DOI: 10.1377/hlthaff.2019.00466
HEALTH AFFAIRS 39,
NO. 1 (2020): 58–66

By Amol S. Navathe, Ezekiel J. Emanuel, Atheendar S. Venkataramani, Qian Huang, Atul Gupta, Claire T. Dinh, Eric Z. Shan, Dylan Small, Norma B. Coe, Erkuan Wang, Xinshuo Ma, Jingsan Zhu, Deborah S. Cousins, and Joshua M. Liao

Spending And Quality After Three Years Of Medicare's Voluntary Bundled Payment For Joint Replacement Surgery

Amol S. Navathe (amol.navathe@gmail.com) is a core investigator at the Corporal Michael J. Cresencz Veterans Affairs Medical Center, in Philadelphia, and an assistant professor in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, and a senior fellow at the Leonard Davis Institute of Health Economics, both at the University of Pennsylvania.

Ezekiel J. Emanuel is the Diane V. S. Levy and Robert M. Levy University Professor, chair of the Department of Medical Ethics and Health Policy, and vice provost for global initiatives, all at the University of Pennsylvania.

Atheendar S. Venkataramani is an assistant professor of medical ethics and of health policy at the Perelman School of Medicine, University of Pennsylvania.

Qian Huang is a statistical analyst in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania.

Atul Gupta is an assistant professor in the Department of Health Care Management, Wharton School, University of Pennsylvania.

ABSTRACT **Medicare has reinforced its commitment to voluntary bundled payment by building upon the Bundled Payments for Care Improvement (BPCI) initiative via an ongoing successor program, the BPCI Advanced Model. Although lower extremity joint replacement (LEJR) is the highest-volume episode in both BPCI and BPCI Advanced, there is a paucity of independent evidence about its long-term impact on outcomes and about whether improvements vary by timing of participation or arise from patient selection rather than changes in clinical practice. We found that over three years, compared to no participation, participation in BPCI was associated with a 1.6 percent differential decrease in average LEJR episode spending with no differential changes in quality, driven by early participants. Patient selection accounted for 27 percent of episode savings. Our findings have important policy implications in view of BPCI Advanced and its two participation waves.**

Bundled payment has the potential to improve the value of care by holding hospitals accountable for episode-specific quality and costs. For example, participation in bundled payment programs from the Centers for Medicare and Medicaid Services (CMS) has been associated with financial savings and stable quality for lower extremity joint replacement (LEJR) surgery episodes.[1]

CMS has demonstrated particular commitment to voluntary bundled payment. In 2013 the agency scaled voluntary bundled payment for LEJR nationwide via the Bundled Payments for Care Improvement (BPCI) initiative.[1] One peer-reviewed evaluation conducted by CMS and its contractors estimated that there had been a 3.9 percent reduction in LEJR episode spending without adverse effects on quality after one year of BPCI.[2] This finding was further corroborated at longer durations by subsequent non-peer-reviewed evaluations conducted by CMS contractors.[3,4] Consequently, CMS elected to continue voluntary bundled payment in October 2018 through the national five-year BPCI Advanced model, in which participating hospitals and physician groups accept bundled payment for up to thirty-seven clinical episodes that correspond to various surgical procedures and medical conditions. BPCI Advanced is the only new advanced alternative payment model established under the administration of President Donald Trump. LEJR is the most commonly selected episode in BPCI Advanced, which has already enrolled nearly 1,300 participants in the first of its two participation waves.[5]

Despite growing participation in voluntary bundled payment for LEJR, two policy salient issues remain unaddressed. First, to our knowledge, there are no studies of BPCI's nationwide impact beyond evidence produced by CMS and its contractors, which has resulted in an overall

paucity of evidence about the long-term impact of BPCI on spending and patient outcomes—particularly based on the timing of organizations' entry into BPCI. This issue is relevant to BPCI Advanced because the program enrolls organizations over time in two waves, separated by fifteen months. Furthermore, evaluations of other payment reforms suggest that benefits can require several years to emerge.[6,7] However, no independent peer-reviewed evaluations have addressed the effects of voluntary bundled payment for LEJR over multiple years of follow-up, and none have examined the timing of participation and benefits as key policy concerns.

Second, voluntary programs such as BPCI and BPCI Advanced are susceptible to patient selection—that is, changes in the case-mix of patients cared for after organizations begin program participation. Consequently, evaluations of voluntary bundled payment must account for the possibility that cost savings associated with participation may arise from the selection of healthier patients rather than practice improvements.[8] Unfortunately, existing evaluations have been limited in accounting for such selection.

Given these critical policy issues, we used a novel analytic approach to address patient selection while conducting the first independent multiyear evaluation of the association between timing of participation in voluntary bundled payment for LEJR and changes in episode spending and quality.

Study Data And Methods

STUDY PERIODS We used January 2011–September 2013 as the baseline (pre-BPCI) period and October 2013–December 2016 as the intervention (entire BPCI) period. To evaluate performance over time, we divided the intervention period into an early BPCI period (October 2013–June 2015) and a late BPCI period (July 2015–December 2016). In the late BPCI period, data through September 2016 were used to define LEJR episodes to allow for a ninety-day postdischarge period between October and December 2016.

MARKETS AND HOSPITALS BPCI consisted of four separate models of care. Among these, model 2 was the most popular and largest in terms of enrollment. Hospitals in model 2 accepted bundled payment for episodes beginning with hospitalization and spanning up to ninety days of postacute care. We identified hospitals that participated in LEJR bundled payment under model 2 ("BPCI hospitals") by compiling publicly available quarterly BPCI participant lists.[2] This allowed us to define BPCI participation per quarter, which reflected both the time-varying nature of program enrollment and the potential for entry into and exit from the program over time. Markets were defined by hospital referral region[9] and categorized as those that contained at least one BPCI hospital at any time in the program ("BPCI markets") or those that contained no BPCI hospitals ("non-BPCI markets"). Our comparison group of hospitals consisted of those that never participated in model 2 and were located in non-BPCI markets ("non-BPCI hospitals").

To evaluate BPCI performance based on duration of participation, we divided BPCI hospitals into "early entrants" (those that began participating in the early BPCI period) and "late entrants" (those that began participating in the late BPCI period). We obtained hospital and market characteristics using data from Medicare claims, the American Hospital Association, Hospital Compare, and the 2017 CMS Improving Medicare Post-Acute Care Transformation (IMPACT) Act file, as done in prior work.[10–12]

PATIENT POPULATION AND EPISODE CONSTRUCTION We identified patients admitted to BPCI hospitals ("BPCI patients") and non-BPCI hospitals ("non-BPCI patients") for major hip and knee joint replacement or reattachment of lower extremity with and without major complicating or comorbid condition (Medicare Severity Diagnosis-Related Groups 469 and 470, respectively). We used a 100 percent sample of Medicare beneficiaries admitted to BPCI hospitals in 2011–16 and a 20 percent national sample of beneficiaries over the same period to identify patients admitted to non-BPCI hospitals. We excluded people not continuously enrolled in fee-for-service Medicare for the episode and a 180-day lookback period, those with end-stage renal disease, and those who died during the initial hospitalization for LEJR surgery. To increase sample homogeneity,[1,13] we also excluded patients younger than age sixty-five or older than age ninety and those using hospice.

Patient characteristics used to control for baseline health included age; sex; race; dual eligibility status (that is, eligibility for Medicare and Medicaid); Elixhauser comorbidities;[14] and prior use of an acute care hospital, skilled nursing facility (SNF), or inpatient rehabilitation facility (IRF). Consistent with prior literature,[10,13,15] LEJR episodes were constructed beginning with hospitalization for Medicare Severity Diagnosis-Related Group 469 or 470 and spanning ninety days after hospital discharge, using all Medicare data for inpatient, outpatient, SNF, and other claims. In line with BPCI program rules, we excluded overlapping or repeat episodes.[2] We also excluded episodes covered by the Comprehensive Care

Claire T. Dinh is a medical student at Harvard Medical School, in Boston, Massachusetts. She was a research coordinator in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, when this work was completed.

Eric Z. Shan is a research assistant in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania.

Dylan Small is a professor in the Department of Statistics, University of Pennsylvania.

Norma B. Coe is an associate professor in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania.

Erkuan Wang is a data analyst in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania.

Xinshuo Ma is a data analyst in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania.

Jingsan Zhu is associate director of data analytics in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania.

Deborah S. Cousins is a project manager in the Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania.

Joshua M. Liao is medical director of payment strategy, director of the Value and Systems Science Lab, and an assistant professor in the Department of Medicine, all at the University of Washington, in Seattle, and an adjunct senior fellow at the Leonard Davis Institute of Health Economics, University of Pennsylvania.

for Joint Replacement model.[16]

OUTCOMES Our primary spending outcome was the average Medicare payment per LEJR episode. Secondary spending outcomes included the proportion of average episode spending attributable to the index hospitalization, readmissions, SNF care, IRF care, home health agency services, professional services, and all remaining components. Spending estimates were standardized following methods used previously and transformed into 2016 dollars.[10–12,17,18] Quality outcomes included ninety-day postdischarge risk-standardized mortality, unplanned readmission, and emergency department visit rates, as well as LEJR-specific complication rates.[19]

STATISTICAL ANALYSES We compared patient, hospital, and market characteristics between BPCI and non-BPCI hospitals and markets. Chi-square tests were used to compare categorical variables, while *t*-tests and Wilcoxon rank-sum tests were used to compare continuous variables.

We conducted unadjusted analyses that compared changes in outcomes across study periods. In adjusted analyses we used a difference-in-differences method to estimate differential changes in spending and quality outcomes for BPCI hospitals compared to a 1:1 propensity-matched set of non-BPCI hospitals in the pre-BPCI and BPCI periods.[20,21] We used baseline hospital and market characteristics and matched BPCI and non-BPCI hospitals based on propensity score values with identical first two digits. Because hospitals that did not participate in BPCI but were located in BPCI markets could have been affected by the program (for example, as a result of changes in market share or referral patterns after it began), only non-BPCI hospitals in non-BPCI markets were used in propensity score matching (for details about our propensity score matching approach, see the appendix methods 1 section and appendix exhibit 1, both in the online appendix).[22]

Episode spending was analyzed using generalized linear models with a log link and gamma distribution, while secondary spending outcomes and binary quality outcomes were analyzed using ordinary least squares regression. In adjusted analyses all models controlled for patient demographics, clinical conditions (for a list of conditions used in statistical models, see appendix exhibit 2),[22] and time-varying market characteristics; included hospital and quarter fixed effects; and used bootstrapped standard errors.[23,24]

Evaluations of voluntary LEJR bundled payment are susceptible to selection (that is, changes in outcomes may arise from changes in the types of patients receiving LEJR under bundled payment rather than changes in care delivery processes at hospitals after participation starts), which could lead to spurious inferences about program-related savings and quality improvements. Such selection—in which bundled-payment participants shift toward lower-risk patients based partially on characteristics unobservable in claims data, and ostensibly to increase financial gain—has been previously described under BPCI.[25]

Such selection may occur for several reasons. For example, physicians who maintain operating privileges at multiple hospitals could preferentially "sort" lower-risk patients to BPCI hospitals and higher-risk patients to non-BPCI hospitals. Hospitals could also engage in strategies that drive preferential selection, such as practice acquisition or network expansion.

To address such selection and mitigate its effects, we used an instrumental variable together with our difference-in-differences method. The instrumental variable used historical hospital referral patterns before BPCI began, to identify whether in the absence of the BPCI program, a given patient would have been hospitalized for LEJR at a hospital that would later become a BPCI hospital versus one that would remain a non-BPCI hospital (for details about our instrumental variable approach, see the appendix methods 2 section and appendix exhibits 3 and 4).[22] Specifically, we used hospital referral patterns in 2011 to generate the predicted probability of hospitalization for LEJR at an eventual BPCI hospital.

In the first stage of a two-stage least squares instrumental variable regression, we used this predicted probability as an instrument for actual hospitalization for LEJR at a BPCI hospital. In the second stage, instrumented admission to a BPCI hospital was related to spending and quality outcomes. Though selection based on unobservable characteristics could also have occurred before BPCI, changes in patient selection as a result of subsequent BPCI participation could not have been influenced by hospital referral patterns in that period specifically. Therefore, this instrumental variable approach helped mitigate effects of any patient selection that occurred after hospitals began participating in BPCI based on characteristics unobservable in our data.

We evaluated changes in outcomes associated with BPCI participation for the BPCI period overall, as well as by early versus late BPCI periods and early- versus late-entrant hospitals. We also tested the parallel trends assumption. Statistical tests were two-tailed and considered significant at $\alpha = 0.05$. Robust standard errors were corrected for heteroscedasticity.[26]

Analyses were performed using SAS, version

9.4, or Stata, version 15.1. This study was approved by the University of Pennsylvania Institutional Review Board, with a waiver of informed consent.

SENSITIVITY ANALYSES We extended methods from prior CMS contractor evaluations to conduct analyses without an instrumental variable or hospital fixed effects.[1] To account for hospital practice changes that occurred in anticipation of BPCI participation, we also conducted analyses that excluded January–September 2013 from the baseline period. Additionally, we conducted analyses that used a less stringent optimal propensity score matching approach and a caliper of 0.05 that varied the match ratio from 1:1 to 1:3 (for details about our propensity score matching approach, see the appendix methods 1 section in the online appendix),[22] as well as analyses that excluded hospitals that participated in the Comprehensive Care for Joint Replacement model from propensity score matching and subsequent analysis.

LIMITATIONS This study had several limitations. First, it was an observational study, and thus our results could be confounded by omitted variables and both patient and hospital selection. However, these concerns were mitigated by the quasi-experimental design that incorporated a robust set of patient, hospital, and market characteristics. In contrast to prior BPCI evaluations, this analysis also incorporated time-varying participation (that is, hospitals served as controls prior to BPCI participation), hospital fixed effects to address hospital selection based on time-invariant unobservable factors, and an instrumental variable to reduce potential confounding from patient selection.

Second, results from our instrumental variable approach applied only to beneficiaries who received LEJR at BPCI hospitals regardless of the BPCI model, not to all beneficiaries who received LEJR.

Third, while our use of an instrumental variable accounted for unobserved selection in ways that prior studies did not, more work is needed to ensure that all sources of selection are accounted for in policy evaluations. This is particularly true because our analytic approach did not account for all forms of selection—such as the selection of healthier patients for LEJR who would not have received the procedure without bundled payment (that is, in the event that BPCI hospitals induced demand for LEJR).

Fourth, secondary outcomes were analyzed with ordinary least squares regression, in part because of the need to include large numbers of fixed effects. While changes in mean estimates rather than predicted values were of interest, this approach might not have accounted for the skewness of data.

Fifth, results might not be generalizable to medical condition episodes or episodes initiated by physician group practices.

Sixth, this study focused only on model 2 in BPCI. However, that model was the largest in terms of enrollment and the basis for BPCI Advanced.

Seventh, the study findings might not apply to mandatory programs such as the Comprehensive Care for Joint Replacement model.[10]

Study Results

Our sample included 244 BPCI hospitals in 123 BPCI markets during the BPCI period (exhibit 1), with 10,757 BPCI patients per quarter in the early BPCI period and 10,949 in the late BPCI period (data not shown). Our matched comparison group consisted of 244 non-BPCI hospitals operating in 98 non-BPCI markets during the BPCI period. There were 2,819 non-BPCI patients per quarter at non-BPCI hospitals in the early BPCI period and 2,375 in the late BPCI period. Wald tests did not indicate divergent secular trends between BPCI and non-BPCI patients during the pre-BPCI period (for details about our tests of parallel trends, see the appendix methods 3 section in the online appendix).[22]

PATIENT, HOSPITAL, AND MARKET CHARACTERISTICS A number of characteristics differed by study period among both BPCI and non-BPCI patients (exhibit 2; see also appendix exhibit 5).[22] For example, in both cases, the mean Elixhauser comorbidity index of patients was significantly lower in the late BPCI period than the early BPCI or pre-BPCI periods. There were several other patient characteristics with small but significant differences by study period, although the changes did not exhibit meaningful differential trends across BPCI and non-BPCI patients (appendix exhibit 6).[22]

BPCI and non-BPCI hospitals varied with respect to a number of characteristics, with differences similar to those described in prior reports.[8,27] For example, compared to non-BPCI hospitals, BPCI hospitals were larger and more likely to be urban, not for profit, and teaching hospitals (exhibit 1). In both BPCI and non-BPCI markets, Medicare Advantage penetration and the proportion of markets with physician group practices were greater in the late BPCI period compared to the early BPCI and pre-BPCI periods (appendix exhibit 7).[22] Differences between non-BPCI and BPCI hospitals were small after propensity score matching (appendix exhibit 1).[22]

EPISODE SPENDING For BPCI patients, mean episode spending was $23,552 in the pre-BPCI period and $22,129 in the BPCI period (a reduc-

EXHIBIT 1

Baseline characteristics of hospitals that participated in the Bundled Payments for Care Improvement (BPCI) initiative in 2011 and those that did not, after propensity score matching

Characteristic	Non-BPCI hospitals (n = 244)	BPCI hospitals (n = 244)	Standardized difference
HOSPITAL ADMISSIONS			
Mean annual admissions for top 10 BPCI episodes (%)[a]	22.7	22.6	–0.03
Median no. of annual admissions for LEJR	170	179	0.01
Mean share of discharges to the highest-volume SNF (%)	28.9	29.5	0.04
Median share of discharges to the highest-volume IRF (%)	90.0	80.0	–0.10
Median 90-day readmission rate	11.0	11.0	0.12
Median 90-day LEJR episode spending ($)	23,936	24,355	0.09
HOSPITAL ORGANIZATIONS			
Median no. of beds	235	261	0.06
Ownership status (%)			
For profit	17.2	16.4	–0.02
Not for profit	77.1	78.7	0.04
Government owned	5.7	4.9	–0.04
Member of a system (%)	79.1	78.7	–0.01
Teaching status (%)[b]			
Major teaching	13.5	12.7	–0.02
Minor teaching	31.6	34.4	0.06
Nonteaching	54.9	52.9	–0.04
Median ratio of interns and residents to beds	0.0	0.0	0.08
Median DSH payments ($)[c]	2,162,258	2,472,058	0.02
Urban location (%)	99.2	99.2	0.00
Mean Medicare days (% of total patient days)	52.9	52.0	–0.07
Median market share (%)	9.2	9.2	0.03
MARKETS			
Median no. of beneficiaries	1,068,113	1,321,591	0.14
Median income ($)	52,303	53,089	0.05
Median SNF beds	6,124	6,308	–0.01
Median IRF beds	122	134	0.05
Medicare Advantage penetration[d]	24.7	25.4	0.05
Hospital market concentration (HHI)	2,216	2,181	–0.02

SOURCE Authors' analysis of 2011 data on hospital characteristics from the American Hospital Association (AHA) Annual Survey and 2011 Medicare claims data. **NOTES** The exhibit shows the characteristics of hospitals used in propensity score matching. Standardized difference refers to the difference in the mean values of a characteristic between the non-BPCI hospital and BPCI hospital groups, divided by an estimate of the standard deviation of the values of that characteristic. Medians are provided where the data are skewed. A fuller version of this exhibit, showing standardized differences before and after propensity score matching, is in appendix exhibit 1 (see note 22 in text). LEJR is lower extremity joint replacement. SNF is skilled nursing facility. IRF is inpatient rehabilitation facility. HHI is Herfindahl-Hirschman Index. [a]Major joint replacement of the lower extremity; double joint replacement of the lower extremity; revision of the hip or knee; hip and femur procedures except major joint; lower extremity and humerus procedure except hip, foot, and femur; coronary artery bypass graft; acute myocardial infarction; congestive heart failure; simple pneumonia and respiratory infections; and chronic obstructive pulmonary disease or bronchitis or asthma. [b]The AHA Annual Survey defines major teaching hospitals as those that are members of the Council of Teaching Hospitals (COTH), minor teaching hospitals as non-COTH members that reported a medical school affiliation to the American Medical Association, and nonteaching hospitals as all other hospitals. [c]Disproportionate share hospital (DSH) payment percentages are derived from the 2017 CMS Improving Medicare Post-Acute Care Transformation (IMPACT) file. [d]The number of Medicare Advantage enrollees in a market divided by the number of Medicare beneficiaries in that market.

tion of $1,423; $p < 0.001$) (appendix exhibit 8).[22] Mean episode spending among non-BPCI patients was $22,834 in the pre-BPCI period and $22,073 in the BPCI period (a reduction of $761; $p < 0.001$). In unadjusted analyses of raw means, the differential change in mean episode spending for BPCI versus non-BPCI patients was –2.7 percent (a differential reduction of approximately $662; $p < 0.001$).

In an adjusted difference-in-differences analysis that used the instrumental variable, LEJR among patients admitted to BPCI hospitals was associated with a 1.6 percent differential decrease (of approximately $377) in episode spending before versus after hospitals initiated BPCI participation, as compared to changes for patients admitted to non-BPCI hospitals (exhibit 3). In the early BPCI period, BPCI participation was associated with a 1.8 percent differential decrease in mean episode spending. In contrast, BPCI participation was not associated with significant changes in mean episode spending for

EXHIBIT 2

Sample and patient characteristics, by hospital participation in the Bundled Payments for Care Improvement (BPCI) initiative and program period

	Non-BPCI patients			BPCI patients		
Characteristic	**Pre-BPCI**	**Early BPCI**	**Late BPCI**	**Pre-BPCI**	**Early BPCI**	**Late BPCI**
SAMPLE						
Markets	98	98	98	123	123	123
Hospitals	244	244	241	244	244	244
Beneficiaries	20,497	19,734	11,876	75,614	75,297	54,744
PATIENTS						
Mean age (years)	73.4	73.4	73.0	73.3**	73.0**	73.0**
Black race (%)[a]	8.7	8.3	7.9	7.6	7.2	7.1
Female (%)	64.4	63.4	62.2	66.1****	64.2****	64.2****
Dually eligible (%)[b]	13.3	12.5	11.8	15.5**	14.2**	13.5**
Mean Elixhauser comorbidity index[c,d]	5.5****	5.3****	4.4****	5.4****	4.8****	4.3****
Prior acute care hospital use (%)[d]	18.9***	17.5***	16.3***	17.9****	15.7****	15.3****
Prior IRF use (%)[d]	1.3	1.6	1.4	1.4	1.2	1.2
Prior SNF use (%)[d]	5.2	5.3	4.9	5.1**	4.7**	4.6**

SOURCE Authors' analysis of Medicare claims data for 2011–16. **NOTES** Patient characteristics are those of people who received lower extremity joint replacement (LEJR), drawn from a 20 percent Medicare claims sample for non-BPCI hospitals and from a 100 percent sample for BPCI hospitals. The pre-BPCI period was January 2011–September 2013. The early BPCI period was October 2013–June 2015, and the late BPCI period was July 2015–December 2016 (data through September 2016 were used to define LEJR episodes to allow for a ninety-day postdischarge period between October and December 2016). Wilcoxon rank-sum tests or *t*-tests were used to test the differences in continuous variables, and chi-square tests were used for categorical variables. Appendix exhibit 6 contains a fuller version of this exhibit, with differential changes for BPCI and non-BPCI patients (see note 22 in text). IRF is inpatient rehabilitation facility. SNF is skilled nursing facility. [a]Race was broken out as black versus others because of existing disparities in access to LEJR among black patients specifically. [b]Eligible for Medicare and Medicaid (as an indicator of low socioeconomic status). [c]The Elixhauser comorbidity score, an index of severity, ranges from −32 to 92, with increasing scores highly correlated with increased probability of in-hospital death. [d]Calculated using data from the year before LEJR hospitalization. **$p < 0.05$ ***$p < 0.01$ ****$p < 0.001$

the late BPCI period.

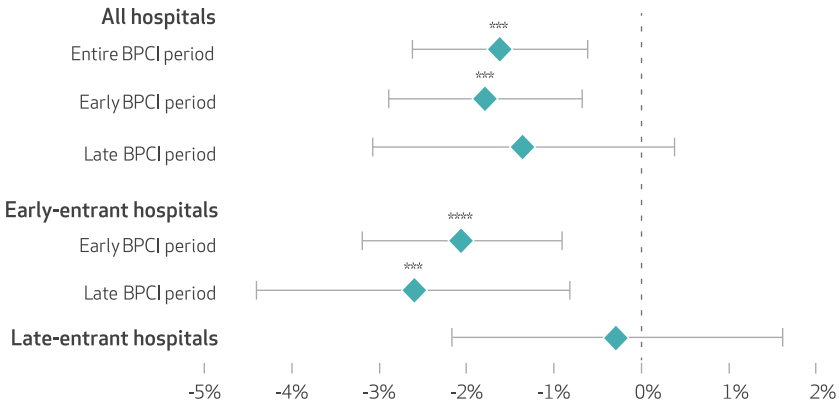
In a secondary analysis, BPCI participation was associated with decreased episode spending for early-entrant hospitals in both the early and late BPCI periods (−2.1 percent and −2.6 percent differential changes, respectively). In contrast, participation was not associated with changes in episode spending among late-entrant hospitals (−0.3 percent differential change).

EPISODE SPENDING COMPONENTS The raw values of several spending components varied by study period (appendix exhibit 9).[22] In adjusted analyses BPCI participation was associated with a differential decrease in episode spending attributable to SNF care (−0.4 percent; $p = 0.02$) and IRF care (−0.7 percent; $p < 0.001$) for BPCI versus non-BPCI patients (appendix exhibit 10).[22] In contrast, spending attributable to index hospitalizations (0.7 percent; $p = 0.001$), home health agency services (0.2 percent; $p = 0.02$), and professional services (0.2 percent; $p = 0.03$) increased differentially for BPCI patients. These associations were generally larger for early versus late BPCI periods and early- versus late-entrant hospitals.

QUALITY OUTCOMES We observed secular trends for a number of quality measures among both BPCI and non-BPCI patients (appendix ex-

EXHIBIT 3

Percent changes in episode spending associated with hospital participation in the Bundled Payments for Care Improvement (BPCI) initiative, by program period and timing of participation



SOURCE Authors' analysis of Medicare claims data for 2011–16. **NOTES** This exhibit shows the results from a series of difference-in-differences models that used an instrumental variable to evaluate the association between participation in BPCI and differential changes in mean episode spending. Separate models were used to evaluate associations between BPCI participation and changes in mean episode spending for the overall cohort and study period (that is, all hospitals and the entire BPCI period, October 2013–December 2016), all hospitals by program period (early versus late BPCI period, defined in the notes to exhibit 2), early-entrant hospitals (those that began participating in the early BPCI period) by program period, and late-entrant hospitals (those that began participating in the late BPCI period) in the late BPCI period. All spending estimates were standardized and transformed into 2016 dollars. Negative estimates indicate savings. The error bars indicate 95% confidence intervals. ***$p < 0.01$ ****$p < 0.001$

hibit 8).[22] In adjusted analyses there were no significant differential changes between BPCI and non-BPCI patients in ninety-day risk-standardized mortality (−0.15 percent; $p = 0.35$), unplanned readmission (0.15 percent; $p = 0.67$), emergency department visit (−0.19 percent; $p = 0.63$), or LEJR-specific complication (0.12 percent; $p = 0.59$) rates (exhibit 4). BPCI participation was not associated with changes in quality outcomes when evaluated specifically by program period or duration of hospital participation.

SENSITIVITY ANALYSES When we extended the approach from prior CMS evaluations, we found that BPCI participation was associated with a 2.2 percent differential decrease in mean episode spending (appendix exhibit 12),[22] a value that was 27 percent lower than estimates from our main analyses (1.6 percent differential decrease in mean episode spending). Analyses that excluded data for January–September 2013, used less stringent propensity matching, and removed hospitals that participated in the Comprehensive Care for Joint Replacement model yielded generally similar results (appendix exhibits 13–15).[22]

Discussion

In this long-term study, participation in the Bundled Payments for Care Improvement initiative was associated with a 1.6 percent decrease in average lower extremity joint replacement episode spending, driven by the performance of early participants, with no changes in quality. This finding has five important implications in light of CMS's decision to continue implementing voluntary bundles through BPCI Advanced.

First, early reductions in episode spending accounted for the savings observed over three years under voluntary LEJR bundled payment. Along with early work that evaluated the impact of mandatory CMS bundled payment programs for LEJR,[11,28] our findings fill an important knowledge gap about bundled payment policy. In particular, this study provides the first rigorous evidence about savings from voluntary LEJR bundled payment that, albeit lower than the 3.9 percent savings demonstrated in early program evaluations and 21 percent savings achieved by high performers,[13] were nonetheless sustained through three years. Longer-term savings support Medicare's decision to expand voluntary bundles through BPCI Advanced.

Second, early 1.8 percent savings came from early entrants that sustained savings over time. These results suggest that observed decreases in SNF and IRF spending continued in the late BPCI period and align with evidence from other value-based payment arrangements such as accountable care organizations[29] and the Alternative Quality Contract[30] that organizations can achieve savings over longer time periods.

Third, despite sustained savings among early entrants, the lack of observed overall LEJR episode savings later in BPCI suggests that savings might not be generalizable across all participants. This may be because of differences between hospitals in the initial and later participation waves. Later participants may have been less able to influence episode spending by systematically changing discharge practices to institutional postacute care or driving practice improvement in activities that require greater time and resource investment (for example, the establishment of clinical protocols for handling potential complications among preferred SNFs). While this study did not evaluate strategies employed by participants, the results collectively suggest that opportunities to achieve savings vary by

EXHIBIT 4

Changes in quality outcomes associated with hospital participation in the Bundled Payments for Care Improvement (BPCI) initiative, by program period and timing of participation

		Program period		Timing of participation		
				Early-entrant hospitals		Late-entrant hospitals
Outcome	Overall	Early BPCI	Late BPCI	Early BPCI	Late BPCI	Late BPCI
Mortality rate	−0.15%	−0.03%	−0.34%	0.01%	−0.27%	−0.37%
Unplanned readmission rate	0.15	0.21	0.03	0.24	−0.10	0.18
ED visit rate	−0.19	−0.22	−0.08	−0.25	0.15	0.54
LEJR-specific complication rate[a]	0.12	0.21	−0.03	0.24	0.003	−0.09

SOURCE Authors' analysis of Medicare claims data for 2011–16. **NOTES** The exhibit shows results from difference-in-differences models that evaluated the association between participation in BPCI and differential changes in quality outcomes, with changes displayed for the overall cohort and intervention period (entire BPCI period) as well as for program period (early and late BPCI, defined in the notes to exhibit 2) and timing of participation (early- and late-entrant hospitals, defined in the notes to exhibit 3). Negative estimates indicate reductions in rates (that is, quality improvements). Emergency department (ED) visits are those without a hospitalization. LEJR is lower extremity joint replacement. Appendix exhibit 11 contains a fuller version of this exhibit (see note 22 in text). [a]Defined by Hospital Compare.

timing of participation without diminishing over time for those able to achieve them—critical insights, given the two participation waves in BPCI Advanced.

Fourth, the 1.6 percent episode savings observed in our primary instrumental variable analyses and the 2.2 percent episode savings observed in our non–instrumental variable sensitivity analyses suggest that while patient selection exists in LEJR bundled payment, it does not fully account for associated savings. The fact that both estimates were smaller than those reported in prior peer-reviewed (−3.8 percent) and CMS-contracted (−3.9 percent) analyses of BPCI over one to three years highlights the ways in which our analysis differs methodologically from those evaluations.[1,4] For instance, CMS contractor evaluations did not account for time-varying entry into BPCI or unobserved differences in characteristics between BPCI and non-BPCI hospitals (for example, they did not include hospital fixed effects). Additional reasons why those results vary in magnitude from ours may include differences in study samples (for example, we excluded episodes covered by the Comprehensive Care for Joint Replacement model) and variables (for example, we incorporated hospital characteristics from the American Hospital Association).

A comparative strength of our analytic approach and use of the instrumental variable is that it mitigated the effects of changes in unobserved patient characteristics. In particular, our instrumental variable analysis yielded estimates that were 27 percent lower than those of our sensitivity analysis without the variable (1.6 percent versus 2.2 percent). While more work is needed to ensure that policy evaluations account for unobserved selection, our results provide early estimates that reflect the fact that even in the absence of observable selection (that is, selection based on characteristics captured in our data), unobserved selection can still occur. For instance, while physicians and hospitals might not select patients for LEJR based on sex or clinical comorbidities, as observed in our analysis, they might still preferentially avoid patients perceived to have poor social support or low treatment adherence. Notably, estimates from our analysis were similar in magnitude to those from evaluations of Medicare's mandatory joint replacement program, which are not as susceptible to patient selection.[28,31]

Taken together, our results corroborate the presence of unobserved patient selection in voluntary bundled payment and demonstrate associated savings in spite of it. These are important insights as Medicare continues to emphasize voluntary bundled payment through BPCI Advanced, whose results are far more likely to mirror those of BPCI than those of mandatory programs—because of organizations' ability to choose whether and when to participate.

Fifth, several measures of quality did not appear to change under LEJR bundled payment, even among hospitals that participated in BPCI for close to two years. On the one hand, these findings indicate that despite generating financial savings, early participants were unable to redesign practice or coordinate care in ways that reduced mortality, complication, readmission, or emergency department visit rates. On the other hand, this analysis suggests that financial savings did not appear to come at the expense of quality. Future work should evaluate other measures of quality, such as patient-reported outcomes, that might not appear in administrative data.

Conclusion

Participation in bundled payment for lower extremity joint replacement under the Bundled Payments for Care Improvement initiative was associated with episode savings over three years, though savings were driven by early participants, and there were no associated changes in quality. These results have important policy implications as CMS continues expanding voluntary bundled payment via BPCI Advanced. ■

Earlier versions of the analyses reported in this article were presented at the AcademyHealth Annual Research Meeting in Washington, D.C., June 5, 2019, and at the Society of General Internal Medicine Annual Meeting in Washington, D.C., May 9, 2019. This work was funded by the Commonwealth Fund and the Leonard Davis Institute of Health Economics at the University of Pennsylvania. Amol Navathe received personal fees from Navvis Healthcare Inc., the National University Health System of Singapore, and Agathos Inc.; personal fees and equity from NavaHealth; personal fees for service as a commissioner of the Medicare Payment Advisory Commission; speaking fees from the Cleveland Clinic; and an honorarium from Elsevier Press, none of which were related to this article. He also serves without compensation as a board member of Integrated Services Inc. The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government.

NOTES

1 Centers for Medicare and Medicaid Services. Bundled Payments for Care Improvement (BPCI) initiative: general information [Internet]. Baltimore (MD): CMS; [last updated 2019 Apr 17; cited 2019 Sep 23]. Available from: https://innovation.cms.gov/initiatives/bundled-payments/

2 Dummit LA, Kahvecioglu D, Marrufo G, Rajkumar R, Marshall J, Tan E, et al. Association between hospital participation in a Medicare bundled payment initiative and payments and quality outcomes for lower extremity joint replacement episodes. JAMA. 2016;316(12):1267–78.

3 Lewin Group. CMS Bundled Payments for Care Improvement (BPCI) initiative models 2–4: year 2 evaluation and monitoring annual report [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; 2016 Aug [cited 2019 Sep 23]. Available from: https://innovation.cms.gov/Files/reports/bpci-models2-4-yr2evalrpt.pdf

4 Lewin Group. CMS Bundled Payments for Care Improvement (BPCI) initiative models 2–4: year 5 evaluation and monitoring annual report [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; 2018 Oct [cited 2019 Sep 23]. Available from: https://downloads.cms.gov/files/cmmi/bpci-models2-4-yr5evalrpt.pdf

5 CMS.gov. BPCI Advanced [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; [last updated 2019 Sep 16; cited 2019 Sep 23]. Available from: https://innovation.cms.gov/initiatives/bpci-advanced

6 Bleser WK, Muhlestein D, Saunders RS, McClellan MB. Half a decade in, Medicare accountable care organizations are generating net savings: part 1. Health Affairs Blog [blog on the Internet]. 2018 Sep 20 [cited 2019 Sep 23]. Available from: https://www.healthaffairs.org/do/10.1377/hblog20180918.957502/full/

7 Muhlestein D, Saunders RS, Richards R, McClellan MB. Recent progress in the value journey: growth of ACOs and value-based payment models in 2018. Health Affairs Blog [blog on the Internet]. 2018 Aug 14 [cited 2019 Sep 23]. Available from: https://www.healthaffairs.org/do/10.1377/hblog20180810.481968/full/

8 Navathe AS, Liao JM, Dykstra SE, Wang E, Lyon ZM, Shah Y, et al. Association of hospital participation in a Medicare bundled payment program with volume and case mix of lower extremity joint replacement episodes. JAMA. 2018;320(9):901–10.

9 Dartmouth Institute. Dartmouth atlas of health care, data by region [Internet]. Lebanon (NH): Trustees of Dartmouth College; [cited 2019 Dec 9]. Available from: https://atlasdata.dartmouth.edu/static/supp_research_data/#crosswalks

10 Navathe AS, Liao JM, Polsky D, Shah Y, Huang Q, Zhu J, et al. Comparison of hospitals participating in Medicare's voluntary and mandatory orthopedic bundle programs. Health Aff (Millwood). 2018;37(6):854–63.

11 Navathe AS, Liao JM, Shah Y, Lyon Z, Chatterjee P, Polsky D, et al. Characteristics of hospitals earning savings in the first year of mandatory bundled payment for hip and knee surgery. JAMA. 2018;319(9):930–2.

12 Liao JM, Emanuel EJ, Polsky DE, Huang Q, Shah Y, Zhu J, et al. National representativeness of hospitals and markets in Medicare's mandatory bundled payment program. Health Aff (Millwood). 2019;38(1):44–53.

13 Navathe AS, Troxel AB, Liao JM, Nan N, Zhu J, Zhong W, et al. Cost of joint replacement using bundled payment models. JAMA Intern Med. 2017;177(2):214–22.

14 Van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care. 2009;47(6):626–33.

15 Liao JM, Holdofski A, Whittington GL, Zucker M, Viroslav S, Fox DL, et al. Baptist Health System: succeeding in bundled payments through behavioral principles. Healthc (Amst). 2017;5(3):136–40.

16 CMS.gov. Comprehensive Care for Joint Replacement model [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; [last updated 2019 Sep 18; cited 2019 Sep 23]. Available from: https://innovation.cms.gov/initiatives/CJR

17 Tsai TC, Joynt KE, Wild RC, Orav EJ, Jha AK. Medicare's bundled payment initiative: most hospitals are focused on a few high-volume conditions. Health Aff (Millwood). 2015;34(3):371–80.

18 Joynt KE, Gawande AA, Orav EJ, Jha AK. Contribution of preventable acute care spending to total spending for high-cost Medicare patients. JAMA. 2013;309(24):2572–8.

19 Medicare.gov. Hospital Compare datasets [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; [last updated 2019 Oct 30; cited 2019 Nov 19]. Available from: https://data.medicare.gov/data/hospital-compare

20 Meyer BD. Natural and quasi-experiments in economics. J Bus Econ Stat. 1995;13(2):151–61.

21 Card D, Krueger A. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. Am Econ Rev. 1994;84(4):772–93.

22 To access the Appendix, click on the Details tab of the article online.

23 Angrist JD, Pischke J-S. Mostly harmless econometrics: an empiricist's companion. Princeton (NJ): Princeton University Press; 2009.

24 Woolridge JM. Econometric analysis of cross section and panel data. Cambridge (MA): MIT Press; 2002.

25 Alexander D. How do doctors respond to incentives? Unintended consequences of paying doctors to reduce costs [dissertation]. Princeton (NJ): Princeton University; 2006.

26 MacKinnon JG, White H. Some heteroskedasticity-consistnet covariance matrix estimators with improved finite sample properties. J Econom. 1985;29(3):305–25.

27 Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. Participation and dropout in the Bundled Payments for Care Improvement initiative. JAMA. 2018;319(2):191–3.

28 Finkelstein A, Ji Y, Mahoney N, Skinner J. Mandatory Medicare bundled payment program for lower extremity joint replacement and discharge to institutional postacute care: interim analysis of the first year of a 5-year randomized trial. JAMA. 2018;320(9):892–900.

29 McWilliams JM, Hatfield LA, Landon BE, Hamed P, Chernew ME. Medicare spending after 3 years of the Medicare Shared Savings Program. N Engl J Med. 2018;379(12):1139–49.

30 Song Z, Safran DG, Landon BE, Landrum MB, He Y, Mechanic RE, et al. The "Alternative Quality Contract," based on a global budget, lowered medical spending and improved quality. Health Aff (Millwood). 2012;31(8):1885–94.

31 Barnett ML, Wilcock A, McWilliams JM, Epstein AM, Joynt Maddox KE, Orav EJ, et al. Two-year evaluation of mandatory bundled payments for joint replacement. N Engl J Med. 2019;380(3):252–62.

Navathe AS, Emanuel EJ, Venkataramani AS, Huang Q, Gupta A, Dinh CT, et al.
Spending and quality after three years of Medicare's voluntary bundled payment for joint replacement surgery.
Health Aff (Millwood). 2020;39(1).

Appendix

Contents

Appendix Exhibit 11. Changes in quality outcomes associated with BPCI participation by program period and timing of hospital participation, 2012-2016 (Full Table)

Appendix Exhibit 12. Sensitivity analysis extending methods from prior CMS evaluations

Appendix Exhibit 13. Sensitivity analysis excluding January to September 2013 from the baseline period

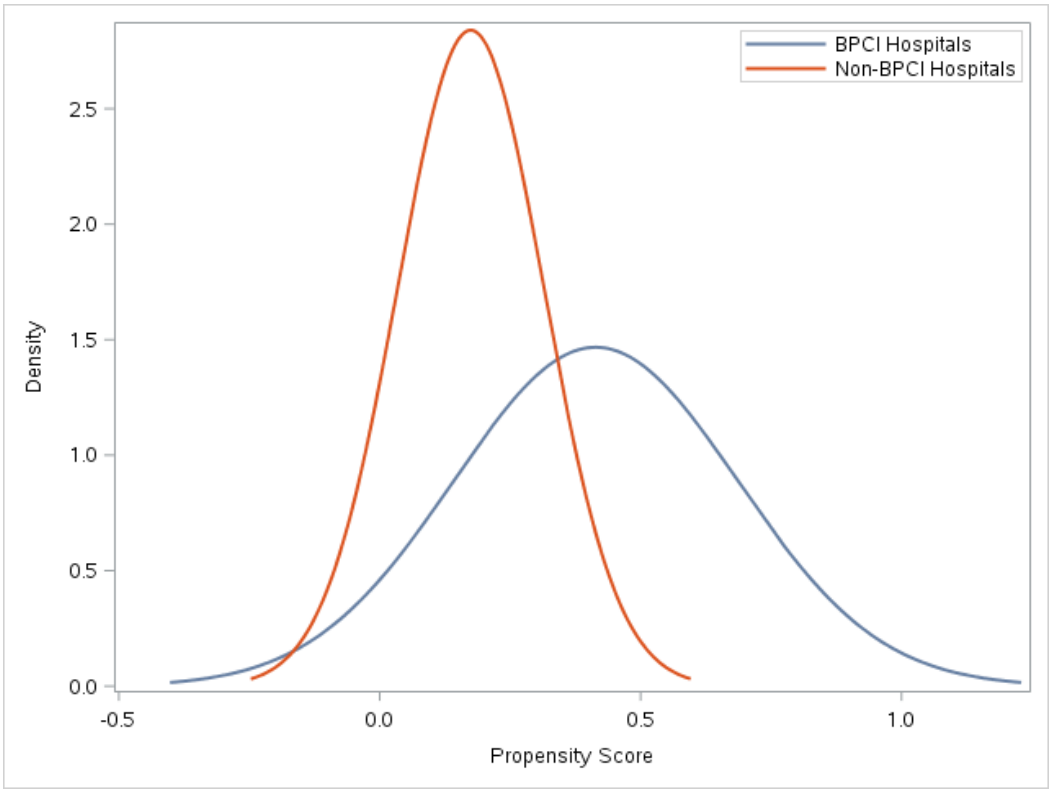
Appendix Exhibit 14. Sensitivity analysis using less stringent propensity matching

Appendix Exhibit 15. Sensitivity analysis excluding CJR hospitals from propensity score matching and analyses

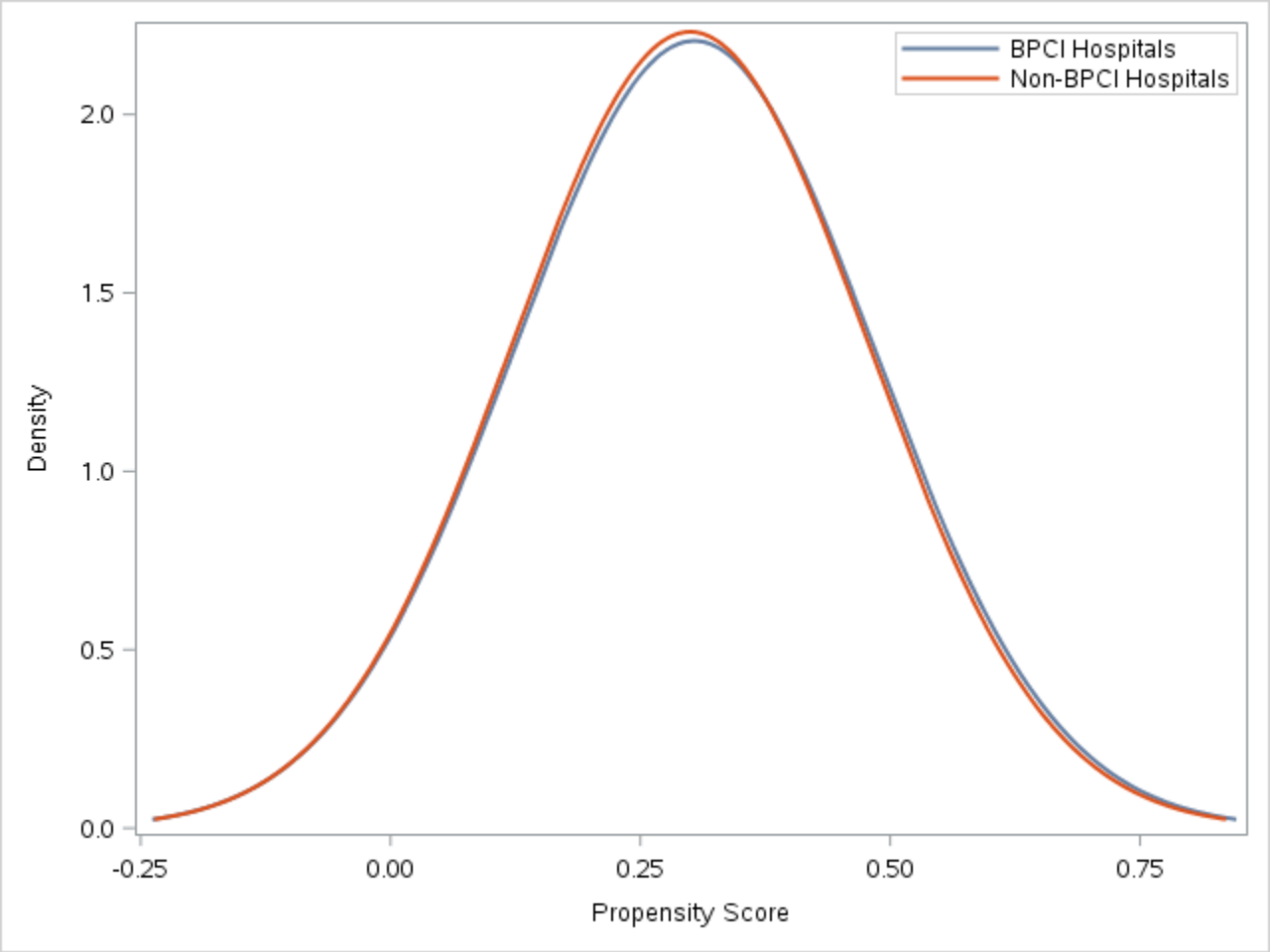
Appendix Methods 1. Propensity score matching of BPCI hospitals to Non-BPCI hospitals

To decrease differences between BPCI and Non-BPCI hospitals, we used propensity score matching on a variables drawn from Medicare claims and AHA Annual Survey data (see Appendix Exhibit 3 below for all the variables). Hospitals not participating in BPCI but located in BPCI markets could have been affected by program (e.g., from changes in market share or referral patterns after BPCI began). Thus, they were not used in the Non-BPCI Hospital group throughout our analyses to avoid potential confounding. Our propensity model used logistic regression with a binary dependent variable of participation in BPCI during any quarter in the BPCI period for the LEJR episode and independent variables listed in Appendix Exhibit 1. Using a minimum of a 2 digit propensity score match, we were able to match 244 BPCI hospitals with 244 Non-BPCI hospitals in Non-BPCI markets in a 1:1 match.[1] Comparison of hospital characteristics between Non-BPCI and BPCI hospitals before and after matching is presented in Appendix Exhibit 1 below. As a sensitivity analysis, we used a less stringent optimal propensity match method by varying the match ratio from 1:1 to 1:3 and using a caliper=0.05 to match 225 BPCI with 628 Non-BPCI hospitals.[2]

Density plot: propensity score distribution of all BPCI and all Non-BPCI Hospitals in Non-BPCI Markets, before matching

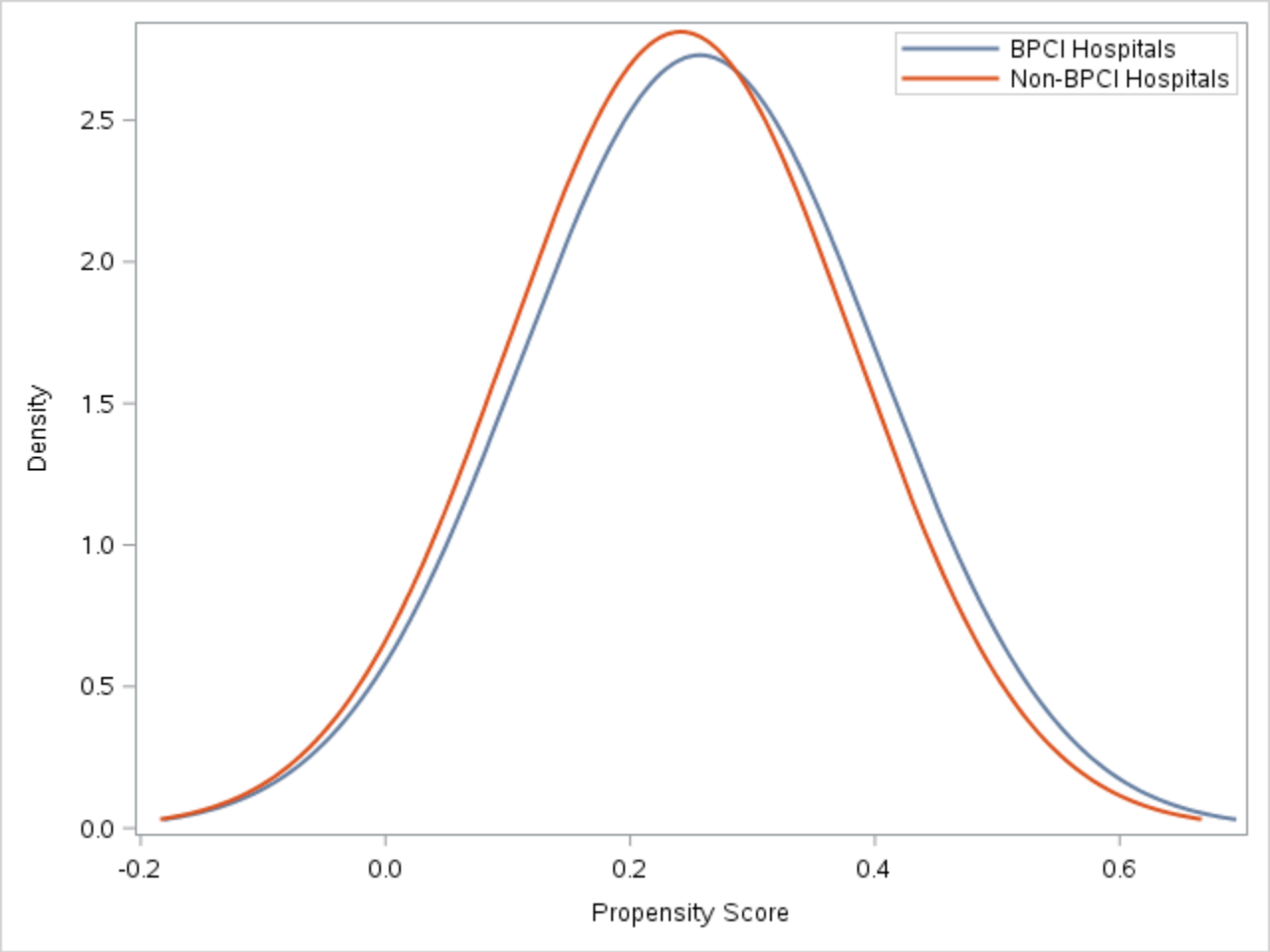


Density plot: propensity score distribution of BPCI (n=244) and Non-BPCI (n=244) Hospitals, after matching



These density plot figures illustrate the area of common support and the distribution of propensity scores before and after matching using a minimum of a 2 digit propensity score match in a 1:1 match.

Density plot: less stringent propensity score distribution of BPCI (n=225) and Non-BPCI (n=628) Hospitals, after matching



These density plot figures illustrate the area of common support and the distribution of propensity scores before and after matching using a less stringent optimal propensity match method by varying the match ratio from 1:1 to 1:3 and using a caliper=0.05. This resulted in 225 BPCI hospitals matched to 628 Non-BPCI Hospitals.

Appendix Exhibit 1. Comparison of baseline characteristics of BPCI hospitals and non-BPCI hospitals before and after propensity matching, 2011

Patient Characteristics						
	Before Propensity Matching			**After Propensity Matching**		
	Non-BPCI Hospitals	**BPCI Hospitals**	**Standardized Difference**	**Non-BPCI Hospitals**	**BPCI Hospitals**	**Standardized Difference**
Annual admissions for top 10 BPCI episodes, mean % (SD)[a]	24.8 (6.7)	22.1 (5.0)	-0.46	22.7 (5.2)	22.6 (5.0)	-0.03
Annual admissions for LEJR, median (IQR)[b]	118 (55 to 229)	175 (94 to 296)	0.39	170 (90 to 314)	179 (96 to 293)	0.01
Proportion of discharges to highest-volume SNF, mean % (SD)	35.7 (19.5)	28.2 (17.0)	-0.41	28.9 (14.8)	29.5 (17.6)	0.04
Proportion of discharges to highest-volume IRF, median % (IQR)[b]	50.0 (0.0 to 100)	80.0 (0.0 to 99.6)	0.22	90.0 (0.0 to 100.0)	80.0 (0.0 to 100.0)	-0.10
90-day readmission rate, median (IQR)[b]	10.5 (5.7 to 16.7)	11.3 (9.1 to 15.3)	0.25	7.1 (16.3)	11.0 (9.0 to 15.0)	0.12
90-day LEJR episode spending, median $ (IQR)[b]	22,996 (20,134 to 26,813)	24,606 (22,343 to 27,523)	0.39	23,936 (21,092 to 27,930)	24,355 (21,977 to 27,309)	0.09
Hospital Characteristics						
	Before Propensity Matching			**After Propensity Matching**		
	Non-BPCI Hospitals	**BPCI Hospitals**	**Standardized Difference**	**Non-BPCI Hospitals**	**BPCI Hospitals**	**Standardized Difference**
Number of beds, median (IQR)[b]	176 (94 to 313)	268 (164 to 402)	0.53	235 (148 to 401)	261 (154 to 391)	0.06
Ownership status, %						
For-profit	19.0	18.2	0.36	17.2	16.4	0.05
Not-for-profit	66.7	78.0		77.1	78.7	
Government	14.4	3.8		5.7	4.9	
Member of a system, %	63.8	80.2	0.37	79.1	78.7	-0.01
Teaching status, %[c]						
Major teaching	7.9	14.7	0.3	13.5	12.7	0.05
Minor teaching	30.1	36.1		31.6	34.4	
Non-teaching	62.0	49.2		54.9	52.9	

Intern and resident to bed ratio, median (IQR)[b]	0.0 (0.0 to 0.03)	0.0 (0.0 to 0.08)	0.33	0.0 (0.0 to 0.06)	0.0 (0.0 to 0.06)	0.08
Disproportionate share hospital payments, median $ (IQR)[d]	1,464,209 (422,560 to 4,464,053)	2,723,772 (521,608 to 7,198,542)	0.22	2,162,258 (390,499 to 6,420,029)	2,472,058 (492,948 to 6,560,881)	0.02
Urban status, %	94.4	99.4	0.29	99.2	99.2	0
Medicare days as % of total patient days, mean (SD)	52.0 (12.8)	51.7 (10.7)	-0.03	52.9 (11.8)	52.0 (10.7)	0.07
Market share, median % (IQR)[b]	7.7 (3.2 to 19.6)	7.9 (3.4 to 18.4)	-0.004	9.2 (3.7 to 23.8)	9.2 (4.4 to 19.7)	0.03
Market characteristics						
	Before Propensity Matching			**After Propensity Matching**		
Total number of beneficiaries, median (IQR)[b]	821,288 (449,198 to 1,523,905)	1,617,708 (796,342 to 2,765,863)	0.61	1,068,113 (591,261 to 2,695,204)	1,321,591 (678,746 to 2,075,609)	0.14
Median income, $ (SD)	48,143 (11,293)	56,104 (16,608)	0.56	52,303 (14,697)	53,089 (14,543)	0.05
SNF beds, median (IQR)[b]	5,026 (2,752 to 9,607)	7,652 (4,043 to 15,361)	0.43	6,124 (3,420 to 13,670)	6,308 (3,578 to 9,646)	-0.01
IRF beds, median (IQR)[b]	97 (54 to 163)	180 (82 to 388)	0.57	122 (64 to 270)	134 (60 to 253)	0.05
MA penetration, % (SD)	24.2 (13.7)	24.8 (11.5)	0.05	24.7 (15.1)	25.4 (11.5)	0.05
Hospital concentration, HHI (SD)	2,471 (1,762)	1,990 (1,564)	-0.29	2,216 (1,735)	2,181 (1,653)	-0.02

[a]Major joint replacement of the lower extremity, Double joint replacement of the lower extremity, Revision of the hip or knee, Hip and femur procedures except major joint, Lower extremity and humerus procedure except hip, foot, and femur, Coronary artery bypass graft, Acute myocardial infarction, Congestive heart failure, Simple pneumonia and respiratory infections, Chronic obstructive pulmonary disease, bronchitis/asthma. [b]Median (IQR) provided where data are skewed. [c]From the AHA Annual Survey, major teaching hospitals are those that are members of the Council of Teaching Hospitals (COTH), minor teaching hospitals are non-COTH members that had a medical school affiliation reported to the American Medical Association, and nonteaching hospitals are all other institutions. [d]Disproportionate share hospital payment percentage derived from the FY2017 CMS IMPACT file.

Appendix Exhibit 2. Clinical conditions used in statistical models

Acquired immune deficiency syndrome
Alcohol abuse
Chronic blood loss anemia
Chronic pulmonary disease
Coagulopathy
Congestive heart failure
Deficiency anemias
Depression
Diabetes with chronic complications
Diabetes without chronic complications
Drug abuse
Fluid and electrolyte disorders
Hypertension with complication
Hypothyroidism
Liver disease
Lymphoma
Metastatic cancer
Obesity
Other neurological disorders
Paralysis
Peptic ulcer disease excluding bleeding
Peripheral vascular disease
Psychoses
Pulmonary circulation disease
Renal failure
Rheumatoid arthritis/collagen vascular disease
Solid tumor without metastasis
Valvular disease
Weight loss

Appendix Methods 2. Instrumental variable approach

We introduce a new instrumental variable (IV) to mitigate confounding from selection of beneficiaries for hospitalization based on unobservable characteristics. This IV is an adaptation of instrumental variables from outside of healthcare.[3-4] Our approach uses historical hospital referral patterns before the beginning of the BPCI program (i.e., to which hospitals patients were hospitalized for LEJR prior to October 2013) to predict which patients would be hospitalized for LEJR at hospitals later participating in the BPCI program (i.e., which patients would be hospitalized at BPCI Hospitals in the BPCI period based on historical hospitalization patterns prior to the BPCI period). Specifically, we used a set of patient characteristics -- age, race, gender, zip code of residence, clinical conditions (listed in Appendix Exhibit Table 2), MS-DRG, as well as prior hospital, SNF, and IRF use -- from 2011 to predict the probability that a given patient would be hospitalized for LEJR at a hospital that later participates in BPCI (BPCI Hospital).

Importantly, we could not use other standard IVs used in health care services research such as the distance to hospital. This is because many hospital systems in bundled payment exerted substantial effort to attract preferred (i.e., lower risk or healthier) types of patients, many of whom were at greater distance from the hospital. These efforts including buying hospitals in outlying suburban areas, buying physician practices, or contracting with surgeons in areas with a greater density of preferred patients.

Because historical hospital referral patterns for elective conditions such as LEJR are not correlated with changes in patient selection after hospital participation in BPCI (conditional on variables we can observe in our data such as patient demographics), they serve as a

reasonable IV for our analysis. In tests of the instrumental variable with respect to observed treatment at a BPCI hospital, we found a strong association (F-measure of 11,835) between predicted and observed BPCI exposure, controlling for time-varying hospital, market, and patient characteristics and including hospital, market, and time fixed-effects. This confirmed that we had a strong instrument that was uncorrelated with the confounder of unobservable patient selection, but highly predictive of our treatment, i.e., admission to a BPCI Hospital (in the BPCI program period). The instrument was constructed using 2011 data and the formula $BPCI_{ever,h} = Zip_{pt} + Cov_{pt} + \varepsilon$, where $BPCI_{ever,h}$ is an indicator describing whether a Medicare beneficiary was admitted for LEJR at a hospital that later joins BPCI (at any point), Zip_{pt} is the beneficiary's ZIP code of residence, Cov_{pt} is a vector of characteristics of the beneficiary (including demographics and clinical comorbidities) and ε is the error term.

In the first-stage regression, we used pre-BPCI period hospital referral patterns in 2011 to generate the predicted probability of hospitalization for LEJR at an eventual BPCI Hospital ($\bar{\pi}_{pt}$), and used that probability as an instrument for actual hospitalization for LEJR at a BPCI Hospital. The first stage is $BPCI\_observed_{pt,h,t} = \bar{\pi}_{pt} + HRR_{FE} + Time_{FE} + Cov_{pt} + Cov_{HRR} + \varepsilon$, where the observed BPCI status for a beneficiary receiving LEJR (whether the hospital the beneficiary was admitted to participated in BPCI in that market-quarter) was regressed on the IV ($\bar{\pi}_{pt}$, the predicted probability for that beneficiary of going to a BPCI hospital based on historical patterns), market (based on patient ZIP code of residence) and time fixed effects, patient characteristic covariates, market time-varying covariates, and an error term. We use the first stage as part of a differences-in-differences design, but to do so the $\bar{\pi}_{pt}$ is set to 0 in the pre-period (i.e., there is no BPCI treatment before BPCI began). Thus, in the second stage regression, we then used

hospitalization to a BPCI hospital as an instrument for observed "treatment" at a BPCI or Non-BPCI hospital in the period after BPCI began as part of the difference-in-differences design, relating treatment to spending and quality outcomes. $y_{pt,h,t} = \alpha + \beta * BPCI\_observed_{pt,h,t} + \gamma * Hosp_{FE} + HRR_{FE} + Time_{FE} + \delta * Cov_{pt} + \theta * Cov_{HRR} + \varepsilon$, where the coefficient of interest is β and captures the average effect of BPCI on outcome y (note that $BPCI\_observed_{pt,h,t}$ is time-varying and is 1 for BPCI episodes in the post-period only and 0 otherwise, which is similar to the interaction term in a usual difference-in-difference model. Together with hospital and time fixed effects, the coefficient β on $BPCI\_observed_{pt,h,t}$ gives the difference-in-differences estimate of the effect of BPCI on outcome *y*. While we show these equations separately, this was estimated simultaneously using 2 stage least squares (2SLS) and not in 2 steps. This instrumental variable approach allowed us to measure the effect of BPCI among patients who received LEJR at a BPCI Hospital regardless of BPCI's existence.

Notably, because Non-BPCI hospitals were also likely affected by BPCI when in the same market as a BPCI Hospital (a BPCI Market), we only use beneficiaries admitted to Non-BPCI Hospitals (propensity-matched per Appendix A in our primary analysis) located in Non-BPCI Markets as the comparison group. We also conducted a Hausman test evaluating for endogeneity, finding $p<0.001$ and therefore rejecting the null hypothesis of equivalence (i.e., no endogeneity from unobserved confounding). This result provided additional rationale for the need for this IV to mitigate confounding from unobserved selection. We used bootstrapped standard errors to account for the fact that the IV is an estimated quantity.[5]

Finally, in response to reviewer suggestions, we directly examined potential selection. In our sample, 2,096 surgeons (5.1% of all surgeons) operated at least 1 BPCI and 1 Non-BPCI hospital. Further, among the 6,712 surgeons who work at multiple hospitals, the 2,096 surgeons corresponded to the 31.2% who operated at BPCI and Non-BPCI hospitals, as opposed to only BPCI or only Non-BPCI hospitals. These values suggest that while not widespread, the dynamic of operating at multiple hospitals exists in our sample and could give rise to selection.

To further examine and illustrate changes in hospital referral patterns, we evaluated how hospital referral patterns changed with BPCI participation for surgeons working at BPCI and Non-BPCI hospitals. Specifically, for the sample of patients receiving LEJR from surgeons operating at both BPCI and Non-BPCI hospitals, we examined differential changes at BPCI hospitals (those that become a BPCI hospital in the BPCI period) vs. non-BPCI hospitals with respect to the mean predicted probability of admission to a BPCI hospital. This difference-in-differences analysis provides insight about whether, among surgeons who operate at both BPCI and Non-BPCI in both the pre-BPCI and BPCI periods, the probability of referring a patient to a BPCI vs Non-BPCI hospital changes after the surgeon begins participating in the program. We found that there was indeed a differential change in likelihood of operating on patients in BPCI vs Non-BPCI hospitals (diff-in-diff estimate 6.6%, p=0.0127) associated with BPCI participation.

More broadly, there are multiple forms of selection that could occur under BPCI. In particular, hospitals can also ‘sort’ patients and drive selection through practice acquisition, development of clinically integrated networks, hiring physicians, and other activities. To

examine selection more broadly (and in a fashion more germane to BPCI and Medicare vis-à-vis the policy question), we conducted an additional difference-in-differences analysis evaluating the effect of BPCI participation on changes in hospital referral patterns across all episodes in our sample. We observed differential changes in the probability of being admitted to a BPCI hospital coincident with hospital participation in BPCI (diff-in-diff estimate 1.1%, $p<0.0001$).

Taken together, these analyses suggest that while the magnitude of selection is not expansive, it does occur on the margin. Our findings also corroborate that big shifts in referral patterns for large amounts of patients may be difficult and perhaps undesirable for BPCI hospitals. Regardless, these results highlight that while the magnitude of selection is small, it has a significant and outsized impact on BPCI program outcomes.

Appendix Exhibit 3. Covariate balance between Non-BPCI and BPCI hospital groups by values of the instrumental variable, using all Non-BPCI episodes

	Original Observed Data			Quartile 1			Quartile 2			Quartile 3			Quartile 4		
	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD
Episodes, No.	243,880	177,756	NA	161,135	6,685	NA	120,051	47,769	NA	73,244	94,576	NA	36,601	131,219	NA
Age, mean year (SD)	73.0*	73.0*	-0.001	73.0*	72.8*	-0.03	73.2***	72.9***	-0.03	73.3***	73.2***	-0.02	73.0*	73.2*	0.02
Elixhauser co-morbidity index, mean (SD)	4.4***	4.2***	-0.02	4.4*	4.9*	0.05	4.6**	4.8**	0.02	4.8***	4.4***	-0.04	4.6***	4.2***	-0.04
Black, %	5.6***	6.6***	0.04	5.9	5.6	-0.01	4.7**	5.0**	0.02	6.1***	7.0***	0.04	7.4	7.2	-0.01
Female, %	62.8***	63.9***	0.02	63.2	62.5	-0.02	63.0**	62.3**	-0.02	63.9**	64.7**	0.02	63.0***	64.9***	0.04
Dual-eligible, %	11.3***	10.5***	-0.03	12.2*	11.3*	-0.03	11.4**	11.9**	0.02	11.2	11.0	-0.01	11.3***	10.4***	-0.03
Prior acute care hospital use, %	16.3***	14.6***	-0.05	16.5	16.6	0.003	16.6	16.5	-0.002	18.1***	15.5***	-0.07	18.4***	14.5***	-0.11
Prior IRF use, %	1.4***	1.2***	-0.01	1.3	1.4	0.02	1.4	1.4	-0.0002	1.4***	1.2***	-0.02	1.7***	1.3***	-0.03
Prior SNF use, %	4.5***	4.0***	-0.03	4.5***	5.6***	0.05	4.4	4.4	0.001	5.0***	4.2***	-0.04	4.8***	3.8***	-0.05

This table SMD=Standardized Mean Difference. *p<0.05, **p<0.01, ***p<0.001

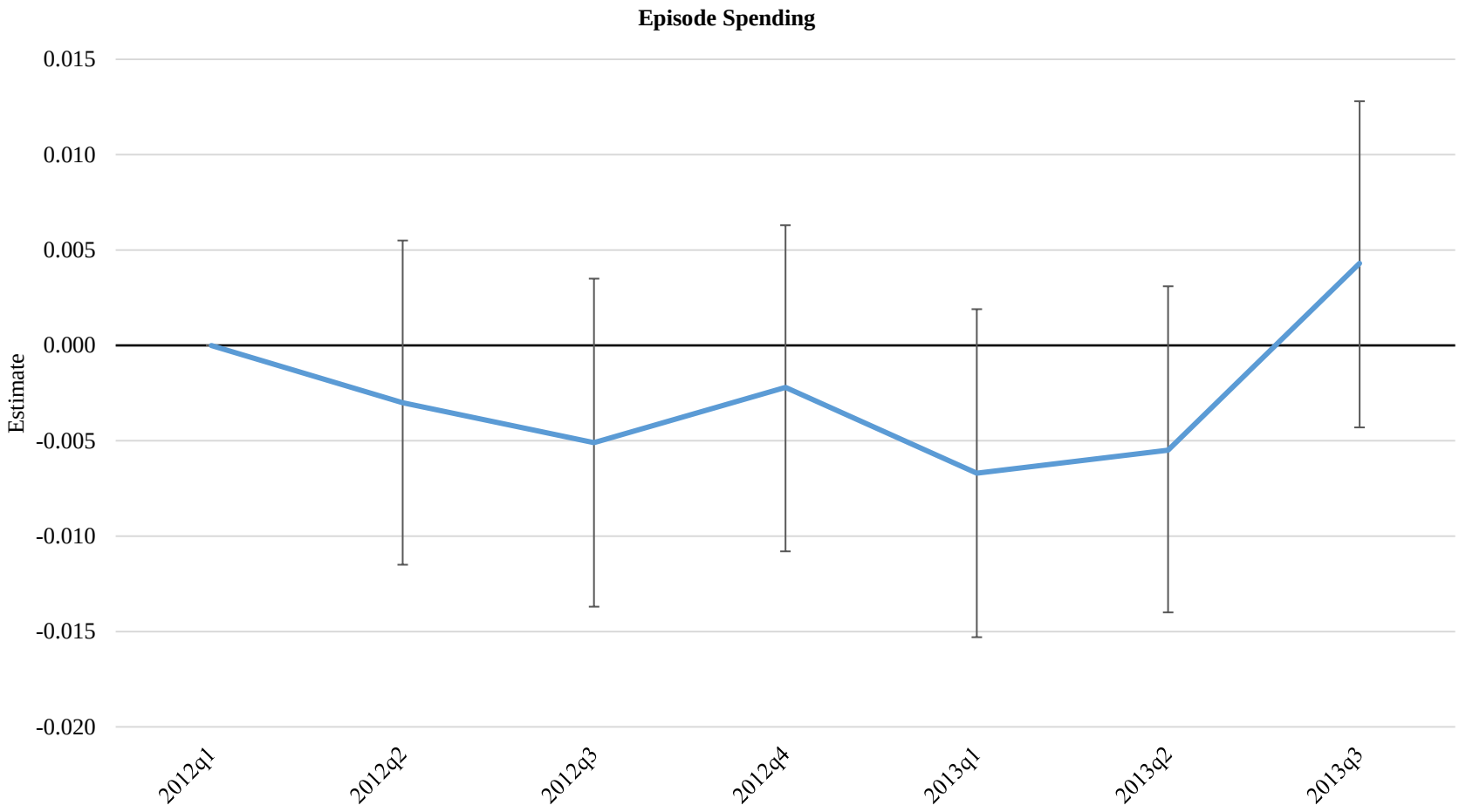
Appendix Exhibit 4. Covariate balance between Non-BPCI and BPCI hospital groups across values of the instrumental variable, using only Non-BPCI episodes in BPCI markets

	Original Observed Data			Quartile 1			Quartile 2			Quartile 3			Quartile 4		
	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD	Non-BPCI	BPCI	SMD
Episodes, No.	131,505	177,756	NA	54,014	23,301	NA	38,309	39,006	NA	25,543	51,773	NA	13,639	63,676	NA
Age, mean year (SD)	73.2***	73.0***	-0.02	73.2***	72.6***	-0.07	73.2***	72.9***	-0.03	73.3*	73.1*	-0.01	73.0	73.1	0.01
Elixhauser co-morbidity index, mean (SD)	4.6***	4.2***	-0.04	4.6	4.6	-0.005	4.5*	4.2*	-0.03	4.8***	4.2***	-0.06	4.5***	4.0***	-0.05
Black, %	5.4***	6.6***	0.05	4.9*	5.4*	0.02	4.5***	5.4***	0.04	7.4	7.6	0.01	6.3***	7.1***	0.03
Female, %	62.9***	63.9***	0.02	62.8***	61.4***	-0.03	62.7*	63.4*	0.01	63.3*	64.2*	0.02	62.4***	64.8***	0.05
Dual-eligible, %	11.4***	10.5***	-0.03	12.3***	11.2***	-0.04	10.2	10.6	0.01	11.9***	11.0***	-0.03	10.3*	9.8*	-0.02
Prior acute care hospital use, %	17.1***	14.6***	-0.07	16.5*	15.9*	-0.02	16.8***	14.7***	-0.06	18.4***	14.9***	-0.09	17.6***	13.7***	-0.11
Prior IRF use, %	1.5***	1.2***	-0.02	1.5	1.4	-0.01	1.4***	1.0***	-0.03	1.5***	1.2***	-0.03	1.8***	1.3***	-0.04
Prior SNF use, %	4.8***	4.0***	-0.04	4.7	4.5	-0.01	4.5***	4.0***	-0.03	5.5***	4.2***	-0.06	4.8***	3.6***	-0.06

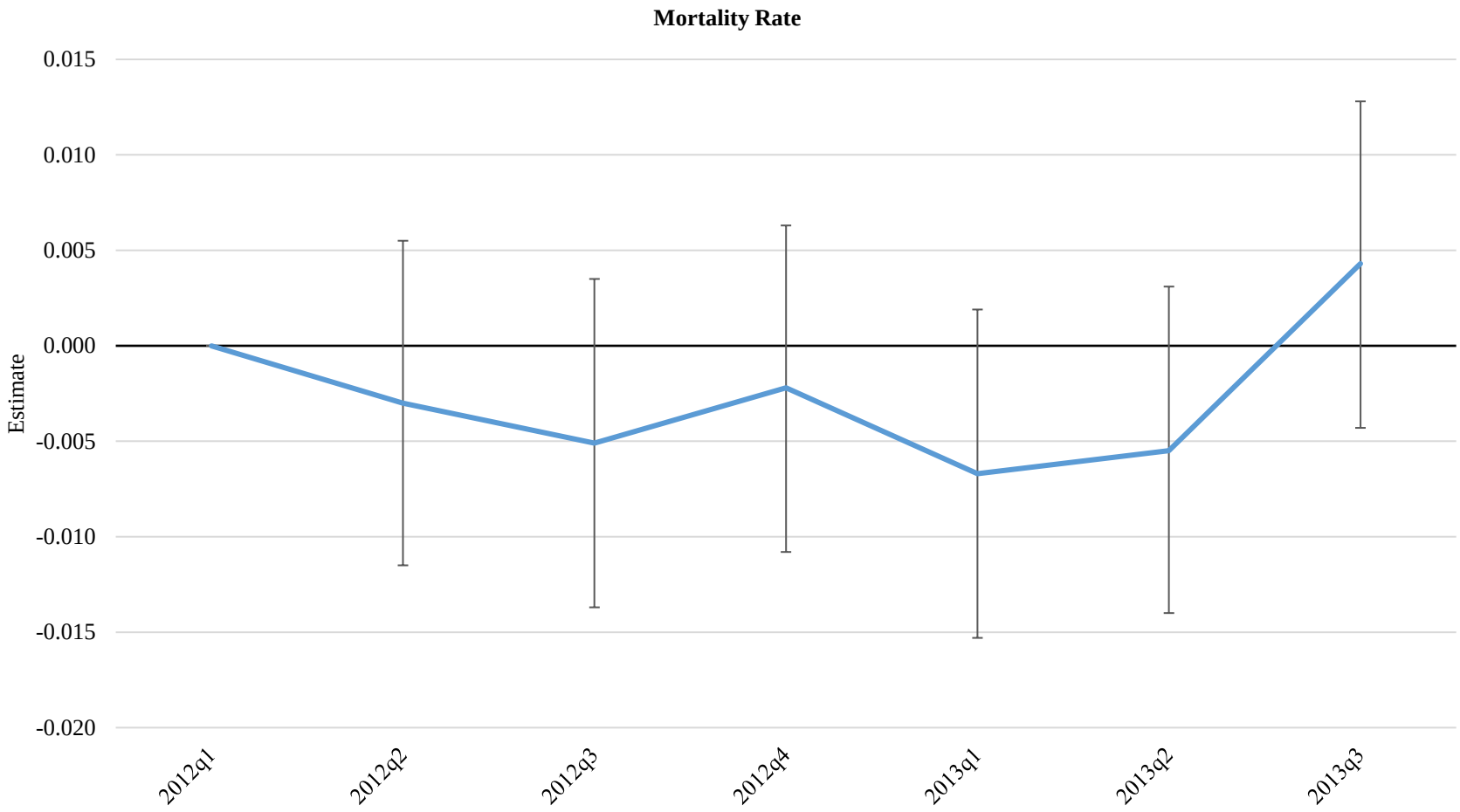
SMD=Standardized Mean Difference. *p<0.05, **p<0.01, ***p<0.001

Appendix Methods 3. Tests of parallel trends between BPCI and Non-BPCI hospitals for spending and quality outcomes

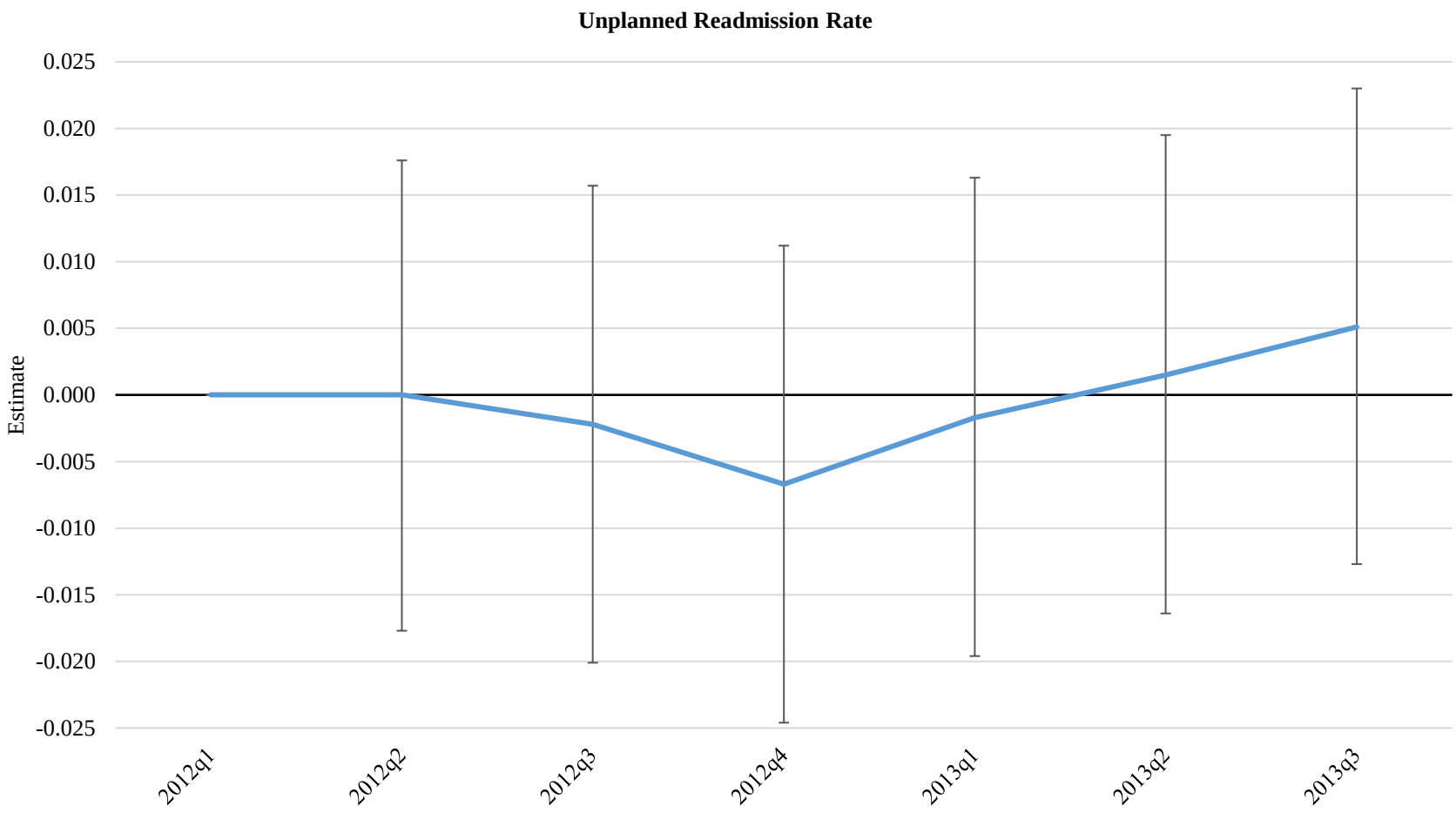
This series of graphs shows the results of generalized linear regression models for each outcome as the dependent variable, and independent variables of a time (quarter) fixed effects, BPCI Hospital indicator variable, and the interaction. The estimates plotted show that the interaction term coefficients are not statistically significant, indicating no divergent trends in the pre-period for any outcome variable. The estimates plotted show that the interaction term coefficients are not statistically significant, indicating a lack of divergent trends in the pre-period for any outcome variable. Similarly, tests conducted using a linear, instead of categorical, time variable demonstrated no statistically significant differential trends.



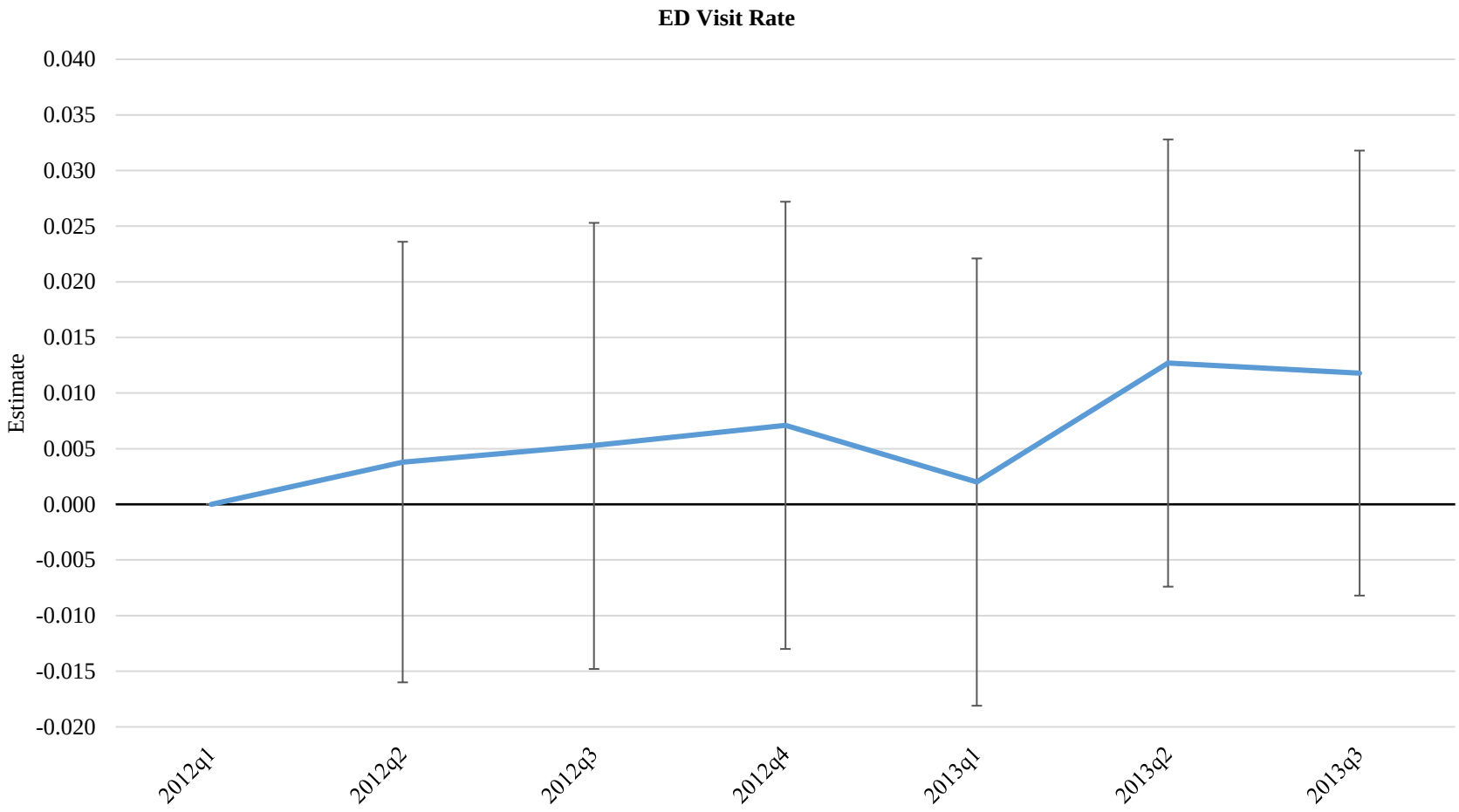
p-values	.	0.85	0.78	0.47	0.62	0.24	0.93



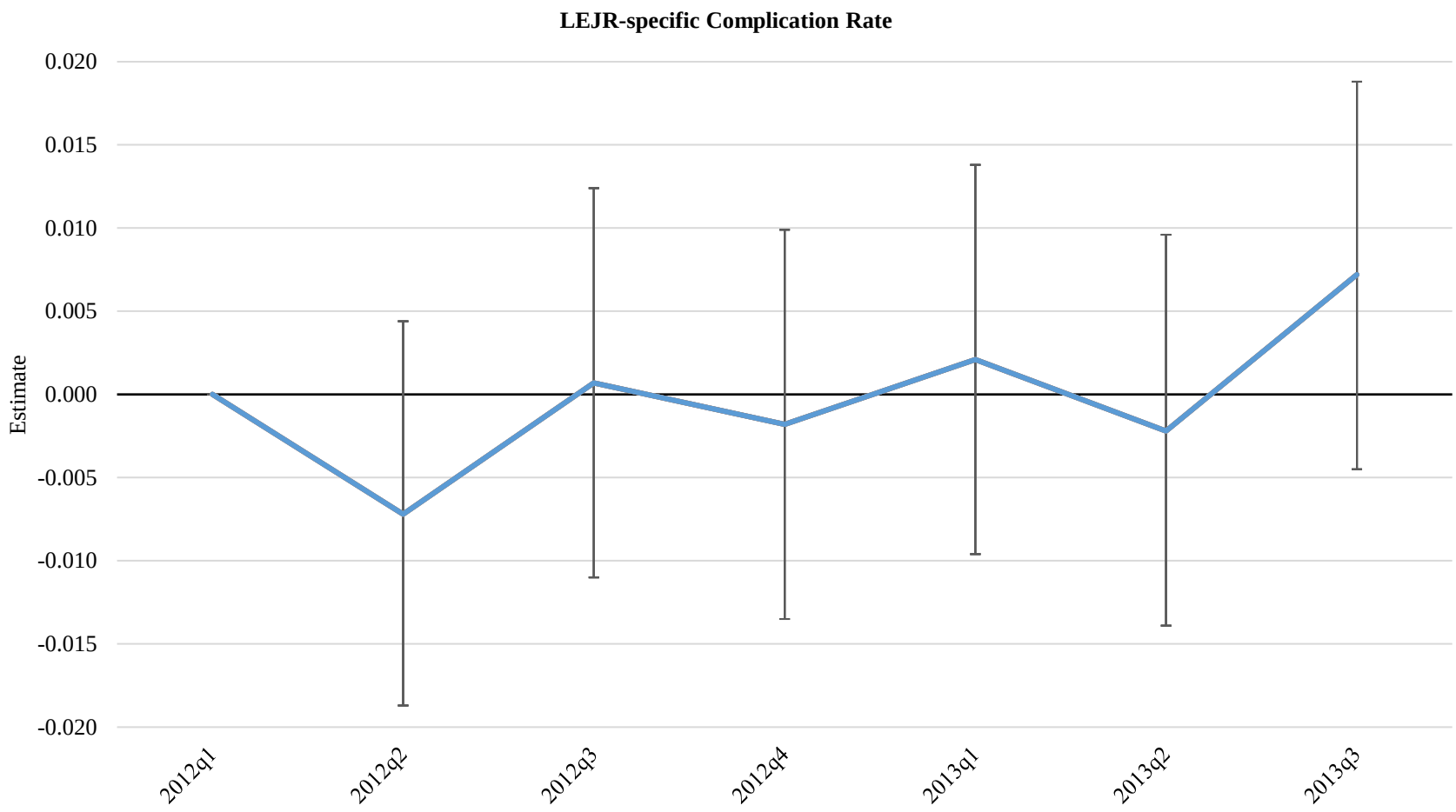
p-values	.	0.49	0.25	0.61	0.13	0.21	0.33



p-values	.	1.00	0.81	0.47	0.86	0.87	0.57



p-values	.	0.71	0.61	0.49	0.84	0.22	0.25



p-values	.	0.22	0.91	0.76	0.73	0.72	0.23

Appendix Exhibit 5. Patient characteristics by BPCI participation and program period, 2012-2016 (Full Table)

Sample Characteristics[a]						
	Non-BPCI Patients			BPCI Patients		
	Pre-BPCI	Early BPCI	Late BPCI	Pre-BPCI	Early BPCI	Late BPCI
Markets, No.	98	98	98	123	123	123
Hospitals, No.	244	244	241	244	244	244
Beneficiaries, No.	20,497	19,734	11,876	75,614	75,297	54,744
Patient Characteristics						
	Non-BPCI Patients			BPCI Patients		
	Pre-BPCI	Early BPCI	Late BPCI	Pre-BPCI	Early BPCI	Late BPCI
Age, mean	73.4 (2.6)	73.4 (2.7)	73.0 (3.3)	73.3* (1.9)	73.0* (1.7)	73.0* (1.8)
Black race, %[b]	8.7 (14.5)	8.3 (13.1)	7.9 (13.6)	7.6 (10.7)	7.2 (10.6)	7.1 (10.9)
Female, %	64.4 (8.5)	63.4 (10.3)	62.2 (13.9)	66.1*** (5.6)	64.2*** (4.8)	64.2*** (5.9)
Dual-eligible, %[c]	13.3 (10.9)	12.5 (11.0)	11.8 (11.0)	15.5* (11.2)	14.2* (10.5)	13.5* (10.3)
Elixhauser comorbidity index, mean[d,e]	5.5*** (3.1)	5.3*** (2.9)	4.4*** (3.3)	5.4*** (2.8)	4.8*** (2.1)	4.3*** (2.1)
Prior acute care hospital use, %[e]	18.9** (10.1)	17.5** (8.9)	16.3** (10.9)	17.9*** (7.1)	15.7*** (4.6)	15.3*** (4.6)
Prior IRF use, %[e,f]	1.3 (2.5)	1.6 (3.2)	1.4 (3.1)	1.4 (1.8)	1.2 (1.6)	1.2 (1.6)
Prior SNF use, %[e,g]	5.2 (5.6)	5.3 (5.2)	4.9 (8.0)	5.1* (3.8)	4.7* (2.7)	4.6 (3.1)

This table describes patient characteristics in the pre-BPCI and BPCI periods for patients receiving LEJR at BPCI Hospitals and Non-BPCI Hospitals. Differential changes for BPCI and Non-BPCI Patients are shown in Appendix Exhibit 3. Pre-Period=January 2011 to September 2013. BPCI Period=October 2013 to December 2016. Early BPCI=October 2013 to June 2015. Late BPCI=July 2015 to December 2016; however, data presented were drawn from LEJR episodes occurring between July 2015 and September 2016 in order to allow for 90-day post-discharge period. Wilcoxon rank-sum or t-tests were used to test the differences in continuous variables and Chi-square tests for categorical variables. *p<0.05, **p<0.01, ***p<0.001. [a]Characteristics for Non-BPCI hospitals and patients were drawn from a 20% Medicare claims sample while characteristics for BPCI hospitals and patients were drawn from a 100% sample. [b]Race was broken out as black versus others because of existing disparities in access to LEJR among black patients specifically. [c]Dual eligible indicates eligibility for both the Medicare and Medicaid programs as an indicator of low socioeconomic status. [d]The Elixhauser comorbidity score is an index of severity with a range of −32 to +92 with increasing scores highly correlated with increased probability of in-hospital death. [e]Calculated using data from the year prior to LEJR hospitalization. [f]Inpatient Rehabilitation Facility. [g]Skilled Nursing Facility.

Appendix Exhibit 6. Changes in patient and market characteristics by BPCI participation and program period, 2012-2016

Characteristics of Beneficiaries Receiving LEJR[a]								
	Non-BPCI Hospitals			BPCI Hospitals				
	Pre-BPCI	BPCI	Difference	Pre-BPCI	BPCI	Difference	DiD[b]	DiD P-Value[b]
Patient Characteristics								
Age, mean year (SD)	73.4 (2.6)	73.2 (2.4)	-0.2	73.3 (1.9)	73.0 (1.6)	-0.3	-0.1	0.61
Black, % (SD)[c]	8.7 (14.5)	8.1 (12.8)	-0.5	7.6 (10.7)	7.2 (10.8)	-0.4	0.1	0.93
Female, % (SD)	64.4 (8.5)	63.0 (8.9)	-1.3	66.1 (5.6)	64.2 (4.1)	-1.8	-0.5	0.61
Dual-eligible, % (SD)[d]	13.3 (10.9)	12.0 (8.9)	-1.3	15.5 (11.2)	13.9 (10.2)	-1.6	-0.3	0.80
Elixhauser comorbidity index, mean (SD)[e,f]	5.5 (3.1)	5.0 (2.6)	-0.5	5.4 (2.8)	4.5 (1.9)	-0.9	-0.3	0.32
Prior acute care hospital use, % (SD)[f]	18.9 (10.1)	16.9 (7.2)	-2.0	17.9 (7.1)	15.6 (3.8)	-2.4	-0.4	0.70
Prior IRF use, % (SD)[f,g]	1.3 (2.5)	1.6 (2.6)	0.2	1.4 (1.8)	1.2 (1.4)	-0.2	-0.5	0.09
Prior SNF use, % (SD)[f,h]	5.2 (5.6)	5.0 (4.1)	-0.3	5.1 (3.8)	4.7 (2.3)	-0.5	-0.2	0.67
Market Structure Characteristics[i]								
	Non-BPCI Hospitals			BPCI Hospitals				
	Pre-BPCI	BPCI	Difference	Pre-BPCI	BPCI	Difference	DiD[b]	DiD P-Value[b]
Quarterly LEJR volume, mean (SD)	303.4 (258.0)	326.5 (281.9)	23.0	375.6 (316.2)	408.2 (348.7)	32.5	9.5	0.87
Hospital beds, mean (SD)	3,124.2 (2,870.4)	3,153.9 (2,887.9)	29.7	4,283.4 (4,633.9)	4,320.5 (4,670.5)	37.1	7.4	0.99
SNF beds, mean (SD)[h]	5,145.8 (4,239.6)	5,118.2 (4,244.8)	-27.6	6,758.1 (6,259.5)	6,753.3 (6,243.9)	-4.8	22.8	0.98
MA penetration, % (SD)[j]	24.2 (13.8)	28.1 (13.6)	3.9	26.8 (12.9)	30.7 (13.4)	3.9	0.0	1.00
ACO penetration, % (SD)[k]	3.8 (4.4)	14.8 (11.5)	11.1	4.7 (4.8)	21.7 (13.7)	17.0	5.9	0.001

Hospital concentration, HHI (SD)[l,m]	3,237.0 (1,950.8)	3,256.8 (1,893.8)	19.8	2,697.4 (1,788.7)	2,748.6 (1,824.0)	51.2	31.4	0.93
SNF concentration, HHI (SD)[h,m]	1,398.8 (1,074.0)	1,239.3 (860.2)	-159.5	1,172.5 (822.6)	1,129.3 (788.9)	-43.1	116.4	0.49
Markets with PGP, %[n]	0.0	46.9	46.9	0.0	53.7	53.7	6.72	0.32

This table describes patient and market characteristics in the pre-BPCI and BPCI periods for patients receiving LEJR at BPCI Hospitals and Non-BPCI Hospitals. Pre-Period=January 2011 to September 2013. BPCI Period=October 2013 to December 2016. Early BPCI=October 2013 to June 2015. Late BPCI=July 2015 to December 2016; however, data presented were drawn from LEJR episodes occurring between July 2015 and September 2016 in order to allow for 90-day post-discharge period. Wilcoxon rank-sum or t-tests were used to test the differences in continuous variables and Chi-square tests for categorical variables. Diff-in-diff Estimate=differences-in-differences estimate. [a]Characteristics for Non-BPCI hospitals and patients were drawn from a 20% Medicare claims sample while characteristics for BPCI hospitals and patients were drawn from a 100% sample. [b]DiD=difference-in-differences estimate. [c]Race was broken out as black versus others because of existing disparities in access to LEJR among black patients specifically. [d]Dual eligible indicates eligibility for both the Medicare and Medicaid programs as an indicator of low socioeconomic status. [e]The Elixhauser comorbidity score is an index of severity with a range of −32 to +92 with increasing scores highly correlated with increased probability of in-hospital death. [f]Calculated using data from the year prior to LEJR hospitalization. [g]Inpatient Rehabilitation Facility. [h]Skilled Nursing Facility. [i]Market characteristics are calculated based on total procedural (episode) volume using a 100% sample rather than based on unique patients as in the rest of the table. [j]Medicare Advantage penetration was determined using the 100% Medicare Beneficiary Summary File and computed at the market–quarter level for the proportion of Medicare beneficiaries enrolled in Medicare Advantage at any time during that quarter. [k]ACO penetration (number of beneficiaries in a hospital referral region attributed to a Medicare ACO out of all Medicare beneficiaries) was determined using data from a random 20% sample of fee-for-service beneficiaries and the CMS ACO Provider-level Research Identifiable File on a yearly basis. [l]Herfindahl-Hirschman Index. [m]Hospital and skilled nursing facility concentration was determined using the Herfindahl-Hirschman index.[6] [n]Physician Practice Group. Markets with Physician Group Practice indicates markets with a physician group practice participating in BPCI for the LEJR condition.

Appendix Exhibit 7. Market characteristics by BPCI participation and program period, 2012-2016

Market Characteristics[a]	Non-BPCI Markets			BPCI Markets		
	Pre-BPCI	Early BPCI	Late BPCI	Pre-BPCI	Early BPCI	Late BPCI
Quarterly LEJR volume, mean	303 (258)	315 (273)	342 (294)	376 (316)	394 (337)	428 (367)
Hospital beds, mean	3,124 (2,870)	3,151 (2,887)	3,158 (2,890)	4,283 (4,634)	4,311 (4,657)	4,334 (4,690)
SNF beds, mean	5,146 (4,240)	5,129 (4,247)	5,104 (4,243)	6,758 (6,260)	6,770 (6,266)	6,730 (6,215)
MA penetration, %[b]	24.2* (13.8)	27.4* (13.6)	29.1* (13.8)	26.8* (12.9)	29.9* (13.2)	31.8* (13.6)
ACO penetration, %[c]	3.8*** (4.4)	13.1*** (11.4)	17.4*** (12.8)	4.7*** (4.8)	19.0*** (13.7)	25.3*** (15.3)
Hospital concentration, HHI[d,e]	3,237 (1,951)	3,253 (1,904)	3,262 (1,885)	2,697 (1,789)	2,738 (1,818)	2,763 (1,841)
SNF concentration, HHI[e,f]	1,399 (1,074)	1,270 (906)	1,197 (818)	1,173 (823)	1,134 (787)	1,124 (806)
Markets with PGP, %[g]	0.0***	26.5***	46.9***	0.0***	27.6***	53.7***

This table describes market characteristics in the pre-BPCI and BPCI periods for patients receiving LEJR at BPCI Hospitals and Non-BPCI Hospitals. Differential changes for BPCI and Non-BPCI Patients are shown in Appendix Exhibit 3. Pre-Period=January 2011 to September 2013. BPCI Period=October 2013 to December 2016. Early BPCI=October 2013 to June 2015. Late BPCI=July 2015 to December 2016; however, data presented were drawn from LEJR episodes occurring between July 2015 and September 2016 in order to allow for 90-day post-discharge period. Wilcoxon rank-sum or t-tests were used to test the differences in continuous variables and Chi-square tests for categorical variables. *p<0.05, **p<0.01, ***p<0.001. [a]Market characteristics are calculated based on total procedural (episode) volume using a 100% sample rather than based on unique patients as in the rest of the table. [b]Medicare Advantage penetration was determined using the 100% Medicare Beneficiary Summary File and computed at the market–quarter level for the proportion of Medicare beneficiaries enrolled in Medicare Advantage at any time during that quarter. [c]ACO penetration (number of beneficiaries in a hospital referral region attributed to a Medicare ACO out of all Medicare beneficiaries) was determined using data from a random 20% sample of fee-for-service beneficiaries and the CMS ACO Provider-level Research Identifiable File on a yearly basis. [d]Herfindahl-Hirschman Index. [e]Hospital and skilled nursing facility concentration was determined using the Herfindahl-Hirschman index (Rhoades SA. The Herfindahl-Hirschman index. *Fed. Res. Bull*. 1993;79:188). [f]Skilled Nursing Facility. [g]Physician Practice Group. Markets with Physician Group Practice indicates markets with a physician group practice participating in BPCI for the LEJR condition.

Appendix Exhibit 8. Unadjusted changes in spending and quality outcomes by BPCI participation, 2012-2016

	Non-BPCI Patients				BPCI Patients					
	Pre-BPCI	**BPCI**	**Difference, %**	**p-value[a]**	**Pre-BPCI**	**BPCI**	**Difference, %**	**p-value[a]**	**Differential change[b]**	**p- value[c]**
Episode spending, $ (SD)	22,834 (13,369)	22,073 (13,177)	-3.3	<0.001	23,552 (13,974)	22,129 (13,752)	-6.0	<0.001	-2.7	<0.001
Mortality rate, %	1.9	1.7	-10.0	0.08	2.0	1.7	-14.8	<0.001	-4.8	0.51
Unplanned readmission rate, %	9.0	8.3	-7.6	0.006	9.5	8.2	-13.2	<0.001	-5.6	0.05
ED visit rate, %[d]	4.8	4.6	-4.4	0.26	13.8	13.7	-0.6	0.6	3.8	0.68
LEJR-specific complica-tion rate, %[e]	3.8	3.8	1.4	0.77	3.8	3.6	-5.8	0.01	-7.2	0.16

Pre-BPCI=January 2012 to September 2013; BPCI=October 2013 to December 2016. All spending estimates were standardized and adjusted for inflation and transformed into 2016 dollars. [a]Obtained from Wilcoxon rank sum tests. [b]Calculated by subtracting the difference between pre-BPCI and BPCI periods among Non-BPCI patients from the difference between pre-BPCI and BPCI periods among BPCI patients. [c]Obtained from two-way ANOVA, with p-value reflecting statistical significance of the interaction term measuring differential change. [d]Emergency Department (ED) visits without hospitalization. [e]Defined by Hospital Compare.

Appendix Exhibit 9. Unadjusted spending components by program period, 2012-2016

Episode Spending Components			
	Pre-BPCI	**Early BPCI**	**Late BPCI**
Index hospitalization, mean % (SD)	64.2 (20.4)	65.7 (19.9)	67.9 (18.8)
Readmissions, mean % (SD)	2.4 (8.0)	2.2 (7.9)	2.1 (7.9)
SNF, mean % (SD)[a]	7.7 (16.1)	7.0 (15.4)	5.1 (13.2)
IRF, mean % (SD)[b]	4.6 (14.1)	3.6 (12.5)	2.3 (9.6)
HHA, mean % (SD)[c]	10.0 (8.6)	10.2 (8.7)	10.6 (9.2)
Professional services, mean % (SD)	7.3 (6.3)	7.6 (6.5)	8.0 (6.8)
Other, mean % (SD)[d]	3.7 (5.7)	3.7 (5.6)	3.9 (5.9)

Pre-BPCI=January 2012 to September 2013; Early BPCI=October 2013 to June 2015; Late BPCI=July 2015 to December 2016. All spending estimates were standardized and adjusted for inflation and transformed into 2016 dollars. [a]Skilled Nursing Facility. [b]Inpatient Rehabilitation Facility. [c]Home Health Agency. [d]Includes durable medical equipment and other outpatient facility and non-professional payments.

Appendix Exhibit 10. Changes in the proportion of episode spending attributable to specific components associated with BPCI participation by program period and timing of hospital participation, 2012-2016

	Overall		Program Period				Timing of Participation					
							Early Entrant Hospitals				Late Entrant Hospitals	
			Early BPCI		Late BPCI		Early BPCI		Late BPCI		Late BPCI	
	DiD[a]	P-Value	DiD	P-Value	DiD	P-Value	DiD	P-Value	DiD	P-Value	DiD	P-Value
Episode spending components												
Index hospitalization, mean % (SD)	0.7 (0.3 to 1.1)	0.001	0.9 (0.5 to 1.4)	<0.001	0.3 (-0.5 to 1.1)	0.45	1.1 (0.6 to 1.6)	<0.001	1.1 (0.2 to 1.9)	0.01	-0.2 (-1.0 to 0.5)	0.56
Re-admissions, mean % (SD)	-0.02 (-0.2 to 0.2)	0.85	-0.02 (-0.2 to 0.2)	0.87	-0.03 (-0.4 to 0.3)	0.84	-0.01 (-0.2 to 0.2)	0.91	-0.1 (-0.4 to 0.3)	0.76	0.02 (-0.3 to 0.4)	0.90
SNF, mean % (SD)[b]	-0.4 (-0.7 to -0.1)	0.02	-0.5 (-0.9 to -0.1)	0.01	-0.3 (-0.9 to 0.3)	0.30	-0.5 (-0.9 to -0.1)	0.01	-0.3 (-0.9 to 0.3)	0.28	-0.3 (-0.9 to 0.3)	0.37
IRF, mean % (SD)[c]	-0.7 (-1.0 to -0.4)	<0.001	-0.5 (-0.9 to -0.2)	0.002	-0.9 (-1.4 to -0.5)	<0.001	-0.7 (-1.0 to -0.3)	<0.001	-1.6 (-2.1 to -1.1)	<0.001	-0.3 (-0.7 to 0.2)	0.28
HHA, mean % (SD)[d]	0.2 (0.04 to 0.4)	0.02	-0.2 (-0.4 to 0.002)	0.05	1.0 (0.7 to 1.4)	<0.001	-0.3 (-0.5 to -0.1)	0.01	0.8 (0.4 to 1.1)	<0.001	1.0 (0.6 to 1.3)	<0.001
Professional services, mean % (SD)	0.2 (0.01 to 0.3)	0.03	0.2 (0.1 to 0.4)	0.01	0.04 (-0.2 to 0.3)	0.76	0.3 (0.1 to 0.5)	0.002	0.1 (-0.1 to 0.4)	0.31	-0.2 (-0.4 to 0.1)	0.24
Other, mean % (SD)[e]	0.002 (-0.1 to 0.1)	0.97	0.1 (-0.1 to 0.2)	0.43	-0.1 (-0.3 to 0.2)	0.48	0.1 (-0.1 to 0.2)	0.42	0.01 (-0.2 to 0.3)	0.95	-0.1 (-0.3 to 0.2)	0.53

This table shows results from difference-in-differences models evaluating the association between BPCI participation and differential changes in episode spending attributable to specific components. Changes are displayed for the overall cohort and study period (Overall) as well as program period (Early BPCI and Late BPCI) and timing of participation (Early Entrant and Late Entrant hospitals). Negative estimates indicate decreases in the proportion of episode spending for the relevant component. Early BPCI=October 2013 to June 2015; Late BPCI=July 2015 to December 2016. BPCI=Early BPCI + Late BPCI (October 2013 to December 2016). All spending estimates were standardized and transformed into 2016 dollars. [a]DiD=differences-in-differences estimate. [b]Skilled Nursing Facility. [c]Inpatient Rehabilitation Facility. [d]Home Health Agency. [e]Includes durable medical equipment and other outpatient facility and non-professional payments.

Appendix Exhibit 11. Changes in quality outcomes associated with BPCI participation by program period and timing of hospital participation, 2012-2016 (Full Table)

	Overall		Program Period				Timing of Participation					
							Early Entrant Hospitals				Late Entrant Hospitals	
			Early BPCI		Late BPCI		Early BPCI		Late BPCI		Late BPCI	
	DiD[a]	95% CI	DiD	95% CI	DiD	95% CI	DiD	95% CI	DiD	95% CI	DiD	95% CI
Mortality rate, %	-0.15	-0.5 to 0.2	-0.03	-0.4 to 0.3	-0.34	-3.1 to 0.4	0.01	-0.4 to 0.4	-0.27	-0.8 to 0.3	-0.37	-0.9 to 0.2
Unplanned re-admission rate, %	0.15	-0.5 to 0.8	0.21	-0.5 to 1.0	0.03	-0.9 to 0.2	0.24	-0.5 to 1.0	-0.10	-1.3 to 1.1	0.18	-1.0 to 1.4
ED visit rate, %[b]	-0.19	-1.0 to 0.6	-0.22	-1.1 to 0.7	-0.08	-1.1 to 1.1	-0.25	-1.2 to 0.6	0.15	-1.1 to 1.4	0.54	-0.8 to 1.9
LEJR-specific complica-tion rate, %[c]	0.12	-0.3 to 0.6	0.21	-0.3 to 0.7	-0.03	-1.3 to 1.1	0.24	-0.2 to 0.7	0.003	-0.8 to 0.8	-0.09	-0.9 to 0.7

This table shows results from difference-in-differences models evaluating the association between BPCI participation and differential changes in quality outcomes, with changes displayed for the overall cohort and study period (Overall) as well as program period (Early BPCI and Late BPCI) and timing of participation (Early Entrant and Late Entrant hospitals). Negative estimates indicate reductions in rates (i.e., quality improvements). Early BPCI=October 2013 to June 2015; Late BPCI=July 2015 to December 2016. [a]DiD =differences-in-differences estimate. [b]Emergency Department (ED) visits without hospitalization. [c]Defined by Hospital Compare.

Appendix Exhibit 12. Sensitivity analysis extending methods from prior CMS evaluations

Adjusted results				
	Sensitivity analysis (without IV or hospital fixed effects)		Primary analysis (with IV[a] and hospital fixed effects)	
	DiD[b] (95% CI)	p-value	DiD[b] (95% CI)	p-value
Episode spending, $ (%)	-2.2 (-2.9 to -1.5)	<0.001	-1.6 (-2.6 to -0.6)	<0.001
Mortality rate, %	-0.3 (-0.5 to 0.0)	0.04	-0.2 (-0.5 to 0.2)	0.35
Unplanned readmission rate, %	-0.2 (-0.7 to 0.4)	0.59	0.1 (-0.5 to 0.8)	0.67
ED visit rate, %[c]	0.2 (-0.5 to 0.8)	0.65	-0.2 (-1.0 to 0.6)	0.63
LEJR-specific complication rate, %[d]	0.2 (-0.2 to 0.5)	0.43	0.1 (-0.3 to 0.6)	0.59

[a]Instrumental variable approach. [b]DiD=difference-in-differences estimate. [c]Emergency Department (ED) visits without hospitalization. [d]Defined by Hospital Compare.

Appendix Exhibit 13. Sensitivity analysis excluding January to September 2013 from the baseline period

	Early BPCI		Late BPCI		Overall BPCI	
	DiD[a] (95% CI)	**p-value**	**DiD[a] (95% CI)**	**p-value**	**DiD[a] (95% CI)**	**p-value**
Episode spending, %	-1.6 (-2.7 to -0.5)	0.01	-0.9 (-2.5 to 0.7)	0.27	-1.4 (-2.3 to -0.4)	0.01
Mortality rate, %	-0.01 (-0.4 to 0.4)	0.97	-0.4 (-0.9 to 0.1)	0.16	-0.2 (-0.5 to 0.2)	0.34
Unplanned readmission rate, %	0.2 (-0.6 to 1.0)	0.60	0.2 (-0.9 to 1.4)	0.70	0.2 (-0.5 to 0.9)	0.57
ED visit rate, %[b]	-0.3 (-1.2 to 0.6)	0.57	0.002 (-1.3 to 1.3)	>0.99	-0.2 (-1.0 to 0.6)	0.63
LEJR-specific complication rate, %[c]	0.2 (-0.3 to 0.7)	0.34	0.1 (-0.7 to 0.9)	0.81	0.2 (-0.3 to 0.6)	0.41

Pre-BPCI=January 2012 to September 2013; Early BPCI=October 2013 to June 2015; Late BPCI=July 2015 to December 2016. All spending estimates were standardized and adjusted for inflation and transformed into 2016 dollars. [a]DiD=difference-in-differences estimate. [b]Emergency Department (ED) visits without hospitalization. [c]Defined by Hospital Compare.

Appendix Exhibit 14. Sensitivity analysis using less stringent propensity matching

	Early BPCI		Late BPCI		Overall BPCI	
	DiD[a] (95% CI)	p-value	DiD[a] (95% CI)	p-value	DiD[a] (95% CI)	p-value
Episode spending, %	-2.4 (-3.6 to -1.2)	<0.001	-1.9 (-3.2 to -0.7)	<0.001	-2.1 (-3.1 to -1.2)	<0.001
Mortality rate, %	-0.1 (-0.5 to 0.2)	0.41	-0.3 (-0.7 to 0.04)	0.08	-0.3 (-0.5 to 0.03)	0.08
Unplanned readmission rate, %	-0.1 (-0.9 to 0.6)	0.72	-0.2 (-1.0 to 0.6)	0.63	-0.2 (-0.8 to 0.4)	0.60
ED visit rate, %[b]	-0.3 (-1.2 to 0.6)	0.48	0.1 (-0.9 to 1.0)	0.87	-0.1 (-0.8 to 0.6)	0.76
LEJR-specific complication rate, %[c]	0.1 (-0.5 to 0.5)	0.85	-0.4 (-0.9 to 0.2)	0.18	-0.2 (-0.6 to 0.2)	0.38

Pre-BPCI=January 2012 to September 2013; Early BPCI=October 2013 to June 2015; Late BPCI=July 2015 to December 2016. All spending estimates were standardized and adjusted for inflation and transformed into 2016 dollars. [a]DiD=difference-in-differences estimate. [b]Emergency Department (ED) visits without hospitalization. [c]Defined by Hospital Compare.

Appendix Exhibit 15. Sensitivity analysis excluding CJR hospitals from propensity score matching and analyses

	Early BPCI		Late BPCI		Overall BPCI	
	DiD[a] (95% CI)	p-value	DiD[a] (95% CI)	p-value	DiD[a] (95% CI)	p-value
Episode spending, %	-1.5 (-2.7 to -0.2)	0.02	-1.4 (-3.0 to 0.3)	0.10	-1.5 (-2.5 to -0.4)	0.01
Mortality rate, %	-0.1 (-0.5 to 0.3)	0.70	-0.4 (-0.8 to 0.1)	0.16	-0.2 (-0.5 to 0.1)	0.23
Unplanned readmission rate, %	0.4 (-0.4 to 1.2)	0.33	-0.3 (-1.4 to 0.8)	0.60	0.1 (-0.6 to 0.8)	0.79
ED visit rate, %[b]	-0.1 (-1.0 to 0.9)	0.92	0.1 (-1.2 to 1.3)	0.90	-0.02 (-0.9 to 0.8)	0.97
LEJR-specific complication rate, %[c]	0.1 (-0.4 to 0.7)	0.62	-0.5 (-1.2 to 0.3)	0.20	-0.1 (-0.6 to 0.4)	0.60

Pre-BPCI=January 2012 to September 2013; Early BPCI=October 2013 to June 2015; Late BPCI=July 2015 to December 2016. All spending estimates were standardized and adjusted for inflation and transformed into 2016 dollars. [a]DiD=difference-in-differences estimate. [b]Emergency Department (ED) visits without hospitalization. [c]Defined by Hospital Compare.

References

1. Parsons LS. Performing a 1:N case control match on propensity score. Poster presented at: SAS Users Group International 29; May, 2004, Montreal, Canada.
2. Rosenbaum PR. Optimal Matching for Observational Studies. *J Am Stat Assoc.* 1989;84(408):1024-1032.
3. Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica.* 1994;62(2):467-475.
4. Angrist JD, Krueger AB. Instrumental variables and the search for identification: from supply and demand to natural experiments. *J Econ Perspect.* 2001;15(4):69-85.
5. Angrist J, Pischke J. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press; 2008.
6. Rhoades SA. The Herfindahl-Hirschman index. *Fed Res Bull*. 1993;79:188.