

An Early Look at the Effects of the Medicare Bundled Payments for Care Improvement: Evidence from Presence Health

Nathan Petek, Will Clark, Jonas A. de Souza, Heidi Graham, Svati Patel, Xikuan Wang, David DiLoreto, and Daniel Adelman*

May 2018

Abstract

A recent Medicare bundled payment pilot program gives a single health care provider a financial incentive to keep total payments to all providers during a multi-week episode of patient care under a certain value. We estimate the effect of the pilot for knee and hip replacements using data from 11 Presence Health hospitals. Unconditional on patient characteristics, the program caused a 12 percent decrease in spending. Controls for patient characteristics reduce the effect by half, which is consistent with selection of healthier patients. Direct estimates of the selection effect are typically not statistically significant. The rest of the spending decrease is driven by reduced utilization of inpatient rehabilitation and shifting patients from inpatient rehabilitation to skilled nursing facilities. Spending decreases are concentrated among relatively healthy patients while spending on patients with major complications or comorbidities increases.

Keywords: bundled payments, health care spending, Medicare

JEL Codes: C23, H42, I13

I. Introduction

It is widely believed that US health care spending is inefficiently high (Garber and Skinner 2008, Berwick and Hackbarth 2012). Alternative payment schemes like bundled payments have the potential to increase the efficiency of the health care system by giving providers an incentive to reduce health care utilization and coordinate care across settings (CMS 2015a). Cutler (2014) has argued in favor of a “rapid transition from the existing fee-for-service payment model to bundled

* Petek, Clark, de Souza, Graham, Patel, Wang, and Adelman: University of Chicago; and DiLoreto: Presence Health. Contact: npetek@frc.gov. We are grateful to Matthew Gentzkow, Neale Mahoney, and Devin Pope for helpful comments and to Jessica Cobb for research assistance. This work began as a project in the Healthcare Analytics Laboratory course at the University of Chicago, Booth School of Business and was funded internally by Presence Health and the University of Chicago, Booth School of Business.

payment for episodes of care” because it could reduce the amount of low value health care spending by billions per year.

The Center for Medicare & Medicaid Services’ (CMS) is testing one such program that bundles Medicare expenditures during a patient’s hospital stay at voluntarily participating hospitals with most spending on inpatient and outpatient care during the next 30- to 90-days. If the total spending during the entire time period falls below a hospital-diagnosis specific target price, the hospital where the episode began receives a dollar-for-dollar bonus, and if the total spending exceeds the target price, the hospital pays a dollar-for-dollar penalty (CMS 2015a).

This program gives hospitals a strong incentive to cut spending. However, Medicare spending for an inpatient stay is essentially determined by the patient’s diagnosis, so any decrease in spending must be driven by reduced utilization of post-acute care. Unless the post-acute care providers are owned by the hospital, they have no direct financial incentive to reduce Medicare expenditures. As a result, there are few margins on which the hospital could induce spending reductions and the care at those margins could be valuable. There is also the potential for unintended consequences because one margin at which the hospital could adjust is by selecting relatively healthier patients to treat.

In this paper, we use a difference-in-differences design to estimate the effect of implementing bundled payments for knee and hip replacements at 3 of the 11 hospitals in the Presence Health Network. These surgical procedures are of particular interest because they cost an estimated \$9 billion per year and for a time CMS made bundled payments mandatory for them at about 800 hospitals in 67 metropolitan areas (CMS 2015a, CMS 2015b, Dobson et al. 2012).

We first verify that, consistent with other work on bundled payments, spending fell through reduced utilization of post-acute care (Dummit et al. 2016). Our main contribution is showing the mechanisms through which bundled payments cause physicians and hospitals to reduce the quantity of post-acute care. We check whether the hospitals cut spending by selecting relatively healthier patients to treat, and other margins where a hospital could reduce spending. Understanding how hospitals and physicians cut spending in response to the incentives created by the bundled payment program informs us about how hospitals and physicians respond to financial incentives, where the program is likely to be successful, and whether it is likely to harm patients.

The appropriate specification to use when estimating the effect of bundled payments on spending depends on whether or not physicians select patients they anticipate being lower cost using observable patient characteristics. If physicians can select healthier patients, then including patient characteristics as controls would result in us underestimating the effect of the program on the cost per patient. We find that unconditional on patient characteristics, implementing bundled payments caused a statistically significant 12 percent decrease in the total episode cost, while the effect falls to a marginally significant 6 percent after controlling for patient characteristics.

The substantial decrease in the estimate after conditioning on patient characteristics suggests the magnitude of the selection effect could be large. The sign of the estimated effect of bundled payments on a predicted cost index and four correlates of patient health is consistent with selection of healthier patients. However, only the effect on the Charlson Comorbidity Index is statistically significant and it is driven by the first two quarters after bundled payments are put into place.

The 6 percent decrease in spending after controlling for patient observables is driven by a statistically significant, 18 percentage point decrease in the number of patients who ever use inpatient rehabilitation that is partially offset by an 11 percentage point increase in skilled nursing facility (SNF) use. These results suggest that physician's respond to bundled payments by substituting less intensive for more intensive post-acute care while also increasing discharges to home. Finally, we see significant shifts in patients from non-Presence to Presence-owned skilled nursing homes and home health agencies which would tend to increase Presence revenue and give them more control over the length of post-acute care treatment conditional on admission.

Although bundled payments reduced costs, they could be counterproductive if they result in worse outcomes for patients. We find that the decreases in spending are entirely driven by relatively healthy patients who do not have major complications or co-morbidities, and who we expect ex-ante would be less likely to be harmed by reduced utilization of intensive post-acute care. Spending on patients with major complications and comorbidities actually increases significantly. We find no evidence that the reductions in intensive post-acute care lead to significantly increased readmissions, but our confidence intervals can only reject relatively large increases in the readmission rate.

These results suggest that the participating hospitals and physicians respond to the incentives created by bundled payments on the two primary margins where they have some ability to affect episode costs: the set of patients they treat and where those patients are discharged after their inpatient stay. Operators of non-Presence Health affiliated post-acute care facilities do not have an incentive to reduce spending by cutting the length of stay, and they do not. The choices made by Presence health physicians will tend increase participating physician and Presence Health revenue, but are targeted to limit the potential harm to patients by only cutting spending for relatively healthy patients and by sending a substantial fraction of patients formerly discharged to inpatient rehab to SNFs rather than directly to home. These results suggest bundled payments are most likely to reduce costs in an environment with relatively healthy patients, high utilization of intensive post-acute care, and where it is easier to substitute more intensive for less intensive post-acute care.

Our paper is most closely related to Dummit et al. (2016), a large-scale study of Medicare's bundled payment program for knee and hip replacement patients that found an approximately 3.8 percent decrease in episode costs, conditional on patient observables, that was driven by reduced discharges to post-acute care facilities. They find some evidence that of selection of healthier patients, but do not show the effect of selection on episode costs. There are also a number of small scale studies that, in most cases, simply compared mean costs and admissions pre- and post-implementation with results that are qualitatively similar to Dummit et al. (2016).¹

Our results suggest the effect of bundled payments on medical spending may actually be larger than the Dummit et al. (2016) estimate, which includes controls for an extensive set of patient observables, because there is suggestive evidence that bundled payments cause physicians' to select patients likely to be lower cost ex-ante. Although this source of cost reductions could make patients worse off by reducing access to care for relatively unhealthy patients. Our results also provide some guidance on which environments the program is likely to be effective, and support the Dummit et al. (2016) finding of no effect on health outcomes by showing that spending on relatively unhealthy patients was not affected and there was substitution within the post-acute care category.

¹ See Siddiqi et al. (2017) for a review.

Earlier bundled payment pilot programs had a very different incentive structures that either included little to no post-acute care beyond readmissions in some cases (Cromwell et al. 1997, Urdapilleta et al. 2013, and Alexander 2017) or only included post-acute care within a hospital system (Berry et al. 2009). These programs did create an incentive to select healthier patients and in recent work Alexander (2017) finds substantial evidence of patient selection in a bundled payment pilot in New Jersey.

Our paper relates to studies of other alternative payment models like accountable care organizations (ACOs) that bundle payments at the level of the patient rather than the level of an episode of care like Medicare's Pioneer ACO pilot which resulted in small cost decreases (McWilliams et al. 2015 and Nyweide et al. 2015).

II. Data

A. Overview

Data on procedures with DRG codes 469 and 470 performed on Medicare fee-for-service beneficiaries at Presence Health hospitals from January 1, 2013 to February 27, 2015 were used. We focus on complete 90-day episodes of care by limiting the sample to episodes that start before November 26, 2014. Three Presence Health hospitals participated in the Bundled Payments for Care Improvement (BPCI) Model 2 Major Joint Replacement of the lower extremity: Presence Resurrection Medical Center, Presence Saints Mary and Elizabeth Medical Center, and Presence Saint Joseph Hospital. Eight other Presence Health hospitals did not participate in the BPCI program, and their data were collected for comparison. Individual orthopedic surgeons at the participating hospitals had the choice to participate in the program and potentially benefit from the gains associated with keeping spending below the target price or the losses from spending exceeding the target price. Bundled payments were implemented at Presence on January 1st 2014.

Episodes of care were created by aggregating Medicare claims for patients with the 469 and 470 DRGs across the relevant providers in the period according to the BPCI Model 2 specification manual. Our sample was limited to episodes that started at Presence hospitals, but once the episode began, we had data on all bundled care billed to Medicare regardless of who

owned the facility. Episodes for the four quarters prior to January 1st 2014 were constructed as a baseline control.

The number of episodes at the participating hospitals in our analysis for DRG 469 matched the CMS count in a report provided to Presence and the total number of episodes for DRG 470 differed from the CMS count by 2.5%. Even if this difference is due to processing the data in a way that is inconsistent with the bundled payment program rules, we expect the effect on our estimates to be small because it will be differenced out since it affects episodes at both the participating and non-participating hospitals.

B. Study participants and baseline characteristics

The Medicare data included patients' age, gender, zip code (which was used to construct a proxy for income), race, dual eligibility, anchor stay DRG, and diagnostic ICD-9CM codes from the anchor stay. The Charlson Comorbidity Index was calculated for all patients. These variables allowed us to control for the impact of demographic characteristics and patient comorbidities on the study outcomes.²

We supplemented the CMS claims data set with de-identified health data drawn from the electronic health record (EHR) system maintained by Presence Health. These data sets covered all Medicare patients treated by the 11 hospitals in the Presence system and contained patient marital status and a larger set of diagnoses. We matched the de-identified EHR data to the corresponding de-identified CMS inpatient administrative data using a combination of blinded identifiers, the date of the inpatient anchor surgical stay, and the assigned DRG category of the anchor surgery. We used marital status and a more detailed set of diagnoses codes drawn from these data in supplementary analyses.

C. Outcomes measured

Within each episode of care, we compute the total cost of the episode of care, and the costs for the anchor stay, readmissions, inpatient rehabilitation, outpatient care, Medicare Part B (carrier

² Specifications that include the Charlson Comorbidity index, constructed using ICD9 codes during the anchors stay, as a control require us to assume the bundled payment program does not cause changes the quality of care during the surgery and anchor stay that cause complications that affect a patient's score on this index. This assumption is reasonable because the set of diseases scored by the index are unlikely to be complications of a joint replacement surgery. We also show our results are robust to excluding this control variable.

file), skilled nursing facility, and home health. Costs were defined as the cost of the services to Medicare. Supplementary analyses use alternative measures of the amount of care including number of days of care, admissions, and average length of stay. We also measure how the program affects the amount of medical treatment received by patient characteristics. We use readmissions during the episode as our quality metric. All log transformed outcomes are defined as the log of one plus the outcome variable so as to not lose observations when the outcome variable is zero.

III. Methods

We used a difference-in-differences design to estimate the effect of the bundled payment program on the cost of an episode overall and by source of care, on measures of the health of the population, and on readmissions. This approach removes any unobserved, time-constant differences between the patient populations at the participating and non-participating hospitals.

More formally, we estimate the effect of the bundled payment program on an outcome y_{it} where i indexes individual patient episodes and t indexes time. Our coefficient of interest β is estimated using the interaction of an indicator for participating hospitals, h_i , and an indicator for the year 2014, D_t . Our estimating equation also includes a constant term, δ_0 , and we define ε_{it} as an error term assumed to be orthogonal to the regressors. We initially estimate the model without including controls for patient characteristics because if physicians are systematically able to select healthier patients based on patient observables, including patient characteristics as controls will result in us underestimating the effect of the program on cost.

$$y_{it} = \delta_0 + \delta_1 h_{it} + \delta_2 D_t + \beta h_{it} D_t + \varepsilon_{it} \quad (1)$$

We then add a vector of patient characteristics, X_{it} , that includes third order polynomials of the patient's age and Charlson Comorbidity Index, gender, race, indicators for quartile of income of the patient's zip code, and an indicator for DRG 469. If the estimated β falls substantially after adding controls for patient characteristics it suggests that a lower cost set of patients are being treated post-implementation of the program.

$$y_{it} = \delta_0 + \delta_1 h_{it} + \delta_2 D_t + \beta h_{it} D_t + \Gamma X_{it} + \varepsilon_{it} \quad (2)$$

The key assumption required for a difference-in-differences model to produce consistent estimates of the effect of the program is that the participating and non-participating hospitals would have the same trend in the outcome variables absent the effect of the program. We will check for evidence this assumption holds by estimating the following equation which we will typically estimate after including controls for patient characteristics.

$$y_{it} = \gamma_0 + \gamma_1 h_{it} + \gamma_t Q_t + \sum_{q=1}^8 \alpha_q h_{it} 1[\text{quarter} = q] + \Gamma X_{it} + \varepsilon_{it} \quad (3)$$

This equation is a modified version of our primary model that replaces the indicator for 2014 with a vector of quarter-year fixed effects, Q_t , and adds an estimate the effect of being treated at a participating hospital in each quarter in 2013 and 2014 (note that $\gamma_t Q_t$ for Q4 2013 is normalized to zero). If the estimates of $\alpha_1, \dots, \alpha_4$ do not have trend, it suggests that the common trends assumption holds.

We cluster the standard errors at the operating physician-level. Because we only have 11 hospital-level clusters, we check the robustness of our results to a higher level of clustering using a permutation test that allows for within-hospital correlation in the error term rather than relying on asymptotic standard errors.

IV. Results

A. Effect of program on Medicare episode cost

Table 1 summarizes the patients' characteristics and the cost of care at the participating and non-participating hospitals in 2013 and 2014. There were 1,166 patients treated prior to the BPCI implementation, with 853 treated in non-BPCI participating hospitals and 313 in participating hospitals. After BPCI implementation, there were 1,120 patients, with 814 treated in non-participating hospitals and 306 in participating hospitals.

Prior to the implementation of bundled payments, the participating hospitals were higher cost than the non-participating hospitals and that the cost difference was driven by higher spending on post-acute care. After the bundled payments are implemented, there is a substantial

decrease in the total episode cost at the participating hospitals relative to the non-participating hospitals (Table 1).

A difference-in-differences estimate, unconditional on patient characteristics, indicates that costs fell by a marginally significant \$3,483. If we estimate the same model with log costs as the outcome, we obtain a statistically significant 12 percent decrease in the cost of an episode of care (Table 2). This result suggests that bundled payments substantially reduced the cost of the average knee and hip replacement.

After we add controls for patient characteristics, the estimated effect of bundled payments on episode costs falls to a marginally significant 6 percent (Table 2). The first panel in Figure 1 plots quarterly estimates from this model. There is no evidence of a trend before January 2014 that violates the key assumption of the difference-in-differences model, but none of the quarterly estimates after the program is implemented are statistically significant. The substantial decrease in the estimated effect after adding controls for patient characteristics is consistent with selection of healthier patients or a decrease in the predicted cost of patients at the participating hospitals that is unrelated to bundled payments.

A number of alternative specifications estimated after conditioning on patient characteristics produce cost reductions of 5-10 percent that are either marginally significant or statistically significant (Table 3). The marginally significant estimates include specifications that add quarter-year and hospital fixed effects, compute the p-values using a permutation test that allows for correlated errors at the hospital-level rather than physician-level, and dropping the partially completed Q4 2014. Estimating the model with quarter-year and operating physician fixed effects results in a statistically significant estimate, as does dropping the controls for the Charlson Comorbidity Index, and adding a third-order polynomial of the number of EHR diagnoses and marital status as controls in a smaller sample for which we could match the EHR data to the Medicare data. Using the level of total cost rather than log total cost results in an insignificant estimate.

The cost decreases were concentrated at the highest volume hospital which had a statistically significant 9 percent cost reduction, while the other two hospitals had small or no cost reductions (Online Appendix, Table A1). The program was also only effective in episodes where the operating physician also served as the inpatient attending physician and was a

participant in the program (i.e., received a share of the bonus/penalty). There was no decrease in cost if only the operating physician was a participant in the program (Online Appendix, Table A2).

B. Evidence of selection

The substantially smaller estimate after adding controls for patient characteristics suggests that the patients treated at the participating hospitals after bundled payments were introduced would be predicted to be lower cost based on their observables. This finding could simply be due to chance or it could imply that the participating physicians are reducing costs by selecting healthier patients to operate on at the participating hospitals.³ If it is simply due to chance, it suggests that the marginally significant 6 percent decrease in cost is the better estimate of the effect of the program. However, if it is due to selection, the true effect of the program could be closer to 12 percent, but a substantial portion that effect may be driven by changing the patient population rather than directly reducing costs for whoever comes through the door.

We cannot directly test for selection, but we can assess the evidence for a change in the patient population. We start by estimating the relationship between patient observables and log cost in the pre-period. We then use those estimates to predict log cost in both the pre- and post-period and estimate the difference-in-difference model using the predicted costs. We find a large but statistically insignificant, 5.6 percent decrease predicted costs at the participating hospitals. Although the result is not statistically significant, the sign is consistent with selection of patients.

A closely related approach is to look directly at the effect of the program on correlates of health. Figure 2 shows that there were increases the average health of treated patients as measured by age, dual eligibility status (Medicaid eligibility), number of diagnoses in the EHR data, and the Charlson Comorbidity Index, but only the effect on the Charlson Comorbidity Index was statistically significant (see also Online Appendix, Table A3). The effect on Charlson Comorbidity Index is driven by significant effects in the two quarters following the program's implementation.

³ For example, participating physicians could discourage sicker patients from obtaining the surgery, encourage healthier patients to obtain the surgery, or encourage healthier patients to go to the participating hospitals and sicker patients to go to the non-participating hospitals.

These results suggest that we cannot rule out a large role for selection, but the affirmative case for selection is only suggestive because we do not see significant effects on either a predicted cost based index of patient health or on most individual correlates of health.

C. Mechanisms for the Effects

A large portion of the cost to Medicare of the initial inpatient stay is essentially fixed, so any decreases in spending after conditioning on patient characteristics will tend to be driven by reductions in the cost of post-acute care. The 6 percent decrease in total cost is driven by a sharp and statistically significant decrease in spending on inpatient rehabilitation, with no evidence of an effect on readmissions, outpatient care, home health care, or the anchor stay (Figure 1). There are fairly large, statistically significant increases in physician services and skilled nursing facility expenditures in some quarters, but pooling across quarters the effects are a marginally significant 14 percent and not a significant 75 percent, respectively (Table 2 and Figure 1). The suggestive results for skilled nursing facilities and part B spending (in the carrier file) are consistent with patients substituting less expensive forms of post-acute care in place of inpatient rehabilitation.

One way for a participating physician to reduce the cost of care, is to not discharge patients to an intensive post-acute care facility. Table 4 suggests that this is in fact what happened. The decrease in inpatient rehabilitation spending is driven by an 18 percentage point decrease in the probability of being admitted to inpatient rehab. However, we see an 11 percentage point increase in the probability of ever being sent to a SNF, which indicates some of those patients are instead discharged to SNFs (Table 4). There is no corresponding increase in home health care. Thus, cost decreases were primarily caused by physicians substituting care at SNFs for inpatient rehabilitation and discharging patients to home.

Spending on post-acute care could also be cut by reducing lengths of stay at post-acute care facilities. Although the point estimates are negative, there is no significant decrease in the length of stay at inpatient rehabilitation facilities or SNFs, but there are significant decreases in days of home health care used (Table 4). It is likely more difficult for the participating physician or hospital to induce reductions on the intensive margin of post-acute care, particularly if the facility is not owned by Presence Health. However, shifting patients from inpatient rehabilitation facilities to SNFs could leave both populations unobservably sicker, resulting in increases in the average length of stay, absent any effort to reduce it.

Presence Health could gain more control over utilization along the intensive margin and to increase its revenue is to shift patients to Presence Health owned post-acute care facilities from non-Presence health facilities. Online Appendix, Table A4 shows the effect of the program on spending on care at hospitals, skilled nursing homes, and home health by Presence Health and non-Presence providers. The program caused increases in Presence SNF and home health expenditures and decreases in non-Presence Health SNF and home health expenditures. The effect on SNFs is statistically significant while effect for home health is marginally significant. However, Presence share of both SNF and home health revenue increased significantly (Online Appendix, Table A5).

D. Effect of the program on patient welfare

1. Incidence of effects

We tested whether the reductions in care impacted the patients who are likely to have the greatest need for post-acute care. Table 3 shows that episode costs decreased by 10 percent for DRG-470, while the cost of DRG-469 increased by 40 percent. DRG-469 covers procedures on patients with major co-morbidities or complications, while DRG-470 includes patients without major co-morbidities or complications who may have less need for intensive post-acute care. Our results point to physicians at Presence targeting their post-acute care reductions to avoid affecting the most severely ill patients.

2. Readmissions

Readmissions would be expected to increase if less intensive post-acute care harmed patients' health. Table 2 and Table 4 show that the total cost of readmission, the number of readmissions days, and the probability of ever being readmitted are not significantly affected by the bundled payments. However, the coefficients are imprecisely estimated and the 95% confidence interval can only reject that the bundled payment program caused fairly large percentage changes in the probability of being readmitted. These results are consistent with BPCI not harming patients, but better data on patient outcomes is required to show conclusively whether or not the program negatively affects patient health.

V. Discussion

Our results suggest that the BPCI program successfully reduced the average episode cost for total hip and knee replacement patients at the participating Presence Health facilities by 6 to 12 percent. The 12 percent effect on cost is larger than what has been found in most previous studies of bundled payments (Cromwell et al. 1997, Berry et al. 2009, Hussey et al. 2012, Urdapilleta et al. 2013, Alexander 2017, and Dummit et al. 2016), and another alternative payment model, Medicare's recent Accountable Care Organization pilot programs (McWilliams et al. 2015 and Nyweide et al. 2015).

Around half of the 12 percent unconditional decrease is explained by selecting patients with observables that suggest they are low cost, although our measures of selection are themselves typically not statistically significant. This result suggests that bundled payments may reduce medical spending by more than implied by studies that control for patient characteristics. The remaining portion of the effect is driven by reduced utilization of inpatient rehabilitation at the highest volume hospital for patients without major complications or comorbidities. These results suggest that bundled payments may be more effective in relatively healthy populations with relatively high utilization of intensive post-acute care. High volume hospitals may be in a better position to reduce spending because the returns to putting in place a system to reduce costs is greater and the larger sample of patients reduces the variance of outcomes which incentivizes greater physician effort to reduce costs.

The program may also have led to a more efficient use of health care resources because spending for patients with major complications and comorbidities increased significantly, while spending on care for patients without major complications or co-morbidities fell significantly. Although the possibility of selection of relatively healthier patients into surgery suggests that bundled payments could reduce access to care for relatively sick prospective joint replacement patients.

The reduction in inpatient rehabilitation utilization, increase in SNF use, and the shift from non-Presence-owned to Presence-owned SNFs and home health suggest that physicians have the ability to influence their patients' selection of post-acute providers and services. Although there is little evidence that they could influence the length of time spent in post-acute care facilities. The shift in patients to Presence owned SNFs may also have offset revenue losses

from reduced utilization of Presence-owned inpatient rehabilitation services, and could, in the future, give participating hospitals greater ability to reduce length of stay in post-acute care.

The fact that the spending cuts were concentrated among relatively healthier patients, and the increased utilization of SNFs to partially replace inpatient rehab stays makes it less likely that patient health will suffer because of reduced care. Although it is possible that the incentives could have the adverse effect of encouraging physicians to discharge to low cost post-acute services at the expense of quality of care or patient convenience (Adida et al 2016). There is evidence for patients who experience medical emergencies that more intensive post-acute care utilization is associated with worse outcomes (Doyle et al. forthcoming), and Dummit et al. (2016) found no evidence of bundled payments on readmissions, emergency department visits, and for a smaller sample satisfaction with their recovery. However, larger studies assessing patient reported outcomes and satisfaction, and better measures of patient health are needed to fully assess the impact on the patient population.

An important limitation of our study is that our only measure of patient health outcomes is readmissions. It is possible that the bundled payments could cause worse patient health outcomes without causing increase in readmissions. We also are only able to measure short-run decreases in cost at the participating hospitals. It is possible that longer-term effects on the cost of care differ from the short-term effects on the cost of care. Finally, because we only have data from one hospital system, there may be correlation in the error term at the level of all participating or non-participating hospitals or all hospitals in the system due for a variety of reasons including decisions by upper management.

References

Adida E, Mamani H, and Nassiri S. 2016. “Bundled Payment vs. Fee-for-Service: Impact of Payment Scheme on Performance.” *Management Science* 63(5): 1606-1624.

Alexander, D. 2017. “How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs.” *Working Paper*.

Berry, SA, Doll MC, McKinley KE, Casale AS, Bothe A Jr. 2009. “ProvenCare: Quality improvement model for designing highly reliable care in cardiac surgery. *BMJ Quality & Safety* 18(5):360–368.

Berwick, DM and Hackbarth, AD. 2012. “Eliminating Waste in US Health Care,” *The Journal of the American Medical Association*, 307(14): 1513-1516.

Center for Medicare and Medicaid Services. 2015. “Bundled Payments for Care Improvement (BPCI) Initiative: General Information” [accessed on 12/21/2015]. Available at: <https://innovation.cms.gov/initiatives/bundled-payments/>.

Center for Medicare and Medicaid Services. 2015. “Comprehensive Care for Joint Replacement Model” [accessed on 12/21/2015]. Available at: <https://innovation.cms.gov/initiatives/cjr/>.

Cromwell J, Dayhoff DA, and Thoumaian AH. 1997. “Cost savings and physician responses to global bundled payments for Medicare heart bypass surgery.” *Health Care Financing Review* 19(1): 41-57.

Cutler, D. 2014. *The Quality Cure: How Focusing on Health Care Quality Can Save Your Life and Lower Spending Too* University of California Press: Berkeley, CA.

Dobson A, DaVanzo JE, Heath S, et al. 2012. “Medicare Payment Bundling: Insights from Claims Data and Policy Implications: Analyses of Episode-based Payment.” *American Hospital Association and Association of American Medical Colleges*.

Doyle J, Graves J, and Gruber J. Forthcoming. “Uncovering Waste in U.S. Healthcare.” *Journal of Health Economics*.

Dummit L, Kahvecioglu D, Marrufo G, Rajkumar R, Marshall J, Tan E, Press MJ, Flood S, Muldoon LD, Gu Q, Hassol A, Bott DM, Bassano A, Conway PH. 2016. “Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes.” *The Journal of the American Medical Association* 316(12):1267-1278.

Garber, AM and Skinner, J. 2008. “Is American Health Care Uniquely Inefficient?” *Journal of Economic Perspectives* 22(4): 27-50.

Hussey PS, Mulcahy AW, Schnyer C, Schneider EC. Bundled Payment: Effects on Health Care Spending and Quality. 2012. “Closing the Quality Gap: Revisiting the State of the Science.” Evidence Report/Technology Assessment No. 208. (Prepared by RAND Evidence-based Practice Center under Contract No. 290-2007-10062-I.) AHRQ Publications No. 12-E007-EF. Rockville,

MD: Agency for Healthcare Research and Quality. Available at:
www.effectivehealthcare.ahrq.gov/reports/final.cfm.

McWilliams JM, Chernew ME, Landon BE, Schwartz AL. 2015. "Performance Differences in Year 1 of Pioneer Accountable Care Organizations." *The New England Journal of Medicine* 372(2):1927-1936.

Nyweide DJ, Woolton L, Cuerdon TT, et al. 2015. "Association of Pioneer Accountable Care Organizations vs Traditional Medicare Fee for Service With Spending, Utilization, and Patient Experience." *The Journal of the American Medical Association* 313(21):2152-2161.

Siddiqi A, White PB, Mistry JB, Gwam CU, Nace J, Mont MA, Delanois RE. 2017. "Effect of Bundled Payments and Health Care Reform as Alternative Payment Models in Total Joint Arthroplasty: A Clinical Review." *The Journal of Arthroplasty* 32(8): 2590–2597.

Urdapilleta O, Weinberg D, Pedersen S, Kim G, Cannon-Jones S, Woodward J. 2013. "Evaluation of the Medicare Acute Care Episode (ACE) Demonstration Final Evaluation Report." *Centers for Medicare and Medicaid Innovation*.

Table 1: Summary Statistics

		2013		2014	
		Non- Participating	Participating	Non- Participating	Participating
(1)	Age	73	77	72	76
(2)	Male	0.32	0.27	0.30	0.31
(3)	Married	0.44	0.36	0.39	0.38
(4)	White	0.87	0.87	0.87	0.90
(5)	Black	0.08	0.07	0.07	0.06
(6)	Other	0.06	0.06	0.06	0.05
(7)	Zipcode Income	\$74,435	\$80,314	\$72,741	\$78,049
(8)	Diagnoses	10.12	11.06	10.32	11.05
(9)	Co-Morbidity Index	1.07	1.27	1.13	0.97
(10)	Dual Eligibility	0.21	0.19	0.22	0.13
(11)	DRG 469	0.07	0.10	0.06	0.07
(12)	Total Cost	\$31,095	\$38,976	\$30,751	\$35,149
(13)	Anchor Stay	\$12,065	\$12,712	\$12,144	\$12,384
(14)	Readmissions	\$160	\$167	\$166	\$160
(15)	Inpatient Rehab	\$2,690	\$4,647	\$2,707	\$1,382
(16)	Outpatient File	\$558	\$516	\$576	\$555
(17)	Part B (Carrier) File	\$5,989	\$6,265	\$5,438	\$6,962
(18)	Skilled Nursing Facility	\$6,845	\$11,251	\$7,321	\$10,827
(19)	Home Health Agency	\$2,136	\$2,856	\$2,063	\$2,313
(20)	Number of Patients	853	313	814	306

Data cover the 2013-2014 period. The unit of observation is an episode of care. The sample consists of all patients who have a non-missing value of each variable.

Table 2: Effect of Bundled Payments on Log Episode Costs

	Outcome Variable	Diff-in-Diff Est.	SE	p-value
<i>No controls for patient characteristics:</i>				
(1)	Total Cost	-0.12	(0.06)	[0.045]
<i>Including controls for patient characteristics:</i>				
(2)	Total Cost	-0.06	(0.03)	[0.090]
(3)	Anchor Stay Cost	-0.02	(0.02)	[0.287]
(4)	Readmission Cost	0.21	(0.17)	[0.217]
(5)	Inpatient Rehab Cost	-1.72	(0.38)	[0.000]
(6)	Outpatient Cost	0.38	(0.39)	[0.335]
(7)	Part B Cost	0.14	(0.09)	[0.093]
(8)	SNF Cost	0.75	(0.45)	[0.101]
(9)	Home Health Cost	-0.18	(0.51)	[0.724]

Data cover the 2013-2014 period. The unit of observation is an episode of care. Each row reports the coefficient, standard error, and p-value for an indicator for the participating hospitals interacted with an indicator for 2014. These parameters are estimated in a regression of the logged outcome variable on that interaction term, a participating hospital indicator, a year 2014 indicator, and in rows (2)-(9) a vector of patient characteristics. The standard errors are clustered by operating physician.

Table 3: Robustness Checks

Alternative Specifications	Diff-in-Diff		
	Est.	SE	p-value
(1) Quarter-Year and Hospital FE	-0.05	(0.03)	[0.067]
(2) Permutation Test P-Value	-0.05	(NA)	[0.066]
(3) Quarter-Year and Surgeon FE	-0.06	(0.03)	[0.036]
(4) Drop Q4 2014	-0.06	(0.03)	[0.076]
(5) Add EHR Diagnosis Controls and Marital Status	-0.10	(0.03)	[0.003]
(6) Drop Charlson Comorbidity Index	-0.08	(0.04)	[0.042]
(7) Total Cost in Levels	-\$1,288	(1,091)	[0.240]
(8) Only DRG 469	0.40	(0.14)	[0.006]
(9) Only DRG 470	-0.10	(0.04)	[0.013]
(10) Inpatient Rehab: Only DRG 469	-0.42	(1.60)	[0.793]
(11) Inpatient Rehab: Only DRG 470	-1.81	(0.45)	[0.000]

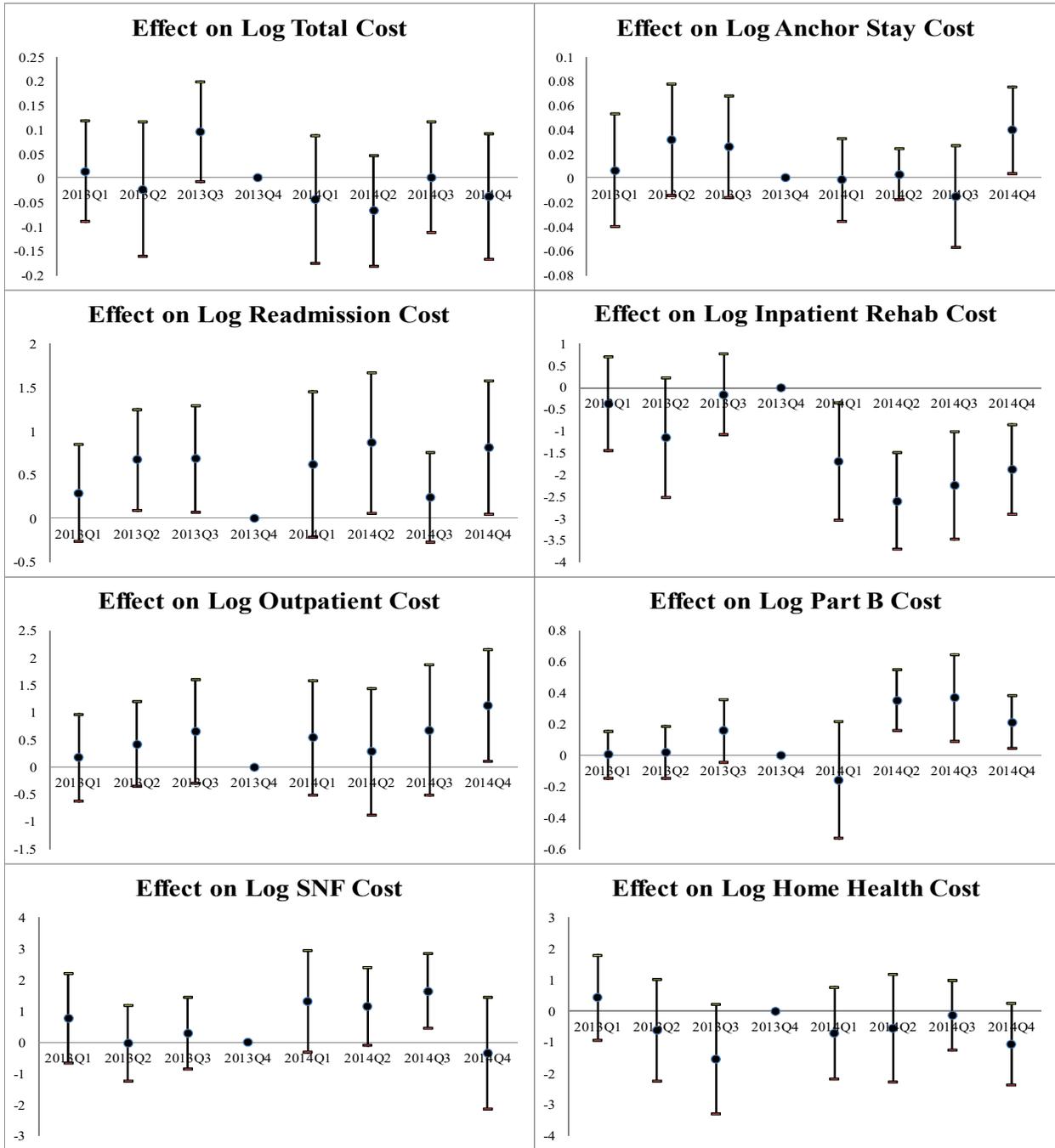
Data cover the 2013-2014 period. The unit of observation is an episode of care. In each row we modify our baseline difference-in-differences regression to test the robustness of the results. In row (1) we replace the year 2014 and participating hospital indicators with year-quarter and hospital fixed effects. In row (2) we use the specification from row (1) but compute the p-value using a permutation test within-hospital to allow for correlation in the errors within hospital and across operating physicians. In row (3) we replace the hospital fixed effects with operating physician fixed effects. In row (4) we drop the fourth quarter of 2014 from our baseline specification. In row (5) we include controls merged from the medical records data (married, a third order polynomial of the number of diagnosis). In row (6) we exclude our controls for the Charlson Comorbidity Index. In row (7) we analyze the episode cost in levels rather than logs. In rows (8) and (10) we limit the sample to DRG 469. In rows (9) and (11) we limit the sample to DRG 470.

Table 4: Effect of Bundled Payments on Other Margins

	Anchor Stay	Readmission	Inpatient Rehab	SNF	Home Health
	(1)	(2)	(3)	(4)	(5)
(1) Days	-0.15 (0.21) [0.473]	0.19 (0.32) [0.556]	-2.30 (0.41) [0.000]	2.28 (1.57) [0.148]	-8.46 (3.63) [0.021]
(2) Ever Admitted		0.02 (0.02) [0.277]	-0.18 (0.04) [0.000]	0.11 (0.05) [0.032]	-0.01 (0.06) [0.862]
(3) Days Conditional on Admission		-1.21 (4.85) [0.804]	-0.88 (1.45) [0.547]	-1.17 (2.36) [0.621]	-10.22 (4.24) [0.018]

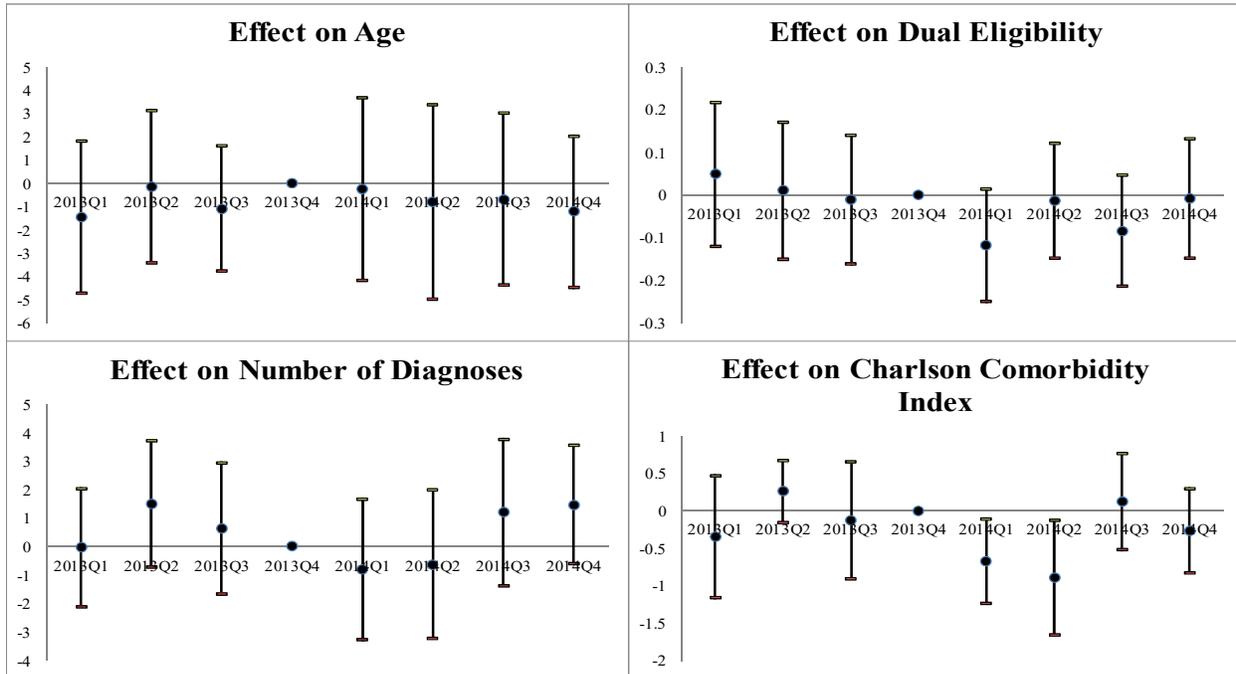
Data cover the 2013-2014 period. The unit of observation is an episode of care. The outcome variables are listed in the rows and the sample is listed in the columns, e.g., the first cell in the table (1,1) reports results from a regression where the outcome variable is the number of days in the anchor stay. Each cell contains the coefficient, standard error (in parenthesis), and p-value (in brackets) for an indicator for the participating hospitals interacted with an indicator for 2014. These parameters are estimated in a regression of the outcome variable on that interaction term, a participating hospital indicator, a year 2014 indicator, and a vector of patient characteristics. The standard errors are clustered by operating physician.

Figure 1: Effect of Bundled Payments on Quarterly Log Episode Cost



Notes: Plots are of coefficients on the interaction between quarter-year indicators and a participating hospital indicator normalized so the value of the Q4 2013 coefficient is zero. The coefficients are from a regression of a log cost variable on a quarter-year participating hospital interaction, a participating hospital fixed effect, quarter-year fixed effects, and controls for patient characteristics and co-morbidities. Error bars show the 95-percent confidence interval from a hypothesis test that the plotted coefficient equals the Q4 2013 coefficient. Standard errors are clustered by the operating physician.

Figure 2: Effect of Bundled Payments on Correlates of Health



Notes: Plots are of coefficients on the interaction between quarter-year indicators and a participating hospital indicator normalized so the value of the Q4 2013 coefficient is zero. The coefficients are from a regression of correlates of health on a quarter-year participating hospital interaction, a participating hospital fixed effect, and quarter-year fixed effects. Error bars show the 95-percent confidence interval from a hypothesis test that the plotted coefficient equals the Q4 2013 coefficient. Standard errors are clustered by the operating physician.