,737
Leukemia	13,856	$44,847	$2,919
Lung cancer	51,913	$40,193	$2,743
Lymphoma	30,666	$45,558	$3,404
Ovarian cancer	10,235	$30,007	$2,833
Pancreatic cancer	9,329	$40,883	$3,025
Prostate cancer	104,725	$12,968	$353
All	330,647	$26,641	$1,776

NOTES: Spending is inflated to 2016. For detailed definitions of cancer types, see the Model Design Report.
SOURCE: Authors' analysis of the episode-level baseline dataset.

We used the number of chemotherapy episodes as a measure of the volume of the physician group's oncology practice. As will be discussed in more detail below, volume plays two important roles in the payment model:

1. Practices with only a handful of oncology episodes are likely not to be true oncology practices. Based on RAND's analysis of practice characteristics and volume of oncology episodes, practices with low volumes (fewer than 50 episodes) tend to be either primary care or urology practices, whereas practices with moderate or high volumes tend to be oncology practices. Primary care

practices or urology practices may initiate a handful of oncology episodes, but those practices may be less likely to participate in the oncology payment model.

2. Oncology volume also matters because it is more difficult to make an accurate comparison of actual per-episode spending versus benchmarks in practices with lower volume. That difficulty reflects the fact that medical spending varies widely from episode to episode, even within a single practice, due to "sampling error": variation in individuals' health care utilization and spending that is impossible to account for fully in a case-mix model. Under the one-sided payment model, sampling error in setting benchmarks and measuring actual performance can result in performance-based payments, even in the absence of any "true" savings (meaning a reduction in utilization that results from the practice participating in the model). Under the two-sided payment model, sampling error creates financial risks for participating practices as well.

As shown in Table 2.3, over 70 percent of practices with attributed oncology episodes had fewer than five attributed episodes in 2010. Many of these low-volume practices are likely not true oncology practices, and these practices only accounted for 6 percent of all episodes and only 4 percent of all chemotherapy spending. Most episodes of chemotherapy were provided by practices with moderate volume, meaning those providing between 50 and 999 oncology episodes. Although only about 5 percent of practices provided 100 or more episodes, those higher-volume practices accounted for the majority of all episodes and spending.

Table 2.3. Distribution of Oncology Practice Volume

Number of Episodes Provided by Practice	Number of Practices (share of practices)	Number of Episodes (share of episodes)	Cumulative Share of Episodes	Spending (millions, 2016, share of total spending)	Cumulative Share of Spending	Mean Spending per Episode
<5	13,726 (71%)	20,167 (6%)	6%	332 (4%)	4%	$16,454
5–9	1,524 (8%)	10,019 (3%)	9%	178 (2%)	6%	$17,759
10–19	1,250 (6%)	17,185 (5%)	14%	353 (4%)	10%	$20,514
20–49	1,281 (7%)	40,562 (12%)	27%	922 (10%)	20%	$22,719
50–99	759 (4%)	54,400 (16%)	43%	1,422 (16%)	36%	$26,134
100–249	560 (3%)	85,792 (26%)	69%	2,484 (28%)	65%	$28,956
250–999	212 (1%)	83,982 (25%)	94%	2,507 (28%)	93%	$29,857
1,000+	10 (<1%)	18,540 (6%)	100%	612 (7%)	100%	$32,993
All	19,322 (100%)	330,647 (100%)		8,809 (100%)		$26,641

NOTES: "Practices" are defined by the attributed TIN (see Model Design Report).
SOURCE: Authors' analysis of the episode-level baseline dataset.

3. Structure of the Simulation Model

For each episode, we simulate a spending amount in 2016, the formula for which is shown in the appendix. The first step in the simulation is to assign a per-episode spending benchmark to each practice. The next step is simulating whether each practice participates in the model. That participation decision reflects the practice's volume of oncology episodes. In a sensitivity test, the probability of participation is modeled as an increasing function of the generosity of the per-episode spending benchmark; this sensitivity test reflects a situation in which practices know, or can anticipate, that their benchmarks are overstated. Once the participation decision has been simulated, we then simulate the behavioral response (i.e., the change, if any, in the utilization of services as a result of participating in the model), the care management payment, and the performance-based payment. The care management payment is the simplest to model: It equals $160 multiplied by the number of months in the episode, up to six.

Simulating Performance-Based Payments

The performance-based payment is simulated based on each practice's per-episode spending target and its actual per-episode spending, taking into account the care management payment and any behavioral response. The comparison of actual and target spending is based on practice-level average spending per episode, not episode by episode. The payment model allows practices to select either a one-sided option (upside only) or a two-sided option beginning in year three of the model (upside and downside, meaning that practices can earn performance payments or owe the Medicare program). Our simulation modeling only incorporates the one-sided option, based on RAND's assessment that few practices will elect the two-sided option. Electing the two-sided option can increase the performance-based payments by up to 1.25 percent of the target (i.e., the difference between 4 percent and 2.75 percent), but it also puts the practice at risk for paying up to 20 percent of the target back to CMS. For many practices, 20 percent of the spending target—i.e., the amount potentially owed to CMS—would exceed their Medicare FFS revenues from providing oncology care. Practices would only elect the two-sided option if they were very confident that their spending would fall below the target. We do not expect that many practices would feel that level of confidence, due to the high degree of variability in spending on chemotherapy episodes, the large amounts potentially owed, and due to the uneven experience with new payment models documented in the literature.

A simulated performance-based payment per episode was calculated for each practice:

$$perf_{i,j} = targ_j * \min\left(ceiling, \max\left(floor, share * \left(\frac{act_j}{targ_j} - (1 + discount)\right)\right)\right) \quad \text{(eq. 1)}$$

Where

$targ_j$ equals the per-episode target spending for practice *j*

ceiling is a limit on the upside performance-based payment as a percent of the target *floor* is a limit on the downside performance-based amounts owed as a percent of the target

share is the share of the “savings” (or excess spending) that is paid to the practice (or owed by the practice)

$\frac{act_j}{targ_j}$ is the ratio of actual per-episode spending by practice *j* to target per-episode spending for practice *j*

discount is a minimum share by which actual spending per episode must fall below the target before attributing savings to the practice.

Based on CMS’s descriptions of the proposed model and the assumption that all practices would elect the one-sided option, RAND assigned *ceiling* a value of 0.2, *floor* a value of 0 (i.e., no downside), *share* a value of 0.9 (reflecting the assumption that 90 percent of participating practices receive a 100 percent performance payment and 10 percent are deemed ineligible to receive a performance payment), and *discount* a value of 0.04.

One of the key steps in the simulation analysis is assigning target spending per episode to each practice, $targ_j$. We assumed that each practice will be assigned a spending target based on the complexity of the cases and patients it treats in the performance year (“case mix”) and on the actual case mix and costs of patients treated by that practice during an earlier historical period. That spending target will be used to determine performance-based payments.

Assignment of a Simulated per-Episode Spending Target to Each Practice

Our approach to simulating the per-episode spending target for each practice was to begin with actual spending per episode in 2016 (i.e., 2010 spending inflated to 2016) for each practice and then apply a “forecast error.” The forecast error is included to account for the fact that projecting health care spending is inevitably imprecise, particularly for small populations. Each practice’s costs per episode—both in the baseline historical period and in the performance period—will include substantial sampling error, due to the high degree of underlying variability in health care spending on individuals receiving chemotherapy.

In the one-sided option, forecast errors will tend to raise Medicare spending; for a discussion of this general phenomenon, see the Congressional Budget Office (1999). In the one-sided option, Medicare will make performance-based payments to practices with overestimated targets but will not recoup any amounts from practices with underestimated targets. In the two-sided option, forecast errors will not tend to raise or lower Medicare spending, but they will create financial risks for participating practices.

Our approach to simulating the forecast error is described in more detail in the appendix. Briefly, we randomly split each practice’s episodes from 2010 into two roughly equally sized groups and then measured the practice’s estimated case-mix-adjusted historical costs for those two groups of patients. By assumption, the practice applied the same practice patterns and drew from the same patient pool for those two groups of patients—in statistical terms, they are two independent samples from the same population. Therefore, we used the differences between estimated per-episode practice costs for those two groups to simulate the forecast error. This approach does not reflect how CMS will actually calculate benchmarks in practice, but it is useful for quantifying the range of magnitude of the forecast errors and their impacts on Medicare spending.

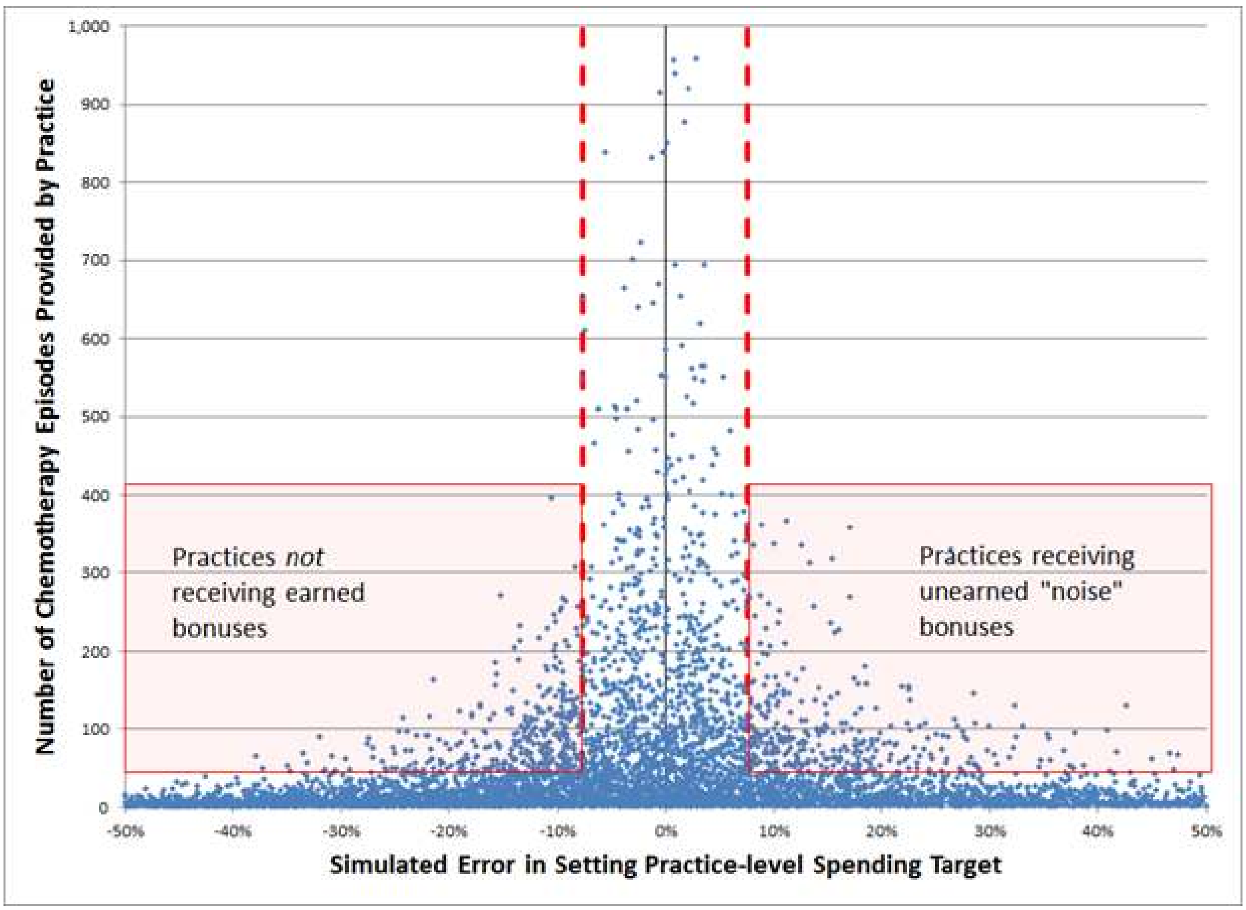
Figure 3.1 illustrates the distribution of simulated forecast errors in setting practice-level per-episode spending targets and the relationship between those errors and the practice's volume of oncology episodes. Forecast errors vary widely, especially for practices with lower volumes of oncology episodes.

Among practices with 50 or more episodes—i.e., those assumed to be large enough to participate in the model—two groups are highlighted in the outlined boxes in the figure. The first group of practices receives unearned "noise" bonuses, meaning that the forecast error overestimates their actual spending by an amount large enough that they would be eligible for performance-based payments even if they did nothing to reduce spending. The second group of practices would not receive earned bonuses, meaning that the forecast error underestimates their actual spending by an amount large enough that they may be ineligible to receive a performance-based payment even if they reduce spending substantially (i.e., enough to offset the case management payments and achieve an additional 4 percent in savings). Among practices providing 50 or more episodes, 11 percent of practices receive "noise" bonuses and a similar share would not receive earned bonuses. In the two-sided option, every participating practice would be impacted by forecast errors—they would either increase the performance-based payments received by the practice or decrease the amounts owed to Medicare by the practice.

CMS is continuing to develop the methodology for assigning benchmarks, and there are several approaches that might reduce forecast errors. Unfortunately, none is perfect. One approach is to use multiple years of historical spending data for each practice. This approach increases the number of episodes used in the calculations of historical spending, but it does not address the fact that spending in the performance year will inevitably include noise due to random variation in patients' medical needs that cannot be captured in a case-mix adjustment system. Including multiple years of historical data also introduces a new source of error due to extrapolation of spending trends over longer periods of time. Another approach is to blend each practice's historical spending with historical spending from other practices that are located in the same region or are similar in some other way. The blended approach could substantially dampen the sampling error in the measurement of historical spending per episode, but there will still be sampling error in spending in the performance year. Blending may also introduce biases if the practices used in the blend tend to be systematically more or less costly than the participating practices. For example, practices that are unusually efficient or that treat an unusually healthy patient population may be the ones that choose to participate in the payment model.

One major potential source of forecast error is a divergence between projected and actual trends in inflation in spending per episode. Practices would, almost certainly, prefer to be informed of a fixed spending benchmark for each type of episode in advance of deciding whether to participate in the payment model. But fixing the benchmark well in advance of the performance period makes the payment model highly vulnerable to errors in inflation forecasts, which have the potential to be large relative to discount factor. The CMS Office of the Actuary (2011) has analyzed the accuracy of its projections of national health expenditures from 1997 through 2009 and found a mean absolute error in projected annual growth rates of around 1 percent, with even larger mean absolute errors in projections of prescription drug spending. Overprojections of that magnitude, if compounded over three or four years, could result in large numbers of practices receiving performance bonuses even if they do nothing to reduce spending on the episodes they provide. Furthermore, for reasons that are not fully understood, annual growth in Medicare spending per beneficiary since 2010 has been several percentage points lower than projections by analysts at CMS and CBO. This surprisingly slow growth demonstrates the possibility that projected spending trends may be significantly overstated. Given the difficulties in projecting spending trends, even over relatively short periods, one alternative is to assign benchmarks to each practice after they have committed to participate but before the performance period; that approach would shorten the period over which inflation is projected and, therefore, the potential magnitude of inflation-based forecast error. A second alternative is not to provide practices with spending benchmarks before they decide to participate and only calculate benchmarks after the performance period is over to reflect actual inflation. This option minimizes errors in inflation forecasts but introduces delays in the calculation and distribution of performance payments, and it creates substantial uncertainty for practices and turns the benchmarks into moving targets.

Figure 3.1. Simulated Errors in Spending Targets: Implications for Bonuses in One-Sided Option



Simulation of Participation Decision

We assumed that participation in the payment model is voluntary. To simulate participation in the model, we first assumed that the very smallest practices (those providing fewer than 50 chemotherapy episodes) would not participate. Those small practices are not true oncology practices in many cases, they would likely face difficulties in meeting the practice capability requirements, and they would likely perceive the potential benefits from participating to be small relative to the administrative effort required to participate.

Among practices providing 50 or more episodes, we assumed a participation rate of 10 percent, with practices drawn randomly; that participation rate was chosen to be roughly consistent with assumptions in CMS's internal planning documents. The actual participation rate could be significantly higher or lower, depending, in part, on details of the payment model that have not yet been finalized. In a sensitivity test, we assumed instead that the probability of participating would be positively related to the forecast error—in other words, practices whose spending target was overstated were assumed to be more likely to participate. Practices might be able to anticipate that benchmarks will be overstated if, for example, those benchmarks are based on a blend of the practice's own historical costs and local average costs, and the practice tends to treat a relatively uncomplicated set of patients.

4. Evidence from the Literature on the Range of Possible Behavioral Responses

There are four lines of research findings that can help inform our expectations regarding the possible impacts of participation in the payment model on total spending (Table 4.1):

1. patient-centered medical homes (PCMHs);
2. oncology treatment guidelines or “clinical pathways” for treatment of specific types of cancer;
3. episode-based payments for oncology; and
4. shared savings models and accountable care organizations.

The proposed payment model for oncology combines some of the elements of a PCMH and a shared savings model. Although the payment model does not incorporate a requirement that participating practices adhere to clinical pathways, experience with those programs offers some insight into the range of possible responses under the proposed model.

Table 4.1. Key Studies to Inform Modeling of Behavioral Responses

Study Types	Study Description	Study Results
PCMH		
Friedberg et al. (2014)	Setting: 32 voluntarily enrolling primary care practices. Intervention: technical assistance and bonus payments for achieving National Committee for Quality Assurance PCMH recognition.	Estimated impact on total spending: increases of 8–9% in total, depending on the year (not statistically significant).
Fifield et al. (2013)	Setting: 18 volunteer primary care practices. Intervention: “tailored practice redesign support” and per member per month bonus payments.	Estimated impact on total spending: “There was no cost savings observed on any cost-of-care measures”
Werner et al. (2013)	Setting: 8 invited primary care practices in New Jersey. Intervention: infrastructure funding, a care management fee, quality- and utilization-based performance payments.	Estimated impact on total spending: increase of 2% (not statistically significant).

Study Types	Study Description	Study Results
Maeng et al. (2012)	Setting: 43 primary care clinics in the Geisenger system. Intervention: clinics converted to PCMH between 2006 and 2010; changes include care management and pay-for-performance reimbursement.	Estimated impact on total spending: Decreases from 3% to 11%, depending on the statistical model and the patient's exposure to the PCMH.
Reid et al. (2010)	Setting: one Seattle clinic chosen as a "proof of concept" by Group Health Cooperative. Intervention: "a whole-practice transformation," including increased staffing and new care management practices.	Estimated impact on total spending: decrease of 2% (marginally statistically significant).
Paustian et al. (2014)	Setting: primary care practices in Michigan. Intervention: practices increased PCMH capabilities from 2009 to 2010.	Estimated impact on total spending (full vs. no capabilities): decrease of 8 percent for adults (marginally statistically significant), and an increase of 8% for children (marginally statistically significant).
Oncology "pathways"		
Kreys et al. (2013)	Setting: oncologists volunteering to participate in a pathways program run by CareFirst BlueCross BlueShield of Maryland. Intervention: participating practices received increased fees based on compliance.	Estimated impacts: savings of around $6,500 per patient treated, calculated by comparing actual spending against projected trend.
Hoverman et al. (2011)	Setting: oncologists in the US Oncology network, or treating patients in the MedStat MarketScan database. Retrospective observational study (no intervention).	Estimated difference in spending ("on-" versus "off-pathway"): total spending per patient per month was 35% lower for adjuvant treatment and 30% lower for therapy for metastatic disease
Neubauer et al. (2010)	Setting: 8 oncology practices treating non-small-cell lung cancer. Retrospective observational study (no intervention).	Estimated difference in spending: "on-pathway" patients had total spending 35% lower than "off-pathway" patients.
Episode-based payments for oncology		

Study Types	Study Description	Study Results
Newcomer et al. (2014)	Setting: 5 oncology groups volunteering to participate in a 3-year pilot operated by UnitedHealthcare. Intervention: Episode-based payments for case management, predefined chemotherapy regimens, and chemotherapy drugs paid at average sales price (ASP): ASP+0%.	Estimated differences in spending (patients treated in the pilot versus predicted FFS costs): reduction of 34% in overall costs, but increase of 179% in chemotherapy drug costs.
Shared savings		
Vats et al. (2013)	Setting: 3 volunteer primary care practices. Intervention: Stipends for practice reform, switch from FFS to bundled payments for episodes of care with performance bonuses.	Estimated impact on total spending: decreases of 6–8%, depending on the year (not statistically significant).
L&M Policy Research (2013)	Setting: 32 Medicare Pioneer accountable care organizations (ACOs). Intervention: ACOs became eligible for performance-based shared savings.	Estimated impact on total spending: decrease of 2%.

SOURCE: Authors' analysis.

In general, the evidence from the literature offers only limited guidance as to what to expect under the proposed payment model. Based on the results from the PCMH interventions, it seems unlikely that offering care management fees for qualifying practices will, by itself, produce significant savings. But the results of the pathways interventions suggest that there are some opportunities for practices to choose lower-cost chemotherapy regimens. And the results from the shared savings interventions suggest that financial incentives, which are similar to the performance-based payments in the proposed oncology model, can result in savings, albeit modest. Two factors in the proposed model strengthen participating practices' financial incentives to achieve savings:

1. the opportunity to earn performance-based payments, which can be sizeable relative to the practice's overall revenues; and
2. the potential loss of care management payments if a practice's participation is terminated by CMS because it fails to demonstrate savings.

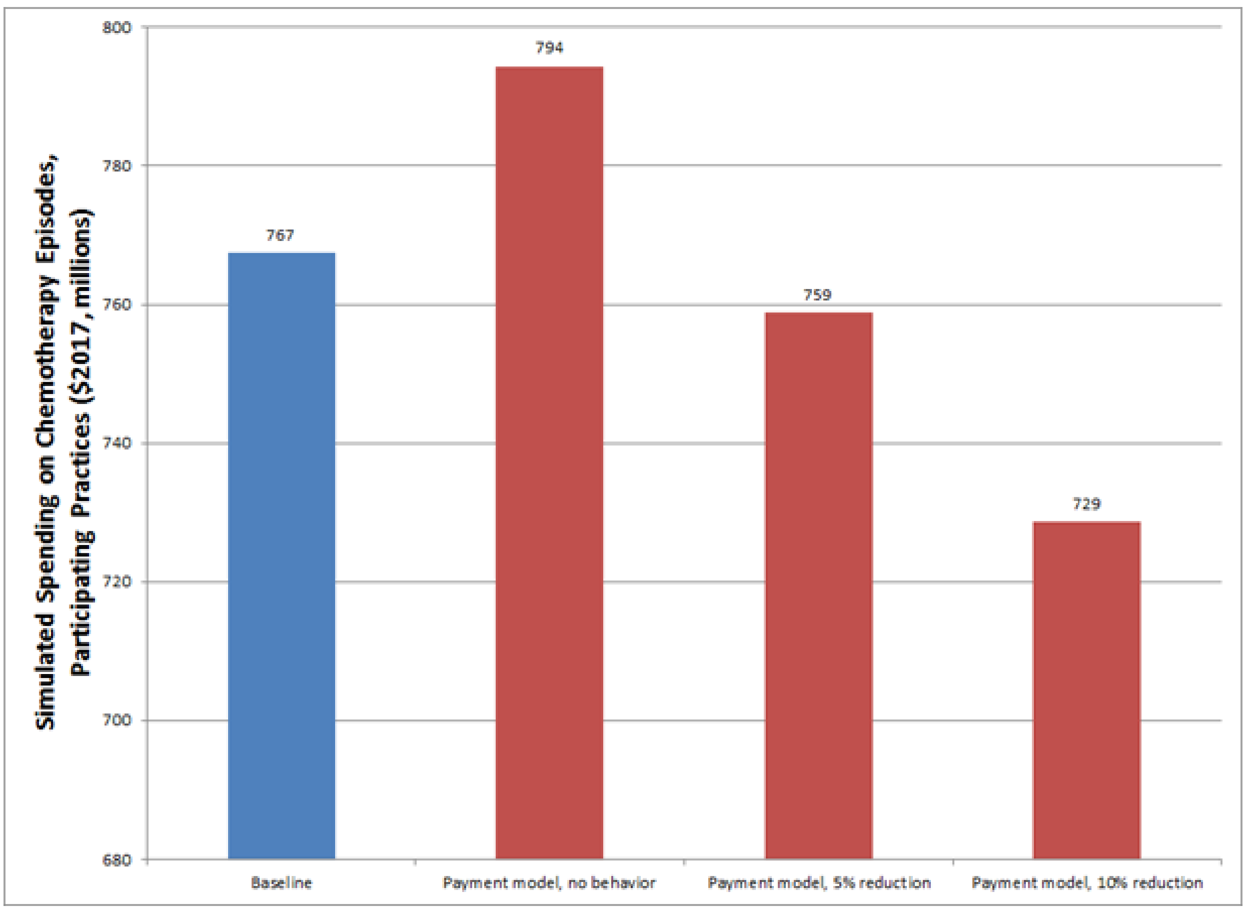
But, at the same time, the care management payments in the proposed model are large enough that they may encourage participation by practices that do not intend to attempt to earn performance-based payments.

Because the evidence is limited, for all of our simulated spending results we present three sets of simulated outcomes using three different assumed behavioral responses: "no behavior," meaning no change in the quantity and intensity of services provided, and "5-percent reduction" and "10-percent reduction," meaning that the quantity and intensity of all services is reduced proportionally across all service categories by those amounts. Ideally, the simulation model would incorporate service category–specific behavioral responses, but the evidence was inadequate to justify that level of granularity.

5. Simulated Spending Impacts

In Figure 5.1, we present the simulated total Medicare spending on episodes of chemotherapy provided by participating practices. Three scenarios are reported: “no behavior” (i.e., oncology practices do not change the quantity or intensity of services provided in response to participating in the model), “5-percent reduction” (i.e., oncology practices reduce the quantity and/or intensity of services in a way that results in 5-percent lower Medicare FFS spending for the episode), and “10-percent reduction.” Under the no-behavior scenario, we estimate that total spending is increased by 4 percent, which reflects the fact that all participating oncology practices are receiving care management payments and some are receiving “noise bonuses,” as described previously. Of the 4-percent increase in total spending in the no-behavior scenario, 90 percent is due to care management payments and the other 10 percent is due to noise (i.e., unearned) bonuses. In the 5-percent-reduction scenario, total spending is slightly below the baseline, indicating that the program has broken even. Total spending in the 10-percent-reduction scenario is about 5 percent lower than baseline spending.

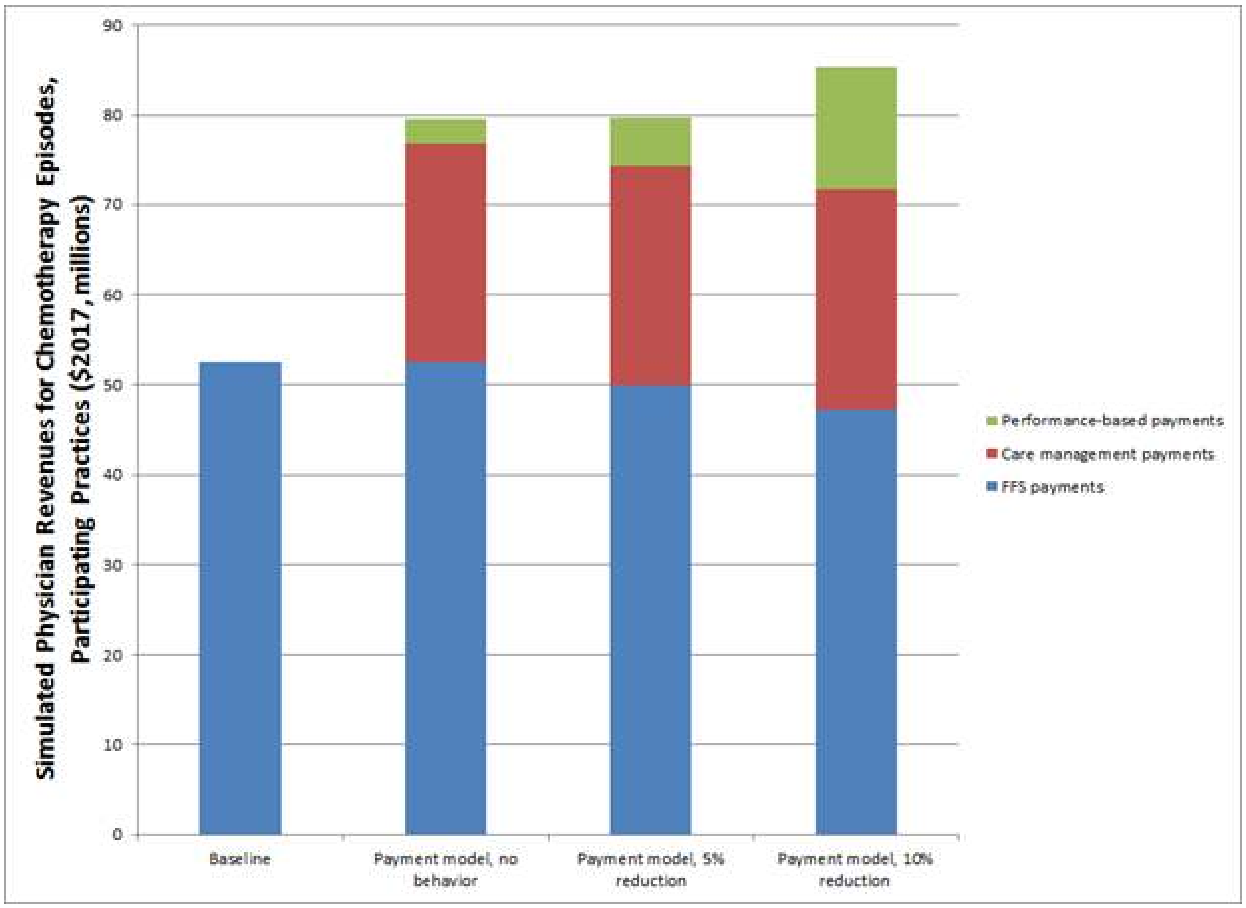
Figure 5.1. Medicare Spending on Chemotherapy Episodes Provided by Participating Practices



SOURCE: Authors’ analysis.

Although the simulated impacts of the payment model on total spending are fairly modest—ranging from +4 percent to -5 percent—the payment model would have a different and much more significant impact on the revenues of participating physician practices. The larger relative impact on physician practice revenues reflects the fact that physician revenues only represent around 7 percent of total spending on chemotherapy episodes. As shown in Figure 5.2, total revenues to physician practices increase substantially under all three behavioral scenarios, due both to the care management payments and the performance-based payments. Consistent with the goals of the proposed payment model, the total revenues to the practice increase with larger behavioral responses, which indicates that the increase in performance-based payments more than offset the loss in FFS revenues to the practice.

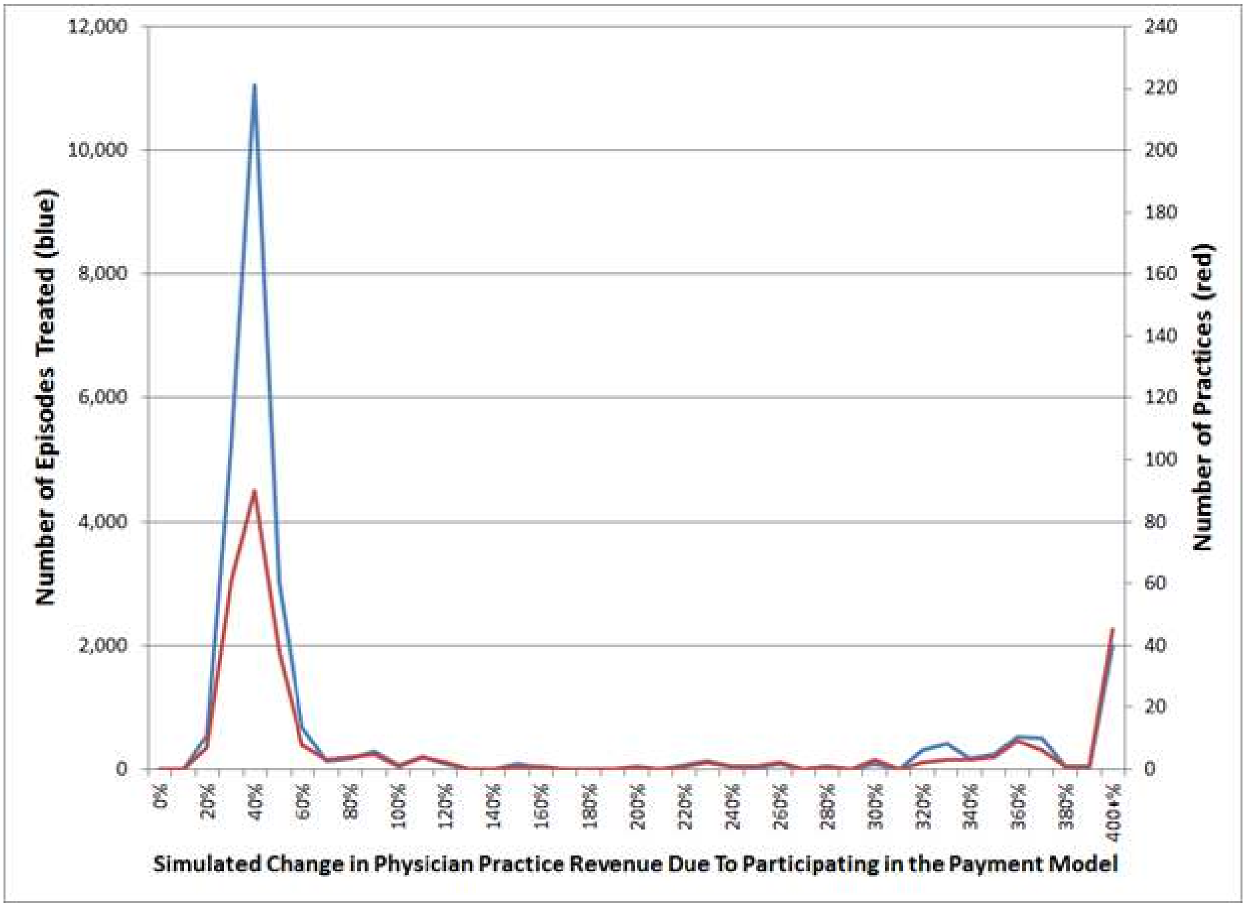
Figure 5.2. Simulated Medicare Revenues at Participating Practices



SOURCE: Authors' analysis.

Overall, simulated revenues to participating physician practices would increase by over 50 percent, from around $2,000 to $3,000, but that aggregate figure masks a very wide distribution among practices.[2] As shown in Figure 5.3, there is a large concentration of practices that would experience an increase in revenues between 20 percent and 60 percent. But there is also a very long right tail, meaning that for a few practices, their revenues would double, triple, or increase even more. These very large changes in revenues occur partly because some practices have very low revenues in the baseline and some practices receive very large simulated performance payments.

Figure 5.3. Distribution of Simulated Changes in Practice-Level Revenues



SOURCE: Authors' analysis.

---

[2] Including all practices, mean projected physician practice revenues per episode are $1,776 in 2016. Mean projected physician practice revenues per episode are higher—$1,989—if the analysis is limited to practices providing 50 or more episodes. The difference between those means reflects the fact that low-volume practices tend to receive lower revenues per episode than high-volume practices.

6. Alternative Model Designs and Participation Assumptions

To explore the implications of alternative model designs, we ran the simulation model using four alternative designs:

1. a minimum practice volume of 100 episodes (rather than the assumed effective minimum of 50 episodes);
2. a performance-based payment share of 50 percent (rather than 100 percent);
3. a 10-percent ceiling on the performance-based payment (rather than 20 percent); and
4. a monthly care management payment of $80 (rather than $160).

Increasing the minimum practice volume reduces the potential for payment of “noise bonuses,” while the second and third alternatives would reduce both “noise bonuses” and earned performance payments (meaning payments due to reductions in utilization and/or intensity). For each of those alternatives, total spending and physician practice revenues were simulated using three behavioral responses: no behavior, 5-percent reduction, and 10-percent reduction.

As shown in Table 6.1, each of the alternative designs reduces the “break-even” point, meaning that a smaller behavioral response would be required in order for total spending to be equal to or less than under baseline. Reducing the care management payment to $80 has the largest impact (the break-even point drops from 3.8-percent savings to 2.2-percent savings), while increasing the minimum practice volume to 100 has the second-largest impact (the break-even point drops from 3.8-percent savings to 3.4-percent savings).

Table 6.1. Alternative Model Designs

Minimum Practice Volume (number of episodes)	Performance-Based Payment Share	Ceiling on Performance-Based Payment	Care Management Payment per Month	Break-Even Point[3]	Behavioral Response	Simulated Impact on Total Spending	Simulated Impact on Physician Practice Revenues
50	90%	20%	$160	3.8%	No behavior	3.5%	51.3%
	(base case)				5% reduction	–1.1%	51.6%
					10% reduction	–5.1%	62.1%
100	90%	20%	$160	3.4%	No behavior	3.3%	47.5%
					5% reduction	–1.5%	46.0%
					10% reduction	–5.6%	54.4%
50	45%	20%	$160	3.5%	No behavior	3.3%	48.8%
					5% reduction	–1.5%	46.5%
					10% reduction	–5.9%	49.3%
50	90%	10%	$160	3.7%	No behavior	3.5%	51.0%
					5% reduction	–1.2%	50.7%
					10% reduction	–5.3%	59.1%

[3] The break-even point is the size of the behavioral response that corresponds to the payment model having no impact on total spending, where the behavioral response is an across-the-board reduction in utilization.

Minimum Practice Volume (number of episodes)	Performance-Based Payment Share	Ceiling on Performance-Based Payment	Care Management Payment per Month	Break-Even Point[3]	Behavioral Response	Simulated Impact on Total Spending	Simulated Impact on Physician Practice Revenues
50	90%	20%	$80	2.2%	No behavior	2.0%	29.6%
					5% reduction	–2.5%	31.7%
					10% reduction	–5.9%	50.6%

SOURCE: Authors' analysis.

Selective Participation

In the base case, the simulation model assumes random practice participation, meaning that practices have an equal probability of participating regardless of their spending target. Physician practices with relatively high benchmarks may have a higher probability of participating, if they can anticipate that those benchmarks will be relatively high. In practice, it is unlikely that practices will have access to the data and the analytical resources required to predict their benchmarks before they choose to participate. Practices may, however, be able to anticipate that benchmarks will be relatively high if the inflation factors that are applied to historical spending are excessive or if benchmarks incorporate region-level average spending patterns that differ predictably from the practice's own patterns.

To illustrate the sensitivity of the findings to selective participation, the simulation model was run applying a higher probability of participation to practices with benchmarks that are high relative to their actual spending (Table 6.2). In this sensitivity test, practices whose benchmarks are overstated by 10 percent are around twice as likely to participate as practices whose benchmarks are understated by 10 percent. The selective participation assumption increases the break-even point, as expected, but the magnitude of the change is relatively small. This indicates that the takeaways from the simulation modeling are not an artifact of the random participation assumption.

Table 6.2. Selective Participation

Participation	Break-Even Point[4]	Behavioral Response	Simulated Impact on Total Spending	Simulated Impact on Physician Practice Revenues
Random	3.8%	No behavior	3.5%	51.3%
(base case)		5% reduction	–1.1%	51.6%

[4] The break-even point is the size of the behavioral response that corresponds to the payment model having no impact on total spending, where the behavioral response is an across-the-board reduction in utilization.

Participation	Break-Even Point[4]	Behavioral Response	Simulated Impact on Total Spending	Simulated Impact on Physician Practice Revenues
		10% reduction	–5.1%	62.1%
Selective (practices with higher benchmarks more likely to participate)	4.1%	No behavior	3.7%	54.0%
		5% reduction	–0.8%	56.4%
		10% reduction	–4.5%	70.2%

SOURCE: Authors’ analysis.

7. Conclusion

The simulation described in this report offers several insights into the proposed payment model for oncology.

1. The proposed care management payments are large relative to oncology practices' FFS revenues. The proposed care management payments will equal $960 in 2016 for a full six-month episode, compared to average estimated physician practice revenues per episode in that year of around $2,000 and total spending per episode of around $27,000. By themselves, therefore, the care management payments would increase physician practices' Medicare revenues by roughly 50 percent on average and would increase total Medicare spending by 4 percent.
2. The use of performance-based payments requires the comparison of spending benchmarks and actual spending; both of those spending amounts are subject to sampling error, and benchmarks may also be subject to errors in forecasted trends and other biases. Those sources of error can be lessened in a number of ways by, for example, applying sophisticated case-mix adjustment, by using multiple years of historical data, or by blending a practice's historical costs with regional or national episode cost data. But the counterfactual—i.e., a participating practice's spending per episode if it had not participated—is fundamentally unobservable. Inaccuracies in the benchmarks, combined with one-sided performance payments, mean that the Medicare program will in some cases pay "noise bonuses"—performance-based payments that are due to inaccurately low benchmarks rather than to providers' behavioral responses. The likelihood of noise bonuses would be reduced if the payment model were limited to larger practices.
3. For the Medicare program to break even, participating oncology practices would have to reduce utilization and intensity by roughly 4 percent. "Breaking even" means that Medicare spending per episode is the same with the payment model, including the care management payments and performance-based payments, as it is without the payment model. Based on research on other similar payment models, a behavioral response of that magnitude may be possible, but is by no means certain.
4. The break-even point can be reduced—i.e., less of a behavioral response would be required for the Medicare program to achieve savings—if the care management payments were reduced or if the performance-based payments were reduced.

The introduction of a new payment model for oncology may have several possible unintended consequences that are not included in the simulation model but are potentially important. One possibility is that oncologists participating in the payment model might increase the number of episodes of treatment they initiate due to of the availability of sizeable care management payments and performance-based payments. That type of "demand inducement" response to changes in payments for oncology has been demonstrated in other contexts (Jacobson et al., 2010; Colla et al., 2012; Elliot et al., 2010). Another possible response is "cherry picking," meaning that practices seek to treat patients who are relatively healthy and whose course of treatment is less intensive than is typical in order to receive performance-based payments. The use of case-mix adjustment in the calculation of performance-based payments can mitigate the incentive to engage in cherry picking. Another possible response is stinting, meaning switching patients to less expensive, but inappropriate, treatment regimens, or recommending against clinically beneficial services, in order to receive performance-based payments. CMS has proposed tracking quality metrics as part of the payment model. Conditioning performance-based payments on adequate performance on those metrics can mitigate the risk of stinting.

References

Bach, P.B., J. N. Mirkin, and J. J. Luke, "Episode-Based Payment for Cancer Care: A Proposed Pilot for Medicare," *Health Affairs*, Vol. 30, No. 3, 2011, pp. 500–509.

Centers for Medicare & Medicaid Services, “Accuracy Analysis of the Short-Term (11-Year) National Health Expenditure Projections,” Office of the Actuary, 2011. As of October 7, 2014: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/ProjectionAccuracy.pdf

Colla, Carrie H., Nancy E. Morden, Jonathan S. Skinner, J. Russell Hoverman, and Ellen Meara, “Impact of Payment Reform on Chemotherapy at the End of Life,” *American Journal of Managed Care,* Vol. 18, No. 5, May 2012, pp. e200–e208.

Congressional Budget Office, *Estimating the Costs of One-Sided Bets: How CBO Analyzes Proposals with Asymmetric Uncertainties*, Washington, D.C., October 1999. As of October 7, 2014: http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/15xx/doc1589/onesided.pdf

Congressional Budget Office, “Lessons from Medicare’s Demonstration Projects on Disease Management, Care Coordination, and Value-Based Payment,” Issue Brief, Washington, D.C., January 2012. As of October 7, 2014: http://cbo.gov/sites/default/files/cbofiles/attachments/01-18-12-MedicareDemoBrief.pdf

Elliott, S. P., S. L. Jarosek, T. J. Wilt, and B. A. Virnig, “Reduction in Physician Reimbursement and Use of Hormone Therapy in Prostate Cancer,” *Journal of the National Cancer Institute,* Vol. 102, No. 24, 2010, pp. 1826–1834.

Fifield, Judith, Deborah Dauser Forrest, Joseph A. Burleson, Melanie Martin-Peele, and William Gillespie, “Quality and Efficiency in Small Practices Transitioning to Patient Centered Medical Homes: A Randomized Trial,” *Journal of General Internal Medicine,* Vol. 28, No. 6, June 2013, pp. 778–786.

Friedberg, Mark W., Eric C. Schneider, Meredith B. Rosenthal, Kevin G. Volpp, and Rachel M. Werner, “Association Between Participation in a Multipayer Medical Home Intervention and Changes in Quality, Utilization, and Costs of Care,” *Journal of the American Medical Association,* Vol. 311, No. 8, 2014, p. 815.

Hoverman, J. Russell, Thomas H. Cartwright, Debra A. Patt, Janet L. Espirito, Matthew P. Clayton, Jody S. Garey, Terrance J. Kopp, Michael Kolodziej, Marcus A. Neubauer, Kathryn Fitch, Bruce Pyenson, and Roy A. Beveridge, “Pathways, Outcomes, and Costs in Colon Cancer: Retrospective Evaluations in Two Distinct Databases,” *Journal of Oncology Practice,* Vol. 7, No. 3S, 2011, pp. 52s–59s.

Huckfeldt, Peter J., Chris Chan, Samuel Hirshman, Aaron Kofner, Jodi L. Liu, Andrew W. Mulcahy, Ioana Popescu, Clare Stevens, Justin W. Timbie, and Peter S. Hussey, *Specialty Payment Model Opportunities and Assessment: Oncology Model Design Report*, Santa Monica, Calif.: RAND Corporation, RR-763-CMMS, 2014, forthcoming.

Jacobson, M., C. C. Earle, M. Price, and J. P. Newhouse, “How Medicare’s Payment Cuts for Cancer Chemotherapy Drugs Changed Patterns of Treatment,” *Health Affairs,* Vol. 29, No. 7, 2010, pp.

1391–1399.

Kreys, E. D., and J. M. Koeller, "Documenting the Benefits and Cost Savings of a Large Multistate Cancer Pathway Program from a Payer's Perspective," *Journal of Oncology Practice,* Vol. 9, No. 5, 2013, pp. e241–e247.

L&M Policy Research, *Evaluation of CMMI Accountable Care Organization Initiatives*, Washington, D.C., November 3, 2013. As of October 7, 2014: http://innovation.cms.gov/Files/reports/PioneerACOEvalReport1.pdf

Maeng, Daniel D., Jove Graham, Thomas R. Graf, Joshua N. Liberman, Nicholas B. Dermes, Janet Tomcavage, Duane E. Davis, Frederick J. Bloom Jr., and Glenn D. Steele Jr., "Reducing Long-Term Cost by Transforming Primary Care: Evidence from Geisinger's Medical Home Model," *American Journal of Managed Care,* Vol. 18, No. 3, 2012, pp. 149–155.

McClellan, M., Kavita Patel, John O'Shea, Jeffrey Nadel, Andrea Thoumi, and Judith Tobin, *Specialty Payment Model Opportunities Assessment and Design—Environmental Scan for Oncology (Task 2): Final Version*, Washington, D.C.: The Brookings Institution, 2013. As of October 7, 2014: http://www2.mitre.org/public/payment_models/Brookings_Oncology_TEP_Environ_Scan.pdf

Neubauer, M. A., J. R. Hoverman, M. Kolodziej, L. Reisman, S. K. Gruschkus, S. Hoang, A. A. Alva, M. McArthur, M. Forsyth, T. Rothermel, and R. A. Beveridge, "Cost Effectiveness of Evidence-Based Treatment Guidelines for the Treatment of Non-Small-Cell Lung Cancer in the Community Setting," *Journal of Oncology Practice,* Vol. 6, No. 1, 2010, pp. 12–18.

Newcomer, Lee N., Bruce Gould, Ray D. Page, Sheila A. Donelan, and Monica Perkins, "Changing Physician Incentives for Affordable, Quality Cancer Care: Results of an Episode Payment Model," *Journal of Oncology Practice*, published online ahead of print on July 8, 2014.

Paustian, Michael L., Jeffrey A. Alexander, Darline K. El Reda, Chris G. Wise, Lee A. Green, and Michael D. Fetters, "Partial and Incremental PCMH Practice Transformation: Implications for Quality and Costs," *Health Services Research,* Vol. 49, No. 1, 2014, pp. 52–74.

Reid, R. J., K. Coleman, E. A. Johnson, P. A. Fishman, C. Hsu, M. P. Soman, C. E. Trescott, M. Erikson, and E. B. Larson, "The Group Health Medical Home at Year Two: Cost Savings, Higher Patient Satisfaction, and Less Burnout for Providers," *Health Affairs,* Vol. 29, No. 5, 2010, pp. 835–843.

Vats, Sonal, Arlene S. Ash, and Randall P. Ellis, "Bending the Cost Curve? Results from a Comprehensive Primary Care Payment Pilot," *Medical Care,* Vol. 51, No. 11, November 2013, pp. 964–969.

Werner, Rachel M., Mark Duggan, Katia Duey, Jingsan Zhu, and Elizabeth A. Stuart, "The Patient-Centered Medical Home," *Medical Care,* Vol. 51, No. 6, 2013, pp. 487–493.

Appendix

Inflating Spending from 2010 to 2016

To inflate spending, we applied the following formula:

$$S^{base}_{i,j,t1} = \sum_c \left(s_{i,j,c,t0} \frac{i_{c,t1}}{i_{c,t0}} \right) = \sum_c s_{i,j,c,t1} \qquad \text{(eq. 2)}$$

where

i indexes oncology episodes

j indexes practices (episode *i* is attributed to practice *j*)

c indexes service categories (e.g., hospital, chemotherapy drugs, etc.)

t0 represents a base year (2010)

t1 represents a simulation year (2016)

$S^{base}_{i,j,t1}$ is spending on episode *i* treated by practice *j* in year *t1* under baseline (i.e., in the absence of the payment model demonstration)

$s_{i,j,c,t0}$ is actual spending on episode *i* for service category *c* treated by practice *j* in base year *t0*

$\frac{i_{c,t1}}{i_{c,t0}}$ represents the inflation in spending per episode from *t0* to *t1*, due to increases in unit prices

$s_{i,j,c,t1}$ is projected spending on episode *i* for service category *c* treated by practice *j* in simulation year *t1*.

The inflation factors, which are shown in Table 2.1, are calculated using commodity-level producer price indexes from the Bureau of Labor Statistics for chemotherapy ("Pharmaceuticals affecting neoplasms, the endocrine system, & metabolic diseases"), physician care, hospital inpatient care, hospital outpatient care, and home health and hospice care. Observed price growth from 2010 through 2014 is extrapolated through 2016.

Simulated Spending Amount

For each episode, we simulate a spending amount in 2016 equal to

$$s^{model}_{i,j} = (1 - partic_j) s^{base}_{i,j} + partic_j \left[s^{base}_{i,j} (1 + \Delta h) + mgmt_{i,j} + perf_{i,j} \right] \qquad \text{(eq. 3)}$$

where

i indexes oncology episodes

j indexes practices

$partic_j$ equals 1 if practice *j* participates in the model and 0 if not

$s_{i,j}^{base}$ equals baseline spending in 2016 (i.e., assuming no payment model)

Δh equals the behavioral response, equal to the percentage change in health spending as a result of participation in the payment model

$mgmt_{i,j}$ equals the care management payment paid to practice *j* for episode *i*

$perf_{i,j}$ equals the performance-based payment, if any, to practice *j* for episode *i*.

Calculating the simulated spending amount requires simulating the participation decision ($partic_j$), the behavioral response (Δh), the care management payment ($mgmt_{i,j}$), and the performance-based payment ($perf_{i,j}$).

Simulated Practice-Level Spending Targets

To simulate a spending target for each practice, we will start with the practice's baseline spending on all episodes in the projection year and apply a noise factor. The noise factor is designed to simulate the random variation in health care utilization that will get incorporated into each practice's historical cost factor. To estimate the noise factor, we will take historical episodes provided by each practice and randomly split them into two groups based on the last digit of their beneficiary ID. By assumption, the practice applied the same practice patterns to these historical episodes, so any spending variation that we observe between the two groups of episodes (above and beyond what can be explained by case mix) reflects measurement error.

The noise factor will equal

$$noise_j = \frac{praccost_j^{odd}}{praccost_j^{even}} shrinkage \quad \text{(eq. 4)}$$

and

$$targ_{j,t1}^{sim} = epicost_j noise_j \quad \text{(eq. 5)}$$

where

$$praccost_j^{odd} = \left(\frac{\sum_{i=m_j}^{n_j} \left[1(i \text{ ends in an odd \#}) tot_i^{trunc} \right]}{\sum_{i=m_j}^{n_j} \left[1(i \text{ ends in an odd \#}) tothat_i^{trunc} \right]} \right)$$

(eq. 6)

$$praccost_j^{even} = \left(\frac{\sum_{i=m_j}^{n_j} \left[1(i \text{ ends in an even \#}) tot_i^{trunc} \right]}{\sum_{i=m_j}^{n_j} \left[1(i \text{ ends in an even \#}) tothat_i^{trunc} \right]} \right)$$

(eq. 7)

and where

m_j is the first episode treated by practice *j*

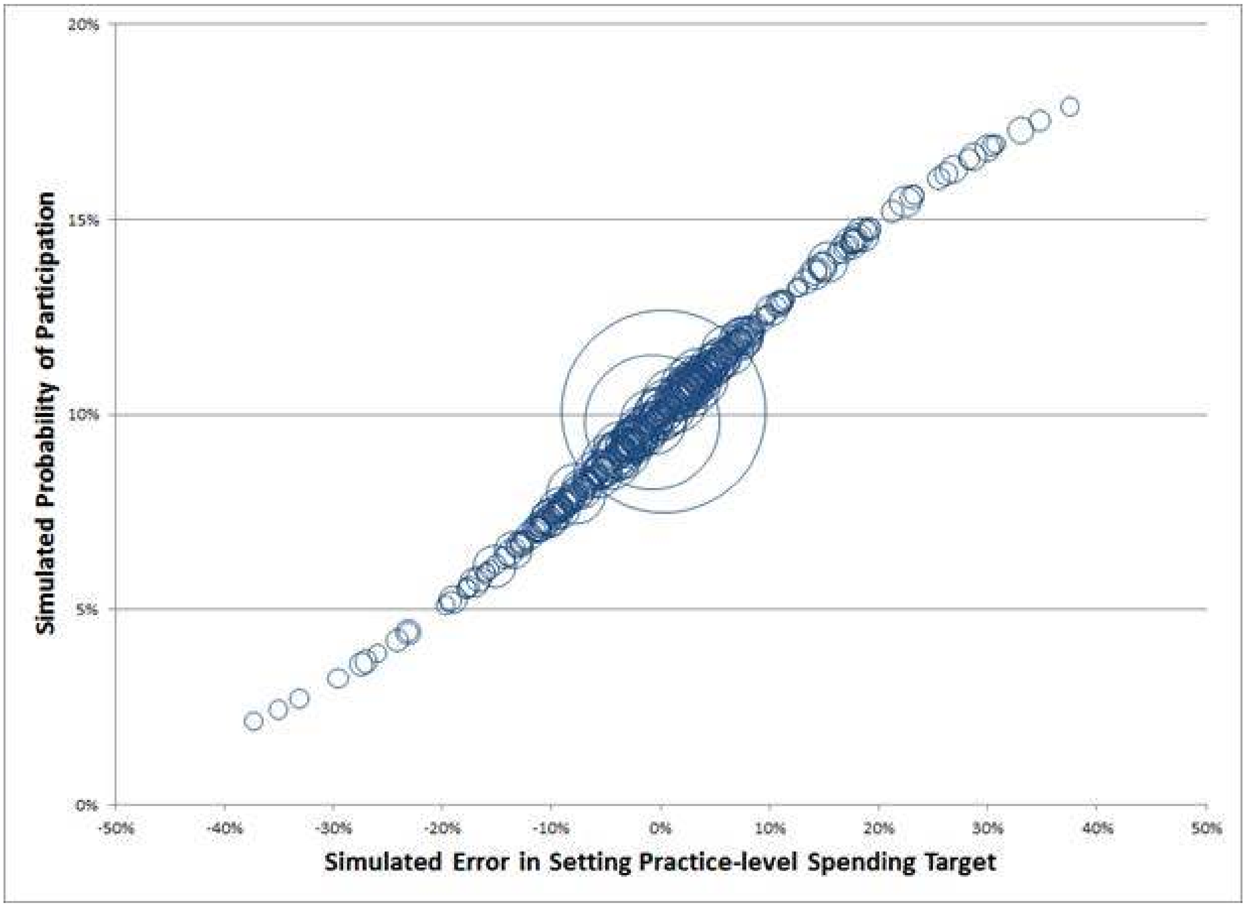
n_j is the last episode treated by practice *j*

and where tot_i^{trunc} is spending on episode *i* with truncation applied at the 95th percentile, and $tothat_i^{trunc}$ is the predicted truncated spending on episode *i* from a case-mix model. A uniform correction factor is applied to the predicted truncated spending for all episodes so that mean predicted spending exactly equals the mean actual spending. The case-mix model includes the patient's age, state of residence, dual status, and type of cancer. The *shrinkage* term is set to 0.64 and is included to account for the fact that splitting the historical episodes into two groups halves the number of episodes. To calculate the shrinkage term, we calculated the absolute value of $\left(\frac{praccost_j^{odd}}{praccost_j^{even}} - 1 \right)$ and then fit a power trendline. The value of 0.64 equals 2 raised to the power of the estimated coefficient from the power trendline, which is an estimate of the proportional difference in the absolute forecast error if the number of episodes were twice as large.

Self-Selected Participation

In the self-selected participation scenario, practices were assigned a probability of participating that varied depending on the practice's simulated forecast error. Practices in the noise bonus range were assigned participation probabilities between 12 and 18 percent, and practices in the range in which they might not receive earned bonuses were assigned participation probabilities from 2 percent to 8 percent. (See Figure A.1.)

Figure A.1. Simulated Participation Probability



NOTES: Each bubble represents a physician practice, and the size of the bubble represents the number of chemotherapy episodes provided by the practice.