Performance of Physician Groups and Hospitals Participating in Bundled Payments Among Medicare Beneficiaries

Joshua M. Liao, MD, MSc; Qian Huang, MPH; Erkuan Wang, MS; Kristin Linn, PhD; Torrey Shirk, BA; Jingsan Zhu, MS; Deborah Cousins, MS; Amol S. Navathe, MD, PhD

Abstract

IMPORTANCE Hospital participation in bundled payment initiatives has been associated with financial savings and stable quality of care. However, how physician group practices (PGPs) perform in bundled payments compared with hospitals remains unknown.

OBJECTIVES To evaluate the association of PGP participation in the Bundled Payments for Care Improvement (BPCI) initiative with episode outcomes and to compare these with outcomes for participating hospitals.

DESIGN, SETTINGS, AND PARTICIPANTS This cohort study with a difference-in-differences analysis used 2011 to 2018 Medicare claims data to compare the association of BPCI participation with episode outcomes for PGPs vs hospitals providing medical and surgical care to Medicare beneficiaries. Data analyses were conducted from January 1, 2020, to May 31, 2022.

EXPOSURES Hospitalization for any of the 10 highest-volume episodes (5 medical and 5 surgical) included in the BPCI initiative for Medicare patients of participating PGPs and hospitals.

MAIN OUTCOMES AND MEASURES The primary outcome was 90-day total episode spending. Secondary outcomes were 90-day readmissions and mortality.

RESULTS The total sample comprised data from 1,288,781 Medicare beneficiaries, of whom 696,710 (mean [SD] age, 76.2 [10.8] years; 432,429 [59.7%] women; 619,655 [85.5%] White individuals) received care through 379 BPCI-participating hospitals and 1,441 propensity-matched non−BPCI-participating hospitals, and 592,071 (mean [SD] age, 75.4 [10.9] years; 527,574 [86.6%] women; 360,835 [59.3%] White individuals) received care from 6,405 physicians in BPCI-participating PGPs and 24,758 propensity-matched physicians in non−BPCI-participating PGPs. For PGPs, BPCI participation was associated with greater reductions in episode spending for surgical (difference, −$1368; 95% CI, −$1648 to −$1088) but not for medical episodes (difference, −$101; 95% CI, −$410 to $206). Hospital participation in BPCI was associated with greater reductions in episode spending for both surgical (−$1010; 95% CI, −$1345 to −$675) and medical (−$763; 95% CI, −$1139 to −$386) episodes.

CONCLUSIONS AND RELEVANCE This cohort study and difference-in-differences analysis of PGPs and hospital participation in BPCI found that bundled payments were associated with cost savings for surgical episodes for PGPs, and savings for both surgical and medical episodes for hospitals. Policy makers should consider the comparative performance of participant type when designing and evaluating future bundled payment models.

Key Points

Question Do physician group practices participating in bundled payments among Medicare beneficiaries exhibit similar or different changes in episode outcomes compared with participating hospitals?

Findings This cohort study with a difference-in-differences analysis found that physician group practices participating in bundled payments had associated savings with surgical but not medical episodes, whereas participating hospitals had savings associated with both episode types.

Meaning The findings of this cohort study suggest that policy makers should consider the comparative performance of participant type when designing and evaluating future bundled payment models.

Open Access. This is an open access article distributed under the terms of the CC-BY License.
Introduction

Health care policy makers and payers are working to improve results under alternative payment models, in part by determining which organization types to engage in episode-based bundled payments.\(^1,2\) This effort requires an understanding of how patient outcomes are affected when patients receive care under different participant types, such as hospitals and physician group practices (PGPs).

Although hospital participation in bundled-payment programs has been well studied,\(^3-9\) little is known about how PGP participation affects outcomes, particularly compared with hospital participation. Along with hospitals, PGPs participated in model 2 of Medicare's Bundled Payments for Care Improvement (BPCI) initiative,\(^10\) assuming accountability for the quality and costs of medical and surgical episodes spanning hospital admission and up to 90 days of postacute care. Although PGP participation in BPCI was associated with reduced Medicare payments and improved quality for joint replacement episodes,\(^11\) data are lacking for the hundreds of groups participating in other medical and surgical episodes.

To coordinate participation in future payment models, policy makers must understand the dynamics of PGP vs hospital performance, particularly given the evidence from other payment models indicating that physician groups may perform differently than hospitals in managing quality and costs.\(^12\) We addressed this knowledge gap by evaluating the association between PGPs and hospitals participating in BPCI model 2 on episode outcomes and comparing PGP vs hospital performance.

Methods

This study was approved by the University of Pennsylvania Institutional Review Board and informed consent was waived because only historical data with minimal risk of harm were used. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.\(^13\)

Data Collection

We identified PGPs and hospitals participating in BPCI model 2 and dates of entry by participant-episode from BPCI Initiative Episode Analytic Files from the US Centers for Medicare & Medicaid Services. These data were linked to Provider Enrollment, Chain, and Ownership System files, from which we identified participating physicians' National Provider Identification numbers from each participating PGP. Using a 20% random national sample of 2011 to 2018 Medicare claims data, we identified patients receiving care through PGPs or hospitals participating in BPCI model 2 for 1 of 10 episodes of interest, as well as patients receiving care through nonparticipating PGPs and hospitals.

Data from the 2011 to 2012 American Hospital Association Annual Survey were used to capture hospital characteristics, and 2011 to 2017 Medicare Provider of Service, Beneficiary Summary, American Community Survey, and Accountable Care Organization files were used to obtain market and additional hospital characteristics. Data from the Medicare Data on Provider Practice and Specialty file,\(^14\) Physician Compare, and Compendium of US Health Systems\(^15\) were used to obtain physician characteristics.

Study Periods, Sample, and Episode Construction

The study period spanned from January 1, 2011, through December 31, 2017, and encompassed a baseline period prior to the start of BPCI (January 1, 2011-September 30, 2013) and a subsequent intervention period (October 1, 2013-December 31, 2017). The study sample included Medicare fee-for-service beneficiaries receiving care through BPCI PGPs and hospitals for 1 of 10 episodes, each defined by a set of Medicare Severity-Diagnosis Related Group codes: the top 5 highest-volume medical episodes (congestive heart failure, pneumonia, sepsis, chronic obstructive pulmonary...
disease, and other respiratory) and the top 5 highest-volume surgical episodes (lower extremity joint replacement, hip and femur procedures except major joint, percutaneous coronary intervention, upper extremity joint replacement, and spinal fusion). We excluded beneficiaries with end-stage kidney disease or insurance coverage through Medicare Advantage, as well as beneficiaries who had any non-Inpatient Prospective Payment System claims, lacked continuous primary Medicare fee-for-service coverage during or in the 12 months preceding the episode, or died during the index hospital admission.

We constructed episodes beginning with hospital admission and spanning 90 days after hospital discharge, capturing episodes beginning on or before September 30, 2017. To avoid bias arising from Medicare precedence rules for overlapping episodes between participating PGPs and hospitals, we followed prior methods and constructed naturally occurring episodes by assigning overlapping ones to the earlier hospitalization.3 We excluded episodes between January 1, 2013, and September 30, 2013, to account for the transitional period during which PGPs and hospitals may have implemented care process changes in anticipation of BPCI.

Study Groups
We used Medicare claims data (2011-2018) to identify BPCI PGPs and BPCI hospitals. Groups and hospitals that never participated in any of the 48 episodes in BPCI were categorized as non-BPCI PGPs and non-BPCI hospitals. We used propensity scores to match BPCI with non-BPCI PGPs and BPCI with non-BPCI hospitals (eMethods 1 in Supplement 1). Markets were defined using hospital referral regions12 and those with any BPCI PGPs or BPCI hospitals were defined as BPCI Markets. We categorized patients based on hospitalization for 1 of 10 episodes of interest through BPCI PGPs, non-BPCI PGPs, BPCI hospitals, or non-BPCI hospitals. Patients receiving care through BPCI PGPs and hospitals were categorized as “BPCI-both.”

Exposure and Covariates
The study exposures were dichotomous indicators of PGP participation and hospital participation in BPCI. To reflect participant entry into BPCI at different times, participation indicators were time-varying and specific to each PGP or hospital. Groups and hospitals were considered as participants after enrolling in BPCI, regardless of subsequently dropping out. Covariates were chosen based on prior studies and included patient demographic and clinical variables, such as age, sex, and disease severity (defined using Elixhauser comorbidities), as well time-varying market variables, such as Medicare population size and Medicare Advantage penetration.3,4,16,17 Race and ethnicity data were obtained from Medicare claims data, which included Black and White race categories that were used directly in the analyses; however, Asian, Hispanic, Native American (including American Indian, Alaska Native, Native Hawaiian, or other Pacific Islander), and other responses were classified in the “other” category.

Study Outcomes
The study’s primary outcome was 90-day total episode spending (actual spending adjusted for inflation to 2018 US dollars). Secondary outcomes were 90-day readmissions and 90-day mortality. We also evaluated a number of exploratory spending and utilization outcomes (eMethods 2 in Supplement 1).

Statistical Analysis
Characteristics between the propensity-matched BPCI and non-BPCI PGPs and BPCI and non-BPCI hospitals were compared using standardized differences of means and proportions.18 We used a difference-in-differences (DID) method to mimic a 2 × 2 factorial experiment comparing BPCI participation among hospitals and PGPs. Per the factorial design, we classified the treatment groups to reflect patient exposure to BPCI hospitals, BPCI PGPs, both (BPCI-both), and neither (non-BPCI).
This approach enabled the comparison of episode performance for BPCI PGPs (vs non-BPCI) and BPCI hospitals (vs non-BPCI), as well as between BPCI PGPs and BPCI hospitals.

In adjusted analyses, DID models included PGP- and hospital-specific indicators of BPCI participation as treatment to reflect the time-varying nature of BPCI participation—that is, the fact that PGS and hospitals could start participating at different times. This approach contrasted with traditional DID models, in which the baseline and treatment periods are fixed regardless of timing of actual contract initiation. Nonparticipating PGS and hospitals were assigned the same treatment indicators as their propensity matched organizations. We evaluated medical and surgical episodes separately because they involve different care processes that may have different associations with outcomes.

We used generalized linear models with identity links and normal distributions for all outcomes. All models included episode (Medicare Severity-Diagnosis Related Group code) and market (hospital referral regions) fixed effects to generate within-episode type, within-market estimates that addressed time-invariant episode type, and geographic differences. Models also included time (calendar quarter) fixed-effects to account for secular trends. Robust standard errors were clustered at the hospital level.

We took steps to examine assumptions of our DID method for the primary outcome. First, we examined but did not observe divergent trends for the primary outcome in the baseline period for study groups (eFigures 1 and 2 in Supplement 1). Second, given concerns regarding time-varying DID designs, we assessed the relationship between the primary outcome and potentially time-varying market-level covariates and how those covariates changed over time across the 4 study groups (eMethods 2 in Supplement 1). Covariates were deemed potential confounders if their association with an outcome varied over time or if their distributions changed differentially over time for BPCI vs non-BPCI PGPs or hospitals in the propensity matched samples. We identified 3 time-varying confounders: Medicare Advantage penetration, acute care organization penetration, and hospital concentration as measured by the Herfindahl-Hirschman index, and accounted for each by including interaction terms between each and a time variable in models. Third, given the potential for time-varying patient selection based on factors that were unobservable in study data, we assessed but did not observe changes in the association between patient Elixhauser comorbidity score and the primary outcome over time across study groups.

Final model specifications for the primary outcome included only covariates deemed as potential confounders according to this process (eMethods 3 in Supplement 1). The Holm–Bonferroni method was used to correct for multiple testing across study groups for the primary study outcome. Under this approach, statistical significance was defined for total episode spending as \( P = .0125, .017, .025, \) and \( .05 \) for a ranked set of 4 comparisons. Analyses were conducted from January 1, 2020, to May 31, 2022, using SAS, version 9.4 (SAS Institute).

Sensitivity and Other Analysis

We accounted for skewness in spending by using nonlinear models with log links and gamma distributions to analyze spending. Additionally, to reflect BPCI initiative precedence rules, we used models that considered patients in the group BPCI-both (those receiving care from both PGP and hospital participants) as being in the BPCI PGP group. Finally, we repeated analyses for the mortality outcome by including individuals who died during index hospitalization.

In response to an association observed between BPCI participation and differential changes in mortality that, to our knowledge, has not been previously described in the literature, we conducted post hoc analyses to explore robustness. First, we described selection based on observable and unobservable patient characteristics. Second, we tested approaches to mitigate bias from unobserved clinical severity of patients. Third, we repeated analyses using our modeling approach, but using data from that earlier time period, to assess the ability to replicate those mortality results.
Results

The total study sample comprised 2011 to 2018 Medicare claims data for 1,288,781 beneficiaries, of whom 696,710 patients (mean [SD] age, 76.2 [10.8] years; 432,429 [59.7%] women; 58,854 [7.6%] Black, 619,655 [85.5%] White, and 50,348 [7.0%] individuals of other race and ethnicity) received care through 379 BPCI hospitals and 1441 propensity-matched non-BPCI hospitals; and 592,071 patients (mean [SD] age, 75.4 [10.9] years; 360,835 [59.3%] women; 41,660 [6.8%] Black, 527,574 [86.6%] White, and 39,783 [6.5%] individuals of other race and ethnicity) received care from 6405 physicians in BPCI PGPs and 24,758 propensity-matched physicians in non-BPCI PGPs.

Hospital and Physician Characteristics

Compared with non–BPCI-participating hospitals, participating hospitals tended to be larger, nonprofit, teaching hospitals located in urban areas and markets with larger populations and smaller proportions of low-income individuals (eTable 1 in Supplement 1). Physicians participating in BPCI vs non-BPCI PGPs differed in several ways including that the former tended to be younger, to care for fewer beneficiaries, and to be affiliated with a health system (eTable 2 in Supplement 1). Differences between BPCI-participating and non-BPCI hospitals, as well as BPCI-participating physicians and non-BPCI PGPs, decreased after propensity score matching, with standardized mean differences of 0.1 or less for all variables.

Medical Episodes

For medical episodes, total episode spending increased, whereas 90-day readmissions and mortality decreased over time (Figure 1; eFigure 3 in Supplement 1). In the baseline period, some characteristics differed among patient groups receiving care for medical episodes through BPCI PGP, BPCI hospital, BPCI-both, and non-BPCI groups (Table 1). Market characteristics also varied by group in the baseline period for medical episodes (eTable 3 in Supplement 1).

Between baseline and intervention periods (eTables 5 and 6 in Supplement 1), total episode spending for medical episodes increased in the BPCI PGP and non-BPCI groups and decreased in the BPCI hospital and BPCI-both groups. Study groups differed with respect to unadjusted changes in 90-day readmissions, mortality, and some exploratory outcomes (eTables 5 and 6 in Supplement 1).

In adjusted analysis of medical episodes (eFigure 4A in Supplement 1), patients in the BPCI hospital group had a differential reduction in spending compared with the non-BPCI group (difference, -$763; 95% CI, -$1139 to -$386; P < .001). In contrast, there were no differential changes in spending between BPCI PGP and non-BPCI groups (difference, -$102; 95% CI, -$410 to -$206; P = .52) or between BPCI-both and non-BPCI groups (difference, -$127; 95% CI, -$875 to -$621; P = .74). The BPCI hospitals had differentially greater reductions in total episode spending compared with BPCI PGPs (difference, -$661; 95% CI, -$1118 to -$204; P = .005).

Regarding secondary outcomes for medical episodes (eFigure 4B in Supplement 1), there were differential changes in readmissions between BPCI hospital and non-BPCI groups (difference, -0.8 percentage points [pp]; 95% CI, -1.6 to -0.02 pp; P = .04), but not between BPCI PGP and non-BPCI groups (difference, -0.1 pp; 95% CI, -0.8 to 0.7 pp; P = .83). Compared with the non-BPCI group, there were differential decreases in mortality for BPCI PGPs (difference, -0.7 pp; 95% CI, -1.3 to -0.1 pp; P = .02) and BPCI hospitals (difference, -1.1 pp; 95% CI, -1.8 to -0.5 pp; P < .001). The BPCI hospitals and BPCI PGPs did not exhibit comparatively differential changes in mortality or readmissions. The BPCI hospitals differed from nonparticipants with respect to exploratory outcomes for medical episodes (eTable 7 in Supplement 1).

Surgical Episodes

Total episode spending for surgical episodes increased over time, whereas 90-day readmissions and mortality decreased over time (Figure 2; eFigure 5 in Supplement 1). Study groups differed by patient (Table 2) and market characteristics in the baseline period (eTable 4 in Supplement 1). For surgical
episodes (eTables 8 and 9 in Supplement 1), total episode spending decreased between the baseline and intervention periods in all groups, with the largest decrease among patients in BPCI-both and the smallest decrease among patients in non-BPCI. Across groups, there were some differences in changes in 90-day readmissions, mortality, and exploratory outcomes (eTable 8 and 9 in Supplement 1).

In adjusted analysis (eFigure 6A in Supplement 1), there were differentially greater reductions in total episode spending for surgical episodes for the BPCI hospital group (difference, –$1010 vs non-BPCI; 95% CI, –$1345 to –$675; P < .001), BPCI PGP group (difference, –$1345 vs non-BPCI; 95% CI, –$1674 to –$1016; P < .001), and BPCI-both (difference, –$1584; 95% CI, –$2326 to –$843; P < .001) compared with the non-BPCI group. The magnitude of spending changes did not differ between BPCI hospitals and PGPs (difference, $335; 95% CI, –$97 to $767; P = .13).

Compared with patients in the non-BPCI group, those in the BPCI PGP group had differentially greater changes in 90-day readmissions (–0.6 pp; 95% CI, –1.1 to –0.04 pp; P = .04). In contrast, there were no differential changes in readmissions for the BPCI hospital or BPCI-both groups compared with the non-BPCI group (eFigure 6B in Supplement 1). The BPCI hospitals had differential increase in readmissions compared with changes for BPCI PGPs (0.7 pp; 95% CI, 0.01 to 1.4 pp; P = .047).

Mortality changed differentially for patients cared for through BPCI PGPs (difference, –0.5 pp; 95% CI, –0.8 to –0.2 pp; P = .002) and BPCI hospitals (difference, –0.7 pp; 95% CI, –0.9 to –0.4 pp; P < .001) compared with patients in the non-BPCI group. The magnitude of these changes was not
different for BPCI hospitals vs BPCI PGPs (difference, –0.2 pp; 95% CI, –0.6 to 0.2 pp; P = .26). Compared with nonparticipants, BPCI hospitals and PGPs differed in exploratory surgical episode outcomes (eTable 10 in Supplement 1).

Sensitivity Analysis
Compared with the main study analyses, results of sensitivity analyses were qualitatively similar (eFigures 7-12 in Supplement 1). Post hoc analyses demonstrated differential reductions in observed and unobserved severity over time for the BPCI hospital and PGP vs non-BPCI groups (eMethods 4 in Supplement 1). Although models to adjust for unobserved severity did not produce differential results compared with original models, these approaches were limited in correlation with patient-level observed mortality, thereby limiting the strength of the conclusions. Analyses using older data ranges replicated findings from previous studies3 that found no association between BPCI hospital participation and differential mortality changes.

Discussion
In this cohort study with DID analysis, participation of PGPs and hospitals in BPCI was associated with cost savings for surgical episodes; however, only hospital participation was associated with cost savings for medical episodes. Hospital and PGP participation were associated with different patterns of changes in postacute utilization and mortality. For example, for medical episodes, hospital participation in BPCI was significantly associated with reductions in length of stay at skilled nursing facilities, whereas PGP participation was not. For surgical episodes, PGP participation in BPCI was

Table 1. Characteristics of Patients Admitted for Medical Episodes, by Participation in Bundled Payments, During Baseline Period

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Non-BPCI</th>
<th>BPCI PGP</th>
<th>BPCI hospital</th>
<th>BPCI both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episdes, No.</td>
<td>312 551</td>
<td>44 831</td>
<td>35 172</td>
<td>3994</td>
</tr>
<tr>
<td>Patients, No.</td>
<td>299 246</td>
<td>44 470</td>
<td>34 765</td>
<td>3983</td>
</tr>
<tr>
<td>Female sex, No. (%)</td>
<td>176 766</td>
<td>25 252</td>
<td>19 929</td>
<td>226 065</td>
</tr>
<tr>
<td>Male sex, No. (%)</td>
<td>135 785</td>
<td>19 579</td>
<td>15 243</td>
<td>178 884</td>
</tr>
<tr>
<td>Age, mean (SD), y</td>
<td>76.5 (12.1)</td>
<td>76.2 (12.1)</td>
<td>77.8 (12.0)</td>
<td>77.2 (12.3)</td>
</tr>
<tr>
<td>Race, No. (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>27 702 (8.9)</td>
<td>36 538 (8.2)</td>
<td>35 180 (10.0)</td>
<td>326 (8.2)</td>
</tr>
<tr>
<td>White</td>
<td>262 328 (83.9)</td>
<td>37 757 (84.2)</td>
<td>28 460 (80.9)</td>
<td>333 4 (83.5)</td>
</tr>
<tr>
<td>Othera</td>
<td>22 521 (7.2)</td>
<td>34 217 (7.6)</td>
<td>31 94 (9.1)</td>
<td>334 (8.4)</td>
</tr>
<tr>
<td>Disabled, No. (%)</td>
<td>42 910 (13.7)</td>
<td>64 034 (14.3)</td>
<td>41 077 (11.7)</td>
<td>533 (13.4)</td>
</tr>
<tr>
<td>Medicare−Medicaid dual eligible, No. (%)</td>
<td>84 864 (27.2)</td>
<td>12 633 (28.2)</td>
<td>93 68 (26.6)</td>
<td>1029 (25.8)</td>
</tr>
<tr>
<td>Elixhauser comorbidity index, mean (SD), points</td>
<td>20.1 (14.1)</td>
<td>19.5 (14.0)</td>
<td>20.3 (14.0)</td>
<td>20.4 (14.0)</td>
</tr>
<tr>
<td>Prior use, mean (SD), %b</td>
<td>3.4 (18.0)</td>
<td>3.2 (17.6)</td>
<td>3.6 (18.7)</td>
<td>2.9 (16.8)</td>
</tr>
<tr>
<td>SNF</td>
<td>19.6 (39.7)</td>
<td>17.1 (37.6)</td>
<td>19.8 (39.9)</td>
<td>18.5 (38.8)</td>
</tr>
<tr>
<td>Hospital</td>
<td>52.9 (49.9)</td>
<td>47.6 (49.9)</td>
<td>49.7 (50.0)</td>
<td>45.9 (49.8)</td>
</tr>
<tr>
<td>Episodes, No.</td>
<td>67 322</td>
<td>8387</td>
<td>10 057</td>
<td>968</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>64 925</td>
<td>9967</td>
<td>7093</td>
<td>756</td>
</tr>
<tr>
<td>COPD/bronchitis/asthma</td>
<td>92 256</td>
<td>14 336</td>
<td>10 749</td>
<td>1321</td>
</tr>
<tr>
<td>Pneumonia/pleurisy</td>
<td>30 628</td>
<td>5144</td>
<td>2470</td>
<td>234</td>
</tr>
<tr>
<td>Respiratory infection/inflammation</td>
<td>57 420</td>
<td>6997</td>
<td>4803</td>
<td>715</td>
</tr>
<tr>
<td>Sepsis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: BPCI, Medicare Bundled Payments for Care Improvement initiative; IRF, inpatient rehabilitation facility; COPD, chronic obstructive pulmonary disease; PGP, physician group practice; SNF, skilled nursing facility. a Includes Asian, Hispanic, Native American (American Indian, Alaska Native, Native Hawaiian, other Pacific Islander), and other responses as reported from the Medicare claims data. b Use within the previous 12 months.
associated with reductions in home health use, whereas hospital participation was not. These findings pose 3 implications.

First, these findings underscore the benefit of engaging PGP in episode-based payment models. This analysis adds to prior work by describing the association of PGP participation in BPCI with cost savings for multiple surgical episodes, extending beyond hip and knee replacements. Although it is only one aspect of a payment model’s success, spending reductions are critical because policy makers increasingly judge the viability of bundled payment programs by their cost savings.

Second, these study findings affirm the suitability of hospitals to bundled payment models, specifically highlighting their relative advantage over PGP in achieving cost and potential quality outcomes for medical conditions. These findings contrast with prior research demonstrating that PGP may be more successful than hospitals at reducing spending in population-based payment models, such as acute care organizations. Future work should elucidate drivers underlying this distinction. For example, hospitals may be better positioned to coordinate with postacute care organizations such as skilled nursing facilities given their high volume of shared patients. These strategies may be particularly important for medical conditions where the episode cost savings come from reductions in postacute care facility length of stay rather than reductions in the proportion of individuals discharged. Policy makers may consider these facets of performance when considering participant types in future alternative payment models.

**Figure 2. Changes in Total Episode Spending, Readmissions, and Mortality for Surgical Episodes, by Study Group**

A, The preperiod means for total episode spending were: non-BPCI, $33,557; BPCI PGP, $30,987; BPCI hospitals, $33,135; BPCI both, $31,055. B, The preperiod means for readmissions were: non-BPCI, 14.2%; BPCI PGP, 10.7%; BPCI hospitals, 12.1%; BPCI both, 11.2%. C, The preperiod means for mortality were: non-BPCI, 2.8%; BPCI PGP, 2.6%; BPCI hospital, 3.3%; BPCI both, 2.5%. BPCI indicates Bundled Payments for Care Improvement; PGP, physician group practice.
Third, these study findings emphasize the need for future research on the drivers of cost savings and quality improvements under bundled payments. Our results regarding BPCI-participating PGP's point to the importance of changes in readmissions and postacute care utilization in determining episode savings. Yet as observed from BPCI-participating hospitals, different patterns of utilization changes may drive savings for different episode types. Specifically, in these findings medical episode savings were associated with reductions in length of stay within skilled nursing facilities, whereas surgical episode cost savings came from fewer discharges to skilled nursing facilities.

Additional work is also needed to assess the relationship between bundled payments and quality improvements. Although differential mortality reductions were observed by this study, there was also evidence of observable and unobservable favorable patient selection under bundled payments. This makes definitive conclusions regarding changes in health care quality challenging. Clarity on whether apparent quality changes represent true improvements, patient selection, or measures of both is highly relevant to policy and should be the focus of future studies.

Limitations
This study had some limitations worth noting. Findings may have been subject to residual confounding; however, we mitigated concerns by using a DID design that accounted for unobserved heterogeneity and patient and hospital characteristics. We evaluated the highest-volume episodes under a single program; however, BPCI model 2 was the direct basis for ongoing PGP and hospital participation in BPCI Advanced. Also, we did not include more recent data from BPCI Advanced because physician group participation files were not available. Moreover, we were unable to match physicians or identify episodes using tax identification number-level information owing to a lack of

Table 2. Characteristics of Patients Admitted for Surgical Episodes, by Participation in Bundled Payments, During Baseline Period

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Non-BPCI</th>
<th>BPCI PGP</th>
<th>BPCI hospital</th>
<th>BPCI both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episodes, No.</td>
<td>207,716</td>
<td>27,658</td>
<td>35,480</td>
<td>2,971</td>
</tr>
<tr>
<td>Patients, No.</td>
<td>205,986</td>
<td>27,624</td>
<td>35,446</td>
<td>2,970</td>
</tr>
<tr>
<td>Female sex, No. (%)</td>
<td>131,081</td>
<td>17,555</td>
<td>22,846</td>
<td>1,951</td>
</tr>
<tr>
<td>Male sex, No. (%)</td>
<td>76,635</td>
<td>10,103</td>
<td>12,634</td>
<td>1,020</td>
</tr>
<tr>
<td>Age, mean (SD) y</td>
<td>74.9 (9.6)</td>
<td>74.7 (9.0)</td>
<td>75.1 (9.5)</td>
<td>74.9 (8.7)</td>
</tr>
<tr>
<td>Race, No. (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>11,166 (5.4)</td>
<td>1293 (4.7)</td>
<td>2,170 (6.1)</td>
<td>196 (6.6)</td>
</tr>
<tr>
<td>White</td>
<td>184,823 (89.0)</td>
<td>25,057 (90.6)</td>
<td>30,890 (87.1)</td>
<td>2,640 (88.9)</td>
</tr>
<tr>
<td>Other†</td>
<td>11,727 (5.7)</td>
<td>1308 (4.7)</td>
<td>2,420 (6.8)</td>
<td>135 (4.5)</td>
</tr>
<tr>
<td>Disabled, No. (%)</td>
<td>18,987 (9.1)</td>
<td>2,051 (7.4)</td>
<td>2,924 (8.2)</td>
<td>194 (6.5)</td>
</tr>
<tr>
<td>Medicare/Medicaid dual eligible, No. (%)</td>
<td>27,923 (13.4)</td>
<td>3,041 (11.0)</td>
<td>4,643 (13.1)</td>
<td>298 (10.0)</td>
</tr>
<tr>
<td>Elixhauser comorbidity index, mean (SD), points</td>
<td>7.2 (11.9)</td>
<td>5.8 (11.1)</td>
<td>6.5 (11.4)</td>
<td>5.9 (11.2)</td>
</tr>
<tr>
<td>Prior use, mean (SD), %bh</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRF</td>
<td>2.2 (14.5)</td>
<td>1.4 (11.6)</td>
<td>1.9 (13.7)</td>
<td>1.5 (12.2)</td>
</tr>
<tr>
<td>SNF</td>
<td>7.7 (26.7)</td>
<td>6.1 (23.9)</td>
<td>7.2 (25.8)</td>
<td>7.1 (25.6)</td>
</tr>
<tr>
<td>Hospital</td>
<td>29.8 (45.7)</td>
<td>27.5 (44.7)</td>
<td>25.4 (43.6)</td>
<td>24.2 (42.8)</td>
</tr>
<tr>
<td>Episodes, No.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hip/femur procedure (not major joint)</td>
<td>24,295</td>
<td>3,687</td>
<td>3,412</td>
<td>219</td>
</tr>
<tr>
<td>Lower extremity joint replacement</td>
<td>147,774</td>
<td>19,619</td>
<td>27,782</td>
<td>2,662</td>
</tr>
<tr>
<td>Percutaneous coronary intervention</td>
<td>17,924</td>
<td>1,654</td>
<td>2,489</td>
<td>38</td>
</tr>
<tr>
<td>Spinal fusion</td>
<td>11,737</td>
<td>1,574</td>
<td>1,392</td>
<td>43</td>
</tr>
<tr>
<td>Upper extremity joint replacement</td>
<td>5,986</td>
<td>1,124</td>
<td>405</td>
<td>9</td>
</tr>
</tbody>
</table>

Abbreviations: BPCI, Medicare Bundled Payments for Care Improvement initiative; IRF, inpatient rehabilitation facility; PGP, physician group practice; SNF, skilled nursing facility.

† Includes Asian, Hispanic, Native American (American Indian, Alaska Native, Native Hawaiian, other Pacific Islander), and other responses as reported from the Medicare claims data.

‡ Use within the previous 12 months.
data availability. However, our approach using all episodes per National Provider Identification number and any participation in BPCI was conservatively biased toward the null hypothesis. Furthermore, to our knowledge, mortality reductions have not been previously described, and although we tested the robustness of the findings using a range of sensitivity and post hoc analyses, the analyses suggested the presence of unobservable patient selection that precluded definitive conclusions regarding any changes in health care quality. Despite conducting sensitivity analyses using alternative modeling approaches for episode spending as the primary outcome, future work should assess the use of modeling approaches beyond ordinary least-squares for other outcomes.

Conclusions

This cohort study with DID analysis found that PGP participation in BPCI was associated with cost savings for surgical episodes but not for medical episodes, whereas hospital participation in BPCI was associated with savings for both episode types. Policy makers should consider the comparative performance of participant type when designing and evaluating future bundled payment models.

ARTICLE INFORMATION

Accepted for Publication: November 4, 2022.

Published: December 29, 2022. doi:10.1001/jamahealthforum.2022.4889

Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2022 Liao JM et al. JAMA Health Forum.

Corresponding Author: Amol S. Navathe, MD, PhD, Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, 1108 Blockley Hall, 423 Guardian Dr, Philadelphia, PA 19104 (amol@pennmedicine.upenn.edu).

Author Affiliations: Department of Medicine, University of Washington School of Medicine, Seattle (Liao); Leonard Davis Institute of Health Economics, University of Pennsylvania, Philadelphia (Liao, Navathe); Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Huang, Wang, Shirk, Zhu, Cousins, Navathe); Department of Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia (Linn); Corporal Michael J. Crescenz Veterans Affairs Medical Center, Philadelphia, Pennsylvania (Navathe).

Author Contributions: Dr Navathe had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Liao, Huang, Navathe.

Acquisition, analysis, or interpretation of data: Liao, Huang, Wang, Linn, Shirk, Zhu, Cousins.

Drafting of the manuscript: Liao, Huang, Wang, Shirk, Cousins, Navathe.

Critical revision of the manuscript for important intellectual content: Liao, Huang, Linn, Zhu, Navathe.


Obtained funding: Cousins, Navathe.

Administrative, technical, or material support: Cousins, Navathe.

Supervision: Liao, Navathe.

Conflict of Interest Disclosures: Dr Liao reported personal fees/honoraria from Marcus Evans, Comagine Health, and Brown University outside the submitted work. Dr Linn reported grants from Commonwealth Fund and Robert Wood Johnson Foundation and personal fees from Correlation One outside the submitted work. Dr Navathe reported grants from the Hawaii Medical Service Association, the Commonwealth Fund, the Robert Wood Johnson Foundation, the Donaghue Foundation, the Pennsylvania Department of Health, the Veterans Affairs Administration, Ochsner Health System, United Healthcare, Blue Cross Blue Shield of North Carolina, Blue Shield of California, and Humana; personal fees from Navvis Healthcare, Yale New Haven Hospital/Center for Outcomes Research and Evaluation, Maine Health Accountable Care Organization, Singapore Ministry of Health, Elsevier Press, Medicare Payment Advisory Commission, Cleveland Clinic, the Analysis Group, the VBID Health, Advocate Physician Partners, the Federal Trade Commission, and the Catholic Health Services Long Island; equity from Nava
Health and Clarify Health; and noncompensated board membership from Integrated Services, all outside the submitted work. No other disclosures were reported.

**Funding/Support:** This study was funded by grants from the US National Institutes of Health's National Institute on Minority Health and Health Disparities (Nos. R01MD013859, R01MD015055) and the Agency for Healthcare Research and Quality (No. R01HS027595).

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication. This article does not necessarily represent the views of the US government, the US Department of Veterans Affairs, or the State of Pennsylvania.

**Data Sharing Statement:** See Supplement 2.

**REFERENCES**


SUPPLEMENT 1.
edMethods 1. Propensity Score Matching Approach
edMethods 2. Exploratory Study Outcomes
eFigure 1. Adjusted Parallel Trends for Medical Episodes
eFigure 2. Adjusted Parallel Trends for Surgical Episodes
edMethods 2. Testing for Time-Varying Confounders in Difference-in-Difference Design
edMethods 3. Model Specifications
table 1. Characteristics of BPCI and Non-BPCI Hospitals, Before and After Propensity Score Matching
table 2. Characteristics of Physicians in BPCI and Non-BPCI PGPs, Before and After Propensity Score Matching
table 3. Unadjusted Changes in Medical Episode Outcomes, 2011-2017
table 4. Characteristics of Surgical Episode Patients and Markets, by Study Period and Bundled Payment Status
table 5. Unadjusted Changes in Medical Episode Outcomes, Baseline versus Intervention Period
table 6. Percent Changes in Medical Episode Outcomes, Baseline versus Intervention Period
table 7. Adjusted Changes in Medical Episode Outcomes Associated with Bundled Payment Participation Among PGPs and Hospitals
table 9. Unadjusted Changes in Surgical Episode Outcomes, Baseline versus Intervention Period
table 10. Percent Changes in Surgical Episode Outcomes, Baseline versus Intervention Period
table 11. Adjusted Changes in Surgical Episode Outcomes Associated with Bundled Payment Participation Among PGPs and Hospitals
table 12. Adjusted Changes in Surgical Episode Exploratory Outcomes Associated With Bundled Payments
table 13. Sensitivity Analysis for Changes in Total Episode Spending in Medical Episodes, Using Generalized Linear Models With Log Link and Gamma Distribution
table 14. Sensitivity Analysis for Changes in Total Episode Spending in Surgical Episodes, Using Generalized Linear Models With Log Link and Gamma Distribution
table 15. Sensitivity Analysis for Changes in Total Episode Spending in Medical Episodes, Considering BPCI Both as BPCI PGP episodes
table 16. Sensitivity Analysis for Changes in Total Episode Spending in Surgical Episodes, Considering BPCI Both as BPCI PGP episodes
table 17. Sensitivity Analysis for Changes in 90-day Mortality in Medical Episodes, Including Episodes with Death at Index Hospitalization
table 18. Sensitivity Analysis for Changes in 90-day Mortality in Surgical Episodes, Including Episodes with Death at Index Hospitalization
eMethods 4. Assessment of the 90-day Mortality Outcome

SUPPLEMENT 2.
Data Sharing Statement