


Bundled Payment Episodes Initiated by Physician Group Practices: Medicare Beneficiary Perceptions of Care Quality



Sean R. McClellan, PhD¹ , Matthew J. Trombley, PhD², Jaclyn Marshall, MS³,
Daver Kahvecioglu, PhD⁴, Colleen M. Kummet, PhD⁵, Christine LaRocca, MD⁶,
Laura Dummit, MSPH⁷, and Andrea Hassol, MSPH¹

¹Abt Associates, Cambridge, MA, USA; ²Abt Associates, Durham, NC, USA; ³Bind Benefits, Minneapolis, MN, USA; ⁴Centers for Medicare & Medicaid Services, Baltimore, MD, USA; ⁵General Dynamics Information Technology, West Des Moines, IA, USA; ⁶Telligen, Greenwood Village, CO, USA; ⁷Lewin Group, Falls Church, VA, USA.

BACKGROUND: The Bundled Payments for Care Improvement (BPCI) initiative incentivizes participating providers to reduce total Medicare payments for an episode of care. However, there are concerns that reducing payments could reduce quality of care.

OBJECTIVE: To assess the association of BPCI with patient-reported functional status and care experiences.

DESIGN: We surveyed a stratified random sample of Medicare beneficiaries with BPCI episodes attributed to participating physician group practices, and matched comparison beneficiaries, after hospitalization for one of the 18 highest volume clinical episodes. The sample included beneficiaries discharged from the hospital from February 2017 through September 2017. Beneficiaries were surveyed approximately 90 days after their hospital discharge. We estimated risk-adjusted differences between the BPCI and comparison groups, pooled across all 18 clinical episodes and separately for the five largest clinical episodes.

PARTICIPANTS: Medicare beneficiaries with BPCI episodes ($n=16,898$, response rate=44.5%) and comparison beneficiaries hospitalized for similar conditions selected using coarsened exact matching ($n=14,652$, response rate=46.2%).

MAIN MEASURES: Patient-reported functional status, care experiences, and overall satisfaction with recovery.

KEY RESULTS: Overall, we did not find differences between the BPCI and comparison respondents across seven measures of change in functional status or overall satisfaction with recovery. Both BPCI and comparison respondents reported generally positive care experiences, but BPCI respondents were less likely to report positive care experience for 3 of 8 measures (discharged at the right time, -1.2 percentage points (pp); appropriate level of care, -1.8 pp; preferences for post-discharge care taken into account, -0.9 pp; $p<0.05$ for all three measures).

CONCLUSIONS: The proportion of respondents with favorable care experiences was smaller for BPCI than comparison respondents. However, we did not detect differences in self-reported change in functional status

approximately 90 days after hospital discharge, indicating that differences in care experiences did not affect functional recovery.

KEY WORDS: bundled payments; patient-reported outcomes; value-based payment models; Medicare; evaluation; survey research; functional status; care experiences.

J Gen Intern Med 37(5):1052–9

DOI: 10.1007/s11606-021-06848-9

© Abt Associates 2021

Bundled Payments for Care Improvement (BPCI) was a 5-year initiative of the Centers for Medicare & Medicaid Services (CMS) from 2013 to 2018, under the authority of the CMS Innovation Center, to test whether incentivizing reductions in Medicare payments for an episode of care triggered by a hospitalization and extending for up to 90 days after hospital discharge could reduce Medicare payments while maintaining quality of care.^{1,2} Under Model 2 of the initiative, hospitals and physician group practices (PGPs) could participate in any of 48 clinical episodes. Episodes were attributed to a participating PGP when a physician in the practice was the attending physician or operating surgeon listed on the inpatient claim, and, starting in 2016, a second criterion was added that a physician from the same practice also must treat the patient at least once during the episode after hospital discharge. If Medicare payments for the episode exceeded a target price, the BPCI participant was responsible for paying Medicare the difference, but if payments were less than the target price, the BPCI participant received the difference. In this way, participants had incentives to reduce service use to lower episode payments, raising concerns about whether reductions could jeopardize quality or patient experiences. Assessing the impact of bundled payment models is important given the continued industry-wide shift towards value-based care.^{3–6} A survey of commercial payers in 2018 indicated that 20% of their business was covered under bundled payments, a

Received January 5, 2021

Accepted April 22, 2021

Published online July 28, 2021

proportion that is expected to increase.⁷ In 2018, CMS introduced the BPCI Advanced model, which builds on BPCI Model 2.⁸

Most research to date on BPCI has been on hospitals and post-acute care providers.^{9–14} Only two peer-reviewed studies to date have reported findings related to PGP participation in BPCI, and both have focused on total joint arthroplasty.^{15,16} Under BPCI Model 2, 272 PGPs initiated nearly 300,000 episodes of care in the first four years of the initiative.¹⁷ BPCI-participating PGPs reduced average payments for 16 of 21 clinical episodes with sufficient sample size for evaluation, seven of which were statistically significant.¹⁷ Quality of care based on measures derived from claims indicated that quality was maintained for PGP-initiated episodes, relative to comparison episodes.¹⁷ Claims-based quality measures, however, do not holistically measure quality of care and are especially limited in reflecting patient perspectives.^{18–20} Patient-reported outcomes can help identify potential impacts of bundled payment initiatives caused by premature discharge or inadequate post-acute care that cannot be measured using claims data.²¹

For hospital-initiated episodes in BPCI Model 2, there were few differences in patient-reported functional status or care experiences between BPCI and comparison survey respondents,¹² but the impact of BPCI on these outcomes may not be the same for those with PGP-initiated episodes. Participating PGPs may have different resources and responses to bundled payments than participating hospitals. For example, PGPs may be more likely to have on-going relationships with patients, which could promote continuity of care.^{22–24} Hospitals may have greater resources to devote to enhanced care coordination.^{13,25} Additionally, hospitals and PGPs that are part of a health system may need to consider the financial impact of their responses across the broader enterprise in redesigning clinical pathways for bundled payments.^{26,27}

This study compares patient-reported functional outcomes and care experiences for Medicare fee-for-service beneficiaries in one of 18 clinical episodes initiated by BPCI Model 2 PGP participants, with the outcomes and experiences of a similar group of comparison beneficiaries. We separately analyzed the five highest-volume clinical episodes.

METHODS

Data

Survey Instrument. The BPCI beneficiary survey instrument contained 36 multiple-choice, closed-ended questions in five domains: changes in functional status, overall mental and physical health, care experiences and discharge planning, overall satisfaction with recovery, and personal characteristics. The survey instrument included items adapted from validated survey instruments, including the CARE Tool,²⁸ National Health Interview Survey,²⁹ Short Form 36 Health Survey,³⁰

and the Care Transitions Measure®.³¹ The survey underwent cognitive testing with a convenience sample of Medicare beneficiaries with recent hospital and PAC experience. Beneficiaries with BPCI episodes could only be identified after their hospitalization, so it was not possible to survey them before the episode. Instead, we surveyed beneficiaries only once, at the end of their episode, and asked them to report their functional status at two points in time: the week before their hospitalization and the day they completed the survey (received roughly 90 days after hospital discharge). While it is possible that patients may not have been able to recall their pre-hospitalization functional status with complete accuracy,^{32,33} some studies have found minimal recall bias up to three months following a major health event,^{34–36} and any recall bias would be similar for both the intervention and comparison groups. The survey instrument is available in the [supplementary appendix](#).

Survey Sample. We surveyed beneficiaries with the 18 highest-volume clinical episodes, which represented 89% of all BPCI Model 2 PGP episodes. We selected these 18 clinical episodes because we projected they would have a sufficient sample size. For each clinical episode, we selected a stratified random sample of beneficiaries in the intervention group, who had an episode attributed to a physician associated with a BPCI-participating PGP. Within each clinical episode, we defined sampling cells as unique combinations of presence/absence of a major complication or comorbidity, beneficiary age (<65, 65–74, 75–84, 85+), hospital size (above or below median number of beds), and hospital academic affiliations. We used coarsened exact matching to identify beneficiaries for the comparison group, randomly selecting comparison beneficiaries in equal proportions as the intervention group across cells.³⁷ The survey was mailed to 37,998 beneficiaries with a clinical episode initiated by a BPCI-participating PGP and 31,707 beneficiaries with a clinical episode initiated at comparison hospitals, from February to September 2017. Appendix A has additional information about the survey sample.

Survey Administration. Approximately 90 days after hospital discharge, we mailed sampled beneficiaries a survey and introductory cover letter, followed by a reminder postcard and a follow-up invitation letter with a second survey. We called the remaining non-respondents to complete the survey by phone. Survey data collection materials were approved by the Abt Associates IRB. The data generated during this study are not currently publicly available.

Patient-Reported Measures

For each functional status component, we created binary measures of improvement or maintenance of the highest functional status between respondents' recalled status and reported status at the time of the survey. The variable had a value of one if

beneficiaries reported improved functional status from before their episode to the time of the survey or if beneficiaries recalled having the highest functional status in the week prior to hospitalization and also at the time of the survey. The variable had a value of zero otherwise. We also constructed binary measures indicating affirmative outcomes for measures of care experience and overall satisfaction with recovery. Appendix B provides greater detail about these measures.

Analysis

We estimated the difference in outcomes between the BPCI and comparison samples using logistic regression, weighted to adjust for sampling and nonresponse, with standard errors clustered at the level of the discharging hospital. The association of BPCI with outcomes was estimated as the percentage point difference between the BPCI and comparison respondents. Outcomes were risk-adjusted for age, sex, Medicaid eligibility, hierarchical condition category (HCC) index, health care use in the 90 days before the episode, Medicare Severity-Diagnosis Related Group (MS-DRG), fracture (for major joint replacement of the lower extremity [MJRLE] episodes), characteristics of the discharging hospital (bed size, ownership type, academic status, participation in the Comprehensive Care for Joint Replacement model), the respondent's recalled functional status prior to hospitalization, whether a proxy responded to the survey, and survey wave.³⁸ For each outcome measure, we used *F*-tests to assess model fit. Although this study focuses on PGP-attributed episodes, we adjusted for characteristics of the discharging hospital, because all BPCI clinical episodes were initiated by a hospitalization, and patient-reported functional status and care experiences partly depend on attributes of the hospital where the episode began.

The primary analysis pooled all survey responses for the 18 available clinical episodes.¹² We also estimated differences in outcomes for each of the five largest clinical episodes, which accounted for approximately half of all Model 2 BPCI episodes initiated by PGPs: MJRLE; chronic obstructive pulmonary disease, bronchitis, and asthma (COPD); congestive heart failure (CHF); simple pneumonia and respiratory infections (pneumonia); and sepsis. Appendix C has additional information about the analytic methods.

Although this study includes 16 patient-reported measures in total, we did not adjust for multiple comparisons³⁹ because this would decrease the chance of identifying negative consequences associated with BPCI.⁴⁰

Sensitivity Analyses. We assessed the robustness of our empirical approach in multiple ways. First, not all BPCI beneficiaries in the survey sample received subsequent outpatient care from the PGP that initiated their episode following their hospitalization; it is possible that beneficiaries who did not have such a follow-up visit had different experiences than those that did continue to receive

post-hospital care from physicians in the same PGP. To assess this possibility, we conducted analyses including only beneficiaries who received follow-up during the episode from BPCI-participating practices (Appendix F). Second, we assessed the sensitivity of estimates when using alternative empirical specifications (regressions without weights, weighted differences with no regression adjustment, and raw differences with no weights or adjustment) (Appendix G). Third, the survey was designed to assess both improvement in functional status and patient experiences throughout a 90-day episode of care; because beneficiaries with BPCI episodes could not be identified before the episode began, we measured improvement in functional status using respondents' current self-reported functional status relative to their recalled pre-hospitalization functional status.³³ Given potential concerns about bias in recalled measures of functional status, we assessed the robustness of findings when using only measures of functional status as reported post-hospitalization, at the time of the survey (Appendix H). Lastly, we estimated changes in functional status stratified by pre-hospital function, to account for the possibility that the association between BPCI and probability of functional improvement varied based on respondents' initial function prior to the triggering hospital admission (Appendix I).

RESULTS

The overall response rate for the survey was 45.3% (44.5% for the BPCI group and 46.2% for the comparison group, $p < 0.01$). While the response rate was statistically significantly different between the BPCI and comparison groups, the magnitude of the difference was qualitatively small, less than two percentage points, and BPCI and comparison survey respondents were well-balanced on all demographic characteristics (Table 1). Respondents tended to be healthier than non-respondents, as indicated by differences in rates of hip fracture among MJRLE patients and the average HCC score (Appendix D). However, differences in patient characteristics between respondents and non-respondents were similar between the BPCI and comparison groups.

For the 18 combined clinical episodes, we did not detect significant differences between the BPCI and comparison respondents for any of seven functional status change measures (Table 2). For example, the same proportion of BPCI and comparison respondents reported improvement or maintenance of ability to walk without resting from before their hospitalization to the time of the survey. In addition, a majority of BPCI and comparison respondents reported positive care experiences and overall satisfaction with recovery. For example, 71% of respondents in both groups reported never receiving conflicting medical advice, and 95% of respondents in both groups reported having a good understanding of how to take care of themselves before going home. However, BPCI respondents were less likely to report positive care experience

Table 1 Characteristics of Survey Respondents*

Characteristic	Mean (SD) or %		Mean standardized difference†
	BPCI (N=16,898)	Comparison (N=14,652)	
Age, %			
< 65	11.4	11.7	0.011
65–74	39.0	38.7	0.005
75–84	30.5	30.5	0.000
>85	19.1	19.0	0.002
Female, %	58.4	58.2	0.005
Dually eligible for Medicare and Medicaid, %	14.9	15.7	0.023
Hierarchical condition category (HCC) index,‡ mean (SD)	0.47 (0.84)	0.45 (0.83)	0.016
Prior inpatient days, mean (SD)	0.79 (4.84)	0.52 (4.12)	0.060
Prior skilled nursing facility (SNF) days,§ mean (SD)	1.06 (6.60)	0.86 (6.28)	0.031
Admission from community, %	93.8	94.7	0.038
Major complication or comorbidity, %	55.1	55.1	0.001
Proxy respondent, %			
Response by beneficiary	75.5	74.7	0.018
Response by proxy	11.2	12.0	0.022
Response type missing	13.3	13.3	0.001
Baseline functional status index (5 items), mean (SD)			
Percent of items with high functional status	43.9 (33.7)	43.9 (33.9)	0.002
Percent of items with moderate functional status	37.9 (26.5)	38.0 (27.2)	0.007
Percent of items with low functional status	18.3 (27.7)	18.0 (28.1)	0.009
Missing functional status index, %	9.6	9.7	0.002
Baseline pain and physical or emotional problems index (2 items), mean (SD)			
Percent of items with pain and physical or emotional problems all or most of the time	65.4 (39.0)	65.4 (38.8)	0.000
Percent of items with pain and physical or emotional problems some, a little, or none of the time	34.6 (39.0)	34.6 (38.8)	0.000
Missing pain and physical or emotional problems index, %	4.5	4.3	0.009
Hip fracture (major joint replacement of the lower extremity only), %	9.4	9.1	0.011
Hip fracture (hip and femur only), %	87.1	83.7	0.118
Large vessel ischemic stroke (stroke only), %	19.7	17.3	0.061
Intracerebral hemorrhagic stroke (stroke only), %	6.0	6.1	0.004
Hospital ownership of the discharging hospital, %			
For-profit hospital	14.4	14.0	0.011
Not-for-profit hospital	65.5	64.8	0.013
Government/other hospital ownership	20.2	21.2	0.026
Discharging hospital was an academic medical center, %	45.9	46.3	0.008
Number of beds in the discharging hospital exceeds median, %	79.3	79.1	0.003

*All means in table were weighted by sampling design and nonresponse. Standard deviations for continuous measures are reported in parentheses. BPCI, Bundled Payments for Care Improvement. SD, standard deviation

†The mean standardized difference was calculated as the absolute difference between BPCI and comparison averages, divided by the pooled standard deviation of the outcome among both BPCI and comparison respondents

‡The HCC index in our sample reflects the portion of the CMS-HCC community risk score that corresponds to the HCCs present in the six months preceding hospital admission. HCC index values in our sample range from zero to 9.7 units

§Prior inpatient days and SNF days range from zero up to 90 days

for 3 of 9 measures: discharged at the right time (−1.2 percentage points [pp], 95% confidence interval [CI]: −2.1 to −0.3), appropriate level of care (−1.8 pp, 95% CI: −3.1 to −0.5), and preferences for post-discharge care taken into account (−0.9 pp, 95% CI: −1.7 to −0.1). Null results were precisely estimated. The 95% confidence intervals for our insignificant estimates did not exceed 2.1 percentage points in absolute magnitude, suggesting that if there were true underlying differences that we failed to detect, they were small in magnitude. These findings were broadly consistent with the sensitivity analyses (Appendix F-I).

Functional improvement, care experience, and overall satisfaction for each of the five largest clinical episodes were similar to the aggregated results, with a few notable exceptions (Table 3). BPCI respondents with episodes of MJRLE were more likely to report improvement in use of

stairs from before the surgery to after the episode than were comparison respondents (3.2 PP, 95% CI: 0.1 to 6.2). There were no other statistically significant differences in functional improvement between BPCI and comparison respondents for any of the five largest clinical episodes. BPCI respondents with COPD or pneumonia episodes were significantly less likely than comparison respondents to report having a good understanding of how to take care of themselves before going home (COPD: −2.1 pp, 95% CI: −4.1 to −0.1; pneumonia: −1.9 pp, 95% CI: −3.8 to 0.0), and less likely to report that medical staff clearly explained how to take medications before going home (COPD: −2.2 pp, 95% CI: −4.1 to −0.2; pneumonia: −3.3 pp, 95% CI: −5.5 to −1.1). Other statistically significant differences were consistent with the aggregate findings.

Table 2 Differences in Survey-based Quality Outcomes between BPCI and Comparison Respondents, Aggregate Model 2 Physician Group Practices, February 2017–September 2017*

Survey measure	Risk-adjusted percent of respondents			F-test§
	BPCI (N=16,898)	Comparison (N=14,652)	Difference (95% CI)	
Improvement or maintenance of highest status				
Bathing, dressing, using toilet, or eating	72.8	72.0	0.8 (−0.2 to 1.7)	22.23
Planning regular tasks	61.6	61.5	0.1 (−1.0 to 1.2)	24.96
Use of a mobility device (less likely to use)	49.9	50.4	−0.5 (−1.7 to 0.7)	28.06
Walking without rest	45.3	45.0	0.3 (−1.0 to 1.5)	20.78
Using stairs	45.0	43.8	1.2 (−0.2 to 2.5)	20.58
Physical/emotional problems limit social activities	59.7	60.5	−0.9 (−2.1 to 0.4)	14.14
Pain limits regular activities	57.3	57.3	0.0 (−1.3 to 1.2)	13.89
Care experience				
Never received conflicting medical advice	72.8	73.7	−0.9 (−2.1 to 0.3)	5.64
Services always appropriate for level of care needed	61.9	63.7	−1.8 (−3.1 to −0.5)†	6.51
Patient was discharged at the right time	88.5	89.7	−1.2 (−2.1 to −0.3)†	3.34
Medical staff took patient's preferences into account in deciding post-discharge health care services	92.0	92.9	−0.9 (−1.7 to −0.1)†	3.40
Patient had good understanding of how to take care of self before going home	94.6	94.9	−0.3 (−1.0 to 0.4)	3.23
Medical staff clearly explained how to take medications before going home	94.1	94.2	−0.1 (−0.9 to 0.6)	3.32
Medical staff clearly explained what follow-up appointments would be needed before patient went home	94.8	94.8	0.0 (−0.7 to 0.7)	3.09
Patient was able to manage health needs since returning home	95.9	96.1	−0.2 (−0.8 to 0.4)	3.96
Overall satisfaction				
Extremely or quite a bit satisfied with overall recovery	70.4	71.1	−0.8 (−2.0 to 0.5)	8.32

*Estimated differences between the BPCI and comparison groups were risk-adjusted using logistic regression and reported in percentage point (PP) terms. All regressions were weighted for sampling and non-response. Regressions controlled for sex; age; dual eligibility for Medicare and Medicaid; hierarchical condition category index; prior inpatient and skilled nursing facility days prior to admission; admission from community; major complication or comorbidity; hip fracture (for joint replacement or hip and femur episodes); type of stroke (for stroke episodes); self-reported pre-hospital functional status; response by proxy; Medicare Severity Diagnosis Related Group (MS-DRG); hospital ownership type, academic affiliation, and size; and survey wave fixed effects. BPCI, Bundled Payments for Care Improvement

†Difference between BPCI and comparison respondents was statistically significant at $p < 0.05$

§F-test values for all outcomes were statistically significant at $p < 0.0001$

DISCUSSION

This study evaluated patient-reported experiences with BPCI Model 2 PGP-attributed episodes and included over 16,000 responses from beneficiaries treated by PGPs participating in the BPCI initiative and over 14,000 responses from a matched comparison group. Over 60% of both BPCI and comparison respondents—and as high as 96% for some measures—reported favorable care experiences and the highest levels of overall satisfaction with recovery. BPCI and comparison respondents reported similar care experience on five of eight measures and similar overall satisfaction with recovery. BPCI respondents, however, were less likely than comparison respondents to report being discharged at the right time, having an appropriate level of care, and that their preferences for post-discharge care were taken into account, with differences ranging from one to two percentage points. Since we used binary measures of care experience in this study, the findings imply substantive differences in care experience associated with BPCI that affected relatively few patients, rather than small differences that affected all patients.

We did not observe differences in self-reported functional outcomes about 90 days after hospital discharge between the two groups, under multiple specifications, indicating that concerns about care experiences among BPCI respondents did not

coincide with worse functional recovery relative to comparison respondents. These results were also consistent across the five highest volume clinical episodes. Regardless of the type of episode, BPCI was not associated with differences in patient-reported functional improvement and was associated with worse care experiences on only a few measures.

Our results are similar to findings among beneficiaries with BPCI episodes initiated by acute care hospitals under Model 2. Prior analyses of beneficiaries with episodes initiated by hospitals did not find differences in functional improvement between BPCI respondents and a matched comparison group, but found that BPCI respondents were significantly less likely than comparison respondents to report positive care experiences and high overall satisfaction with recovery.¹² Differences in care experiences between BPCI and comparison respondents were slightly larger and more likely to be significant for those whose episodes were initiated by hospital participants than by PGPs, and differences in satisfaction with recovery were not significant for those with PGP-initiated episodes. This suggests that participating PGPs may have been more effective in communicating with patients, preparing them for hospital discharge, and setting expectations about post-hospital care. Together the two studies suggest that BPCI Model 2 did not adversely impact patient functional status.

Table 3 Differences in Survey-based Quality Outcomes between BPCI and Comparison Respondents, Five Largest Clinical Episodes in Model 2 Physician Group Practices*

Survey measure	Risk-adjusted difference between BPCI and comparison respondents (95% CI)				
	COPD (BPCI n=1455)	CHF (BPCI n=1594)	MJRLE (BPCI n=1732)	Sepsis (BPCI n=2265)	Pneumonia (BPCI n=1447)
Improvement or maintenance of highest status					
Bathing, dressing, using toilet, or eating	0.8 (−2.3 to 3.9)	−0.8 (−3.9 to 2.3)	1.6 (−0.4 to 3.6)	0.1 (−2.7 to 3.0)	−0.4 (−3.4 to 2.6)
Planning regular tasks	−0.1 (−3.4 to 3.3)	0.4 (−2.7 to 3.6)	1.6 (−0.8 to 4.1)	−1.7 (−4.5 to 1.0)	−2.3 (−5.5 to 0.8)
Use of a mobility device (less likely to use)	−2.3 (−5.1 to 0.5)	−1.0 (−3.6 to 1.7)	0.5 (−2.4 to 3.5)	−1.7 (−4.1 to 0.8)	−1.8 (−4.5 to 0.9)
Walking without rest	1.0 (−2.3 to 4.3)	0.1 (−3.0 to 3.1)	0.6 (−2.4 to 3.7)	0.9 (−2.0 to 3.8)	−0.5 (−3.9 to 2.9)
Using stairs	2.1 (−1.3 to 5.5)	−0.4 (−3.7 to 2.9)	3.2 (0.1 to 6.2)†	−0.5 (−3.4 to 2.4)	−1.8 (−5.6 to 2.0)
Physical/emotional problems limit social activities	1.9 (−2.0 to 5.7)	−2.1 (−5.9 to 1.7)	−1.7 (−4.5 to 1.1)	1.2 (−2.3 to 4.6)	−3.4 (−7.2 to 0.5)
Pain limits regular activities	0.3 (−3.4 to 4.0)	2.0 (−1.7 to 5.7)	−0.9 (−3.7 to 1.8)	0.6 (−2.7 to 3.9)	−1.9 (−6.1 to 2.3)
Care experience					
Never received conflicting medical advice	−3.6 (−7.3 to 0.1)	−0.9 (−4.7 to 2.9)	1.0 (−1.5 to 3.5)	−1.1 (−4.6 to 2.5)	−2.0 (−5.9 to 2.0)
Services always appropriate for level of care needed	−2.3 (−6.1 to 1.5)	−3.4 (−7.4 to 0.6)	−0.9 (−3.8 to 2.1)	−2.8 (−6.4 to 0.7)	−4.0 (−8.3 to 0.3)
Patient was discharged at the right time	−2.2 (−4.9 to 0.5)	−1.6 (−4.4 to 1.2)	−0.6 (−2.5 to 1.3)	−2.9 (−5.3 to −0.4)†	−0.4 (−3.1 to 2.4)
Medical staff took patient's preferences into account in deciding post-discharge health care services	−1.1 (−3.5 to 1.3)	−1.0 (−3.3 to 1.2)	−2.1 (−3.8 to −0.4)†	1.4 (−0.9 to 3.7)	−2.2 (−4.7 to 0.2)
Patient had good understanding of how to take care of self before going home	−2.1 (−4.1 to −0.1)†	−0.4 (−2.3 to 1.4)	−0.3 (−1.7 to 1.1)§	0.2 (−2.0 to 2.3)	−2.2 (−4.1 to −0.2)†
Medical staff clearly explained how to take medications before going home	−1.9 (−3.8 to 0.0)†§	−0.7 (−2.7 to 1.3)§	0.7 (−0.7 to 2.2)	−1.0 (−3.1 to 1.2)	−3.3 (−5.5 to −1.1)†
Medical staff clearly explained what follow-up appointments would be needed before patient went home	−1.5 (−3.4 to 0.4)	−0.2 (−2.1 to 1.8)	0.8 (−0.5 to 2.1)	−1.2 (−3.3 to 0.8)	−1.6 (−3.5 to 0.3)
Patient was able to manage health needs since returning home	−1.1 (−2.8 to 0.5)	−1.2 (−2.9 to 0.6)	0.0 (−1.2 to 1.2)	−0.2 (−2.0 to 1.5)	−1.7 (−3.6 to 0.1)
Overall satisfaction					
Extremely or quite a bit satisfied with overall recovery	−1.7 (−5.6 to 2.1)	−0.1 (−4.1 to 3.9)	−0.2 (−2.9 to 2.5)	−3.2 (−6.8 to 0.3)	−1.6 (−5.5 to 2.3)

*Estimated differences between the BPCI and comparison groups were risk-adjusted using logistic regression and reported in percentage point terms. All regressions were weighted for sampling and non-response. Regressions controlled for sex; age; dual eligibility for Medicare and Medicaid; hierarchical condition category index; prior inpatient and skilled nursing facility days prior to admission; admission from community; major complication or comorbidity; hip fracture (for joint replacement or hip and femur episodes); type of stroke (for stroke episodes); self-reported pre-hospital functional status; response by proxy; Medicare Severity Diagnosis Related Group (MS-DRG); hospital ownership type, academic affiliation, and size; and survey wave fixed effects. BPCI, Bundled Payments for Care Improvement. CHF, congestive heart failure. COPD, chronic obstructive pulmonary disease. MJRLE, major joint replacement of the lower extremity. Pneumonia, simple pneumonia and respiratory infections

†Difference between BPCI and comparison respondents was statistically significant at $p < 0.05$

§Regression F-test was not statistically significant ($p > 0.05$). F-tests for all other outcomes were statistically significant at $p < 0.05$

However, both PGPs and hospitals participating in bundled payment initiatives may have room for improvement with regard to the patient experience of care.

This study has limitations. First, because data collection began after the start of the BPCI initiative, this study relies on a post-only with comparison group study design, and we cannot know whether patient-reported outcomes differed between BPCI and comparison participants prior to the initiative. Pre-existing differences may have contributed to the estimated differences between BPCI and

comparison respondents, and the lack of significant differences for most outcomes does not preclude the possibility that BPCI impacted these outcomes for better or worse. Second, because BPCI was a voluntary initiative, these results might not be generalizable to all Medicare fee-for-service beneficiaries treated by all physician groups.^{41–43} Third, approximately half of the sampled beneficiaries did not complete the survey. Response rates and beneficiary characteristics were similar for the BPCI and comparison groups, and weights and risk-adjustment

helped account for non-response. However, both BPCI and comparison respondents tended to be somewhat healthier than non-respondents on average. While we adjusted for health status, to the extent that non-respondents differed from respondents on other unobservable factors, findings may not generalize to all BPCI beneficiaries. Fourth, PGPs participating in BPCI carefully verified the NPIs attributed to their practice, but it was not possible to have non-participants verify NPI lists in the same way; for this reason, we were not able to report or control for the characteristics of PGPs. However, physician specialty and number of physicians participating per BPCI-participating PGP were similar between BPCI survey respondents and non-respondents, except that beneficiaries with episodes attributed to surgeons were more likely to respond to the survey than beneficiaries with episodes initiated by other specialists, primary care physicians, or hospitalists (Appendix E). We adjusted for this difference using non-response weights and clinical episode fixed effects. Lastly, survey data collection only covered approximately one out of five years of BPCI. Given the possibility for attrition of PGPs prior to our survey, as well as changes to the PGP assignment algorithm made after our survey, our results may not generalize to the entire period covered by the model.

In conclusion, survey respondents in bundled payment episodes attributed to BPCI Model 2 PGP participants and a matched comparison group reported similar changes in functional status from before to after their care episodes, but a smaller proportion of BPCI respondents reported positive care experiences for three of eight measures, relative to the comparison group. These results alleviate concerns that episode-based payment models could adversely affect patient-reported functional outcomes among Medicare beneficiaries, but highlight the importance patients place on clear and consistent communication with providers.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11606-021-06848-9>.

Acknowledgements: This study was supported by the Centers for Medicare & Medicaid Services (CMS), under the Evaluation and Monitoring of the Bundled Payments for Care Improvement Initiative, Models 2–4 (contract number HHSM-500-2011-000011, task order HHSM-500-T0007). The evaluation was conducted independently; while CMS retains a contractual right to review the results and conclusion of our manuscript for consistency with publicly available results, CMS did not censor the results or conclusions of this study. This paper was presented at the 2019 AcademyHealth Annual Research Meeting (Washington, DC). We acknowledge Alisha Creel, formerly of the Abt Associates Data Science, Surveys, and Enabling Technologies (DSET) Division, for support in collecting the survey data used for this study. We also acknowledge Grecia Marrufo (of The Lewin Group) and Qian Gu (formerly of Abt Associates) for important contributions to the study design and analysis.

Corresponding Author: Sean R. McClellan, PhD; Abt Associates, Cambridge, MA, USA (e-mail: Sean_McClellan@abtassoc.com).

Declarations:

Conflict of interest: The study team has no other conflicts of interest to report.

REFERENCES

1. Press MJ, Rajkumar R, Conway PH. Medicare's New Bundled Payments: Design, Strategy, and Evolution. *JAMA*. 2016;315(2):131-2. <https://doi.org/10.1001/jama.2015.18161>
2. Centers for Medicare & Medicaid Services. Bundled Payments for Care Improvement Initiative (BPCI) Background on Model 2 for Prospective Participants. Accessed 29 August, 2015. https://innovation.cms.gov/Files/x/BPCI_Model2Background.pdf
3. Burwell SM. Setting value-based payment goals—HHS efforts to improve U.S. health care. *N Engl J Med*. 2015;372(10):897-9. <https://doi.org/10.1056/NEJMp1500445>
4. Navathe AS, Emanuel EJ, Bond A, et al. Association Between the Implementation of a Population-Based Primary Care Payment System and Achievement on Quality Measures in Hawaii. *JAMA*. 2019;322(1):57-68. <https://doi.org/10.1001/jama.2019.8113>
5. Baum A, Song Z, Landon BE, Phillips RS, Bitton A, Basu S. Health Care Spending Slowed After Rhode Island Applied Affordability Standards To Commercial Insurers. *Health Aff (Millwood)* 2019;38(2):237-245. <https://doi.org/10.1377/hlthaff.2018.05164>
6. Barrett KS, Record KL, Haime V. A state-led framework for value-based purchasing to incent integration. *Healthc (Amst)* 2018;6(2):101-103. <https://doi.org/10.1016/j.hjdsi.2017.05.002>
7. Finding The Value In Value-Based Care: The State of Value-Based Care in 2018. Accessed June. https://img04.en25.com/Web/ChangeHealthcare/%7B36f2f1b6-3275-4a4e-a065-171a37b83cc0%7D_Change_Healthcare_2018_AHIP_Presentation.pdf
8. Centers for Medicare & Medicaid Services. Bundled Payments for Care Improvement Advanced (BPCI Advanced) Model. Centers for Medicare & Medicaid Services. Accessed November 27, 2019. <https://innovation.cms.gov/initiatives/bpci-advanced>
9. Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. Evaluation of Medicare's Bundled Payments Initiative for Medical Conditions. *N Engl J Med*. 2018;379(3):260-269. <https://doi.org/10.1056/NEJMsa1801569>
10. Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. How Do Frail Medicare Beneficiaries Fare Under Bundled Payments? *J Am Geriatr Soc*. 2019. <https://doi.org/10.1111/jgs.16147>
11. Maughan BC, Kahvecioglu DC, Marrufo G, et al. Medicare's Bundled Payments For Care Improvement Initiative Maintained Quality Of Care For Vulnerable Patients. *Health Aff (Millwood)* 2019;38(4):561-568. <https://doi.org/10.1377/hlthaff.2018.05146>
12. Trombley MJ, McClellan SR, Kahvecioglu DC, et al. Association of Medicare's Bundled Payments for Care Improvement initiative with patient-reported outcomes. *Health Serv Res*. 2019. <https://doi.org/10.1111/1475-6773.13159>
13. Zhu JM, Patel V, Shea JA, Neuman MD, Werner RM. Hospitals Using Bundled Payment Report Reducing Skilled Nursing Facility Use And Improving Care Integration. *Health Aff (Millwood)* 2018;37(8):1282-1289. <https://doi.org/10.1377/hlthaff.2018.0257>
14. Yee CA, Pizer SD, Frakt A. Medicare's Bundled Payment Initiatives for Hospital-Initiated Episodes: Evidence and Evolution. *Milbank Q* 2020;98(3):908-974. <https://doi.org/10.1111/1468-0009.12465>
15. Siddiqi A, White PB, Murphy W, Terry D, Murphy SB, Talmo CT. Cost Savings in a Surgeon-Directed BPCI Program for Total Joint Arthroplasty. *Surg Technol Int* 2018;33:319-325.
16. Manner PA. Editor's Spotlight/Take 5-2018 John Charnley Award: Analysis of US Hip Replacement Bundled Payments: Physician-initiated Episodes Outperform Hospital-initiated Episodes. *Clin Orthop Relat Res* 2019;477(2):268-270. <https://doi.org/10.1097/CORR.0000000000000626>
17. Dummit LA, Marrufo G, Marshall J, et al. CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 5 Evaluation & Monitoring Annual Report. 2018. October. <https://downloads.cms.gov/files/cmml/bpci-models-2-4-yr5evalrpt.pdf>
18. Fisher ES. Medicare's Bundled Payment Program for Joint Replacement: Promise and Peril? *JAMA*. 2016;316(12):1262-4. <https://doi.org/10.1001/jama.2016.12525>

19. **Wu AW, Snyder C, Clancy CM, Steinwachs DM.** Adding the patient perspective to comparative effectiveness research. *Health Aff (Millwood)* 2010;29(10):1863-71. <https://doi.org/10.1377/hlthaff.2010.0660>
20. **Bentler SE, Morgan RO, Virnig BA, Wolinsky FD.** Do claims-based continuity of care measures reflect the patient perspective? *Med Care Res Rev* 2014;71(2):156-73. <https://doi.org/10.1177/1077558713505909>
21. **Berwick DM.** What 'patient-centered' should mean: confessions of an extremist. *Health Aff (Millwood)* 2009;28(4):w555-65. <https://doi.org/10.1377/hlthaff.28.4.w555>
22. **Hussey PS, Schneider EC, Rudin RS, Fox DS, Lai J, Pollack CE.** Continuity and the costs of care for chronic disease. *JAMA Intern Med* 2014;174(5):742-8. <https://doi.org/10.1001/jamainternmed.2014.245>
23. **Nyweide DJ, Anthony DL, Bynum JP, et al.** Continuity of care and the risk of preventable hospitalization in older adults. *JAMA Intern Med.* 2013;173(20):1879-85. <https://doi.org/10.1001/jamainternmed.2013.10059>
24. **Wolinsky FD, Bentler SE, Liu L, et al.** Continuity of care with a primary care physician and mortality in older adults. *J Gerontol A Biol Sci Med Sci* 2010;65(4):421-8. <https://doi.org/10.1093/gerona/glp188>
25. **Iorio R, Clair AJ, Inneh IA, Slover JD, Bosco JA, Zuckerman JD.** Early Results of Medicare's Bundled Payment Initiative for a 90-Day Total Joint Arthroplasty Episode of Care. *J Arthroplast* 2016;31(2):343-50. <https://doi.org/10.1016/j.arth.2015.09.004>
26. **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):1139-1149. <https://doi.org/10.1056/NEJMsa1803388>
27. **Liao J, Chen C.** Getting "Bundle Ready": How Provider Organizations Can Design Clinical Pathways in the Context of Bundled Payments. *J Clin Pathways* 2019;5(1):49-51.
28. **Gage B, Deutsch A, Smith L, et al.** The Development and Testing of the Continuity Assessment Record and Evaluation (CARE) Item Set: Final Report on CARE Item Set and Current Assessment Comparisons. 2012.
29. **Blackwell DL, Lucas JW, Clarke TC.** Summary health statistics for U.S. adults: Natinal Health Interview Survey, 2012. 2014. Vital and Health Statistics. https://www.cdc.gov/nchs/data/series/sr_10/sr10_260.pdf
30. **Brazier JE, Harper R, Jones NM, et al.** Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ.* 1992;305(6846):160-4.
31. **Coleman EA, Mahoney E, Parry C.** Assessing the quality of preparation for posthospital care from the patient's perspective: the care transitions measure. *Med Care* 2005;43(3):246-55.
32. **Mancuso CA, Charlson ME.** Does recollection error threaten the validity of cross-sectional studies of effectiveness? *Med Care* 1995;33(4 Suppl):AS77-88.
33. **Fischer D, Stewart AL, Bloch DA, Lorig K, Laurent D, Holman H.** Capturing the patient's view of change as a clinical outcome measure. *JAMA.* 1999;282(12):1157-62. <https://doi.org/10.1001/jama.282.12.1157>
34. **Howell J, Xu M, Duncan CP, Masri BA, Garbuz DS.** A comparison between patient recall and concurrent measurement of preoperative quality of life outcome in total hip arthroplasty. *J Arthroplast* 2008;23(6):843-9. <https://doi.org/10.1016/j.arth.2007.07.020>
35. **Middel B, Goudriaan H, de Greef M, et al.** Recall bias did not affect perceived magnitude of change in health-related functional status. *J Clin Epidemiol* 2006;59(5):503-11. <https://doi.org/10.1016/j.jclinepi.2005.08.018>
36. **Schmier JK, Halpern MT.** Patient recall and recall bias of health state and health status. *Expert Rev Pharmacoecon Outcomes Res* 2004;4(2):159-63. <https://doi.org/10.1586/14737167.4.2.159>
37. **Iacus SM, King G, Porro G.** Causal inference without balance checking: Coarsened exact matching. *Polit Anal* 2012;1(1):1-24.
38. **Horton NJ, Fitzmaurice GM.** Regression analysis of multiple source and multiple informant data from complex survey samples. *Stat Med.* 2004;23(18):2911-33. <https://doi.org/10.1002/sim.1879>
39. **Hochberg Y, Tamhane AC.** Multiple comparison procedures. Wiley series in probability and mathematical statistics Applied probability and statistics. Wiley; 1987:xxii, 450 p.
40. **Bilinski A, Hatfield L.** Seeking evidence of absence: Reconsidering tests of model assumptions. *arXiv preprint.* 2018;arXiv:1805.03273
41. **Smith B.** CMS Innovation Center at 10 Years - Progress and Lessons Learned. *N Engl J Med.* 2021. <https://doi.org/10.1056/NEJMsb2031138>
42. **Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM.** Participation and Dropout in the Bundled Payments for Care Improvement Initiative. *JAMA.* 2018;319(2):191-193. <https://doi.org/10.1001/jama.2017.14771>
43. **Oseran AS, Howard SE, Blumenthal DM.** Factors Associated With Participation in Cardiac Episode Payments Included in Medicare's Bundled Payments for Care Improvement Initiative. *JAMA Cardiol.* 2018. <https://doi.org/10.1001/jamacardio.2018.1736>

Publisher's Note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.