



OPEN ACCESS



Check for updates

# Spending and quality after three years of Medicare's bundled payments for medical conditions: quasi-experimental difference-in-differences study

Joshua A Rolnick,<sup>1,2,3</sup> Joshua M Liao,<sup>4,5</sup> Ezekiel J Emanuel,<sup>6</sup> Qian Huang,<sup>6</sup> Xinshuo Ma,<sup>6</sup> Eric Z Shan,<sup>6</sup> Claire Dinh,<sup>6</sup> Jingsan Zhu,<sup>6</sup> Erkuang Wang,<sup>6</sup> Deborah Cousins,<sup>6</sup> Amol S Navathe<sup>1,5,6</sup>

For numbered affiliations see end of the article.

Correspondence to: A S Navathe  
amol@wharton.upenn.edu  
(or @AmolNavathe on Twitter  
ORCID 0000-0001-7182-4988)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;369:m1780  
<http://dx.doi.org/10.1136/bmj.m1780>

Accepted: 7 April 2020

## ABSTRACT

### OBJECTIVE

To evaluate whether longer term participation in the bundled payments for care initiative (BPCI) for medical conditions in the United States, which held hospitals financially accountable for all spending during an episode of care from hospital admission to 90 days after discharge, was associated with changes in spending, mortality, or health service use.

### DESIGN

Quasi-experimental difference-in-differences analysis.

### SETTING

US hospitals participating in bundled payments for acute myocardial infarction, congestive heart failure, chronic obstructive pulmonary disease (COPD), or pneumonia, and propensity score matched to non-participating hospitals.

### PARTICIPANTS

238 hospitals participating in the Bundled Payments for Care Improvement initiative (BPCI) and 1415 non-BPCI hospitals. 226 BPCI hospitals were matched to 700 non-BPCI hospitals.

### MAIN OUTCOME MEASURES

Primary outcomes were total spending on episodes and death 90 days after discharge. Secondary outcomes included spending and use by type of post-acute care. BPCI and non-BPCI hospitals were compared by patient, hospital, and hospital market characteristics. Market characteristics included population size, competitiveness, and post-acute bed supply.

## RESULTS

In the 226 BPCI hospitals, episodes of care totaled 261 163 in the baseline period and 93 562 in the treatment period compared with 211 208 and 78 643 in the 700 matched non-BPCI hospitals, respectively, with small differences in hospital and market characteristics after matching. Differing trends were seen for some patient characteristics (eg, mean age change −0.3 years at BPCI hospitals v non-BPCI hospitals,  $P<0.001$ ). In the adjusted analysis, participation in BPCI was associated with a decrease in total episode spending (−1.2%, 95% confidence interval −2.3% to −0.2%). Spending on care at skilled nursing facilities decreased (−6.3%, −10.0% to −2.5%) owing to a reduced number of facility days (−6.2%, −9.8% to −2.6%), and home health spending increased (4.4%, 1.4% to 7.5%). Mortality at 90 days did not change (−0.1 percentage points, 95% confidence interval −0.5 to 0.2 percentage points).

## CONCLUSIONS

In this longer term evaluation of a large national programme on medical bundled payments in the US, participation in bundles for four common medical conditions was associated with savings at three years. The savings were generated by practice changes that decreased use of high intensity care after hospital discharge without affecting quality, which also suggests that bundles for medical conditions could require multiple years before changes in savings and practice emerge.

## Introduction

Bundled payments are a common form of value based payment.<sup>1-3</sup> Several countries, including the United States, Germany, Sweden, and the Netherlands, have developed bundled payments programmes, in which healthcare organizations receive a fixed payment for services provided in discrete care episodes and are held financially accountable for spending on those episodes.<sup>4-6</sup> The US recently completed a five year trial of the bundled payments for care improvement (BPCI) initiative, a national voluntary programme for hospitals that engaged 1025 total participants<sup>7</sup> in bundling 48 different clinical episodes under four distinct models. The stated purpose of the US Medicare programme was to reduce total healthcare spending on patients with included clinical episodes while maintaining or improving quality through reducing unnecessary care and improving care coordination.<sup>8</sup> BPCI applied to patients in the traditional Medicare programme, a national government initiative covering

## WHAT IS ALREADY KNOWN ON THIS TOPIC

The United States, Germany, the Netherlands, and other countries have tested bundling payments to physicians and hospitals in caring for patients with medical conditions

The US recently completed a national voluntary five year medical bundles programme for hospitals and physician groups (the bundled payments for care improvement initiative)

Preliminary evidence from year 1 found that hospital participation in the US programme was not associated with any changes in outcomes (spending, mortality, readmission) or practice patterns, but longer term independent evaluation is lacking

## WHAT THIS STUDY ADDS

Hospital participation in bundled payments for acute myocardial infarction, congestive heart failure, chronic obstructive pulmonary disease, and pneumonia after three years of the US programme was associated with a decrease in total spending through reductions in spending on skilled nursing in the 90 days after hospital admission in favor of increased home health spending

Mortality did not change

40 million older ( $\geq 65$  years of age) or disabled Americans.

BPCI participation was the highest for BPCI model 2, in which episodes started with a hospital admission and extended for up to 90 days after discharge. For example, in an episode of congestive heart failure, a patient could have been admitted for heart failure, been discharged to a skilled nursing facility for two weeks of rehabilitation, had multiple office visits with a primary care physician and cardiologist after returning home, and then been readmitted to hospital for heart failure two months after initial discharge. Traditionally, those services were paid separately, no adjustments were made for the total amount spent, and no clinical entity was accountable for the cost or quality outcomes. The fundamental change in the BPCI programme was that the initial hospital became responsible for total spending and quality outcomes during the episode, and Medicare awarded a bonus or penalty to it based on how spending compared with its historical episode spending. After BPCI ended in September 2018, the Centers for Medicare and Medicaid Services used model 2 to design and launch the larger nationwide BPCI-Advanced programme—a successor initiative that has engaged more than 2000 participant hospitals and physician groups<sup>9</sup> and started with 33 clinical episodes over a five year period.<sup>10</sup>

Despite the US decision to move forward with a new bundled payment programme for medical and surgical episodes, longer term evidence on the impact of bundled payments remains limited, particularly for patients admitted to hospital under medical episodes. Most bundled payment evaluations have focused on surgical episodes, with hospitals participating in hip and knee joint replacement bundles showing about 4% episode savings without changes in quality, volume, or case mix.<sup>11–13</sup> Though medical episodes account for eight of the top 10 episodes by number of total episodes, knowledge about the impact of medical bundles on outcomes is limited to US governmental contractor evaluations (which showed episode savings for some medical conditions but not for others) and a single peer reviewed study of short term outcomes (which did not find an association between participation in medical bundles and changes in outcomes over an average of seven months' follow-up).<sup>14–20</sup>

Consequently, an independent, long term evaluation of medical bundles is critical to inform policy making.<sup>21</sup> The effects of these bundles might change over time, similar to how financial savings in US Accountable Care Organization models<sup>22–23</sup> and quality improvements in US joint replacement bundles emerged over time, or how quality improvements in the UK primary care pay-for-performance programme attenuated over time.<sup>24</sup> Evaluations of impact over time are particularly important for payment models designed around chronic diseases, such as bundles for congestive heart failure or chronic obstructive pulmonary disease (COPD), which might require substantial redesign of practice and investments in care coordination.

We examined the long term association between hospital participation in BPCI for four high volume medical bundles—acute myocardial infarction, congestive heart failure, chronic obstructive pulmonary disease (COPD), and pneumonia—and changes in spending, mortality, and use over more than three years. In doing so, our approach also dealt with important limitations of previous work, such as extending the baseline period (to avoid contamination from pre-programme changes immediately before BPCI as organizations prepared for contract initiation), removing episodes attributed to BPCI physician group participants from the comparison group (to mitigate bias toward the null and reflect BPCI programme rules), allowing for time varying entry of BPCI participation (to analyze effects based on actual timing of participation), and addressing potential bias from programme rules on construction of episodes.<sup>25</sup>

## Methods

### Data and study period

To identify hospitals participating in BPCI model 2 (BPCI hospitals) and dates of entry by hospital condition, we used publicly available participation files.<sup>26</sup> Our study period spanned 1 January 2011 to 31 December 2016 and comprised a baseline period (January 2011 to September 2013) and treatment period (October 2013 to December 2016). We selected non-participating hospitals (non-BPCI hospitals) from markets with no hospital participants for the four conditions in this study—acute myocardial infarction, congestive heart failure, chronic obstructive pulmonary disease (COPD), and pneumonia (non-BPCI markets), to minimize bias from spillover effects (for example, if BPCI hospitals selected healthier patients, non-BPCI hospitals in the same markets would experience a change in case mix, which could occur on patient characteristics not observable in claims data). Markets with BPCI hospitals were defined as BPCI markets. We used hospital referral regions to define markets.<sup>27</sup>

Our study used 100% Medicare claims data reflecting payments to facilities (Part A) and physicians and non-hospital providers (Part B) for beneficiaries admitted to hospital at BPCI hospitals for one of the four study conditions, and a 20% national sample for beneficiaries admitted to a non-BPCI hospital from 2011 to 2016. Data from the 2011 American Hospital Association annual survey and 2011–16 Medicare Provider of Service, Beneficiary Summary, and Accountable Care Organization files were used to obtain market and hospital characteristics.<sup>28–29</sup> Provider Enrollment, Chain, and Ownership System data were used to identify BPCI physician group participants.

### Study sample

The study sample included Medicare fee-for-service beneficiaries admitted to hospital for acute myocardial infarction (diagnosis related group codes 280–282), congestive heart failure (291–293), COPD (190–192, 202, 203), and pneumonia (177–179 and 193–195).

These conditions were chosen as the four medical bundles with the highest participation rate in the second year of the programme (when study data acquisition began), and together reflect four of the highest volume medical bundles.

The BPCI hospital group included patients receiving care for these four conditions at BPCI hospitals under model 2. We focused on model 2 because it is the most commonly selected model in BPCI and served as the basis for the BPCI-Advanced programme. Although BPCI allowed participants to select from a post-hospital admission episode duration of 30, 60, or 90 days, we used the 90 day duration because more than 96% of participants selected it. We excluded non-BPCI hospitals if they had fewer than 10 hospital admissions for the diagnoses of interest during the treatment period and excluded patients with end stage renal disease and those who died during the index hospital admission or lacked continuous primary Medicare coverage during the episode.

We constructed 90 day episodes of care for BPCI hospitals, pooling hospital and physician group participants for episode identification. Because medical patients experience frequent hospital admissions, programmes contain rules for episode precedence, which determine how overlapping episodes are assigned to hospitals.<sup>30–31</sup> To avoid bias from precedence, we constructed naturally occurring episodes by assigning overlapping episodes to the earlier hospital admission. Methods 1 in the appendix provides details of episode construction. We used propensity scores to match BPCI hospitals to non-BPCI hospitals with replacement by condition using baseline hospital and market characteristics from 2011.<sup>15–16</sup> We conducted separate matches at the hospital level for each medical condition, allowing hospitals to be matched with up to three non-BPCI hospitals, within a caliper of 0.2 of the standard deviation of the log odds propensity score.<sup>32</sup>

### Outcomes

Our two prespecified primary outcomes were total spending on episodes and 90 day mortality. We chose 90 day mortality as a measure of quality because mortality is unequivocally important, regardless of differences in spending. Further, mortality is not uncommon in patients admitted to hospital for the four medical conditions evaluated (eg, 12.5% in the study baseline period).<sup>33</sup>

Our secondary spending outcomes included spending by type of care after discharge: all institutional post-acute care (defined as skilled nursing facility, institutional rehabilitation facility, or long term acute care with associated professional fees), skilled nursing facility only, outpatient professional fees, durable medical equipment (eg, cane, walker), home health services, index hospital admission (including facility and physician professional fees), readmissions (including facility and physician professional fees), and outpatient hospital facility care. Secondary outcomes of use included 90 day readmissions, 90 day

emergency department visits, any institutional post-acute use (skilled nursing, institutional rehabilitation, or long term acute care), index discharge (ie, discharge from the initial hospital admission triggering the episode) to a skilled nursing facility, index discharge with home health services, and the total number of skilled nursing days. All outcomes were prespecified except discharge to skilled nursing, discharge to home health services, and total number of days receiving skilled nursing care, which were added retrospectively to help explain the observed spending shifts. All payments were standardized and adjusted to 2016 dollars.<sup>34–35</sup>

### Covariates

Covariates were chosen based on previous studies.<sup>11–13, 16, 36–38</sup> Patient level covariates included age, sex, race, disability status, Medicare and Medicaid dual eligibility status, Elixhauser comorbidities, and admission diagnosis related group.<sup>39</sup> Time varying market covariates included penetration of Accountable Care Organizations, penetration of Medicare Advantage, and number of Medicare beneficiaries.

### Statistical analysis

We used standardized differences of means and proportions to compare hospital characteristics between the propensity matched BPCI and non-BPCI hospitals.<sup>40</sup> Patient and market characteristics and their trends between the baseline and treatment periods, by BPCI and non-BPCI hospitals, were compared using  $\chi^2$  tests and *t* tests for categorical and continuous outcomes, respectively.

In adjusted analysis, we used difference-in-differences models for each outcome using a time varying indicator of hospital participation in BPCI as the treatment. In this design, the treatment indicator reflected whether an episode occurred at a BPCI hospital after contract initiation for a given medical condition, thereby categorizing episodes occurring at BPCI hospitals before participation as episodes in the non-BPCI group. This more closely reflected the time varying nature of BPCI participation compared with traditional difference-in-difference models in which the baseline and treatment periods are fixed regardless of timing of actual contract initiation.<sup>41</sup> We used an intention-to-treat approach, thus once hospitals began participation in BPCI for a given condition, all subsequent episodes were in the BPCI hospital group regardless of later dropout from the programme. We tested the parallel trends assumption in the baseline period using the same outcomes and predictors previously mentioned. All models included hospital and time (ie, calendar quarter) fixed effects.

We used generalized linear models for all outcomes. For mortality and use we used a normal distribution and an identity function (ordinary least squares), except for the total skilled nursing days use outcome for which we used generalized linear models with log-link function and negative binomial distribution. All spending outcomes used generalized linear models

with a log-link function and gamma distribution. Standard errors were clustered at the hospital level. We used an adjusted  $\alpha$  of 0.025 according to the Bonferroni method for the two primary outcomes and 0.05 for secondary outcomes.<sup>42</sup> We used the Holm-Bonferroni method to adjust for multiple comparisons within individual conditions.<sup>43</sup> After applying inclusion and exclusion criteria, complete case analysis was performed. We also examined for any residual correlation between errors across groups after clustering standard errors, possibly because of the 20% sample for non-BPCI hospitals. All analyses were conducted using SAS (version 9.4, SAS Institute).

### Sensitivity analyses

We conducted six sensitivity analyses. First, because of high attrition in participation over time,<sup>16</sup> we examined whether results differed for hospitals that remained in the programme. We examined the subgroup of hospitals in each condition that had remained in the programme as of the end of 2016, regardless of date of entry. Second, we repeated the analysis removing the year 2013, to remove anticipatory effects as hospitals were preparing for the risk bearing phase of the programme. Third, we repeated analysis for the primary outcomes using an alternate matched comparison group that included non-BPCI hospitals from BPCI markets. Fourth, we repeated our primary analysis for total spending with the effect separated by years since the start of participation, allowing us to determine how spending might have changed over time. Finally, we compared our results to those obtained with an alternate episode construction methodology; one that preferentially assigned overlapping episodes in the treatment period to BPCI hospitals (BPCI precedence).<sup>7 10</sup>

### Patient and public involvement

No patients were involved in the design or execution of this study. Since this study used deidentified data, no direct dissemination to research participants is planned.

### Results

Ninety six hospitals participated in BPCI for acute myocardial infarction, 136 for COPD, 185 for congestive heart failure, and 144 for pneumonia (table 1). In total, 238 hospitals participated in BPCI and 1415 hospitals were in non-BPCI markets. Overall, 226 BPCI hospitals were matched to 700 non-BPCI hospitals. Mean follow-up after participation in BPCI began was 23 months across all conditions (23 months for congestive heart failure or COPD, 22 months for acute myocardial infarction or pneumonia). Overall, BPCI hospitals experienced 354 725 episodes (261 163 in the baseline period, 93 562 in the treatment period) and non-BPCI hospitals experienced 289 851 episodes (211 208 in the baseline period, 78 643 in the treatment period). The smaller number of non-BPCI episodes was related to the 20% sample for the non-BPCI group.

### Hospital characteristics

BPCI hospitals differed from non-BPCI hospitals in hospital, market, and episode characteristics in 2011 at the beginning of the baseline period (table 1). BPCI hospitals were larger and more likely to be non-profit, teaching, and urban. In 2011, compared with non-BPCI markets, BPCI markets were larger in total population, had a lower percentage of individuals on a low income, higher Medicare Advantage penetration, and more skilled nursing facility beds.

These differences were reduced after propensity score matching. All standardized differences were equal to or below the ideal threshold of 0.2 for variables included as controls in regression models except for geographic distribution (which was not included in the propensity matching model and had a post-match standardized difference of 0.6).<sup>44</sup> Characteristics of patients for all 48 BPCI conditions at BPCI and non-BPCI hospitals were similar (appendix table 1). Analysis of baseline trends did not exhibit divergence for any outcome (appendix table 2, and graphically shown for primary outcomes in appendix figures 1 and 2).

### Patient and market characteristics

Mean patient age for BPCI hospitals increased from 77.2 years in the baseline period to 77.4 years in the treatment period and from 75.7 years to 76.2 years for non-BPCI hospitals (differential change -0.3 years) (table 2 and appendix table 3). At baseline, 44.4% (n=115 853) of patients at BPCI hospitals and 44.5% (n=94 069) at non-BPCI hospitals were men, with a differential change of -0.01 percentage points. Patient Elixhauser comorbidity index scores for BPCI hospitals increased from 19.4 in the baseline period to 19.5 in the treatment period, whereas it increased from 19.4 to 19.6 for non-BPCI hospitals (differential change -0.09 percentage points). Market level Accountable Care Organization penetration increased differentially by 3 percentage points for BPCI markets compared with non-BPCI markets from the same baseline penetration of 7%. The Medicare Advantage population differentially increased by 0.5 percentage points in BPCI markets. Mean beneficiary market population decreased differentially by -4735 beneficiaries in BPCI markets.

### Primary outcomes

In unadjusted analysis, total episode spending decreased for BPCI hospitals from \$18 995 in the baseline period to \$18 771 in the treatment period, whereas it decreased from \$18 849 to \$18 797 for non-BPCI hospitals (differential change -\$172; table 3 and appendix table 4). In adjusted difference-in-differences analysis (fig 1 and appendix figure 3), BPCI participation was associated with a differential decrease in total episode spending of 1.2% (95% confidence interval -2.3% to -0.2%).

Unadjusted 90 day mortality decreased from 12.5% in the BPCI group in the baseline period to 12.0% in the treatment period, whereas mortality in the non-BPCI group increased from 11.2% to 11.4% (differential

**Table 1 | Characteristics of US hospitals participating versus not participating in bundled payments for care improvement for four medical conditions before and after propensity score matching, 2011. Values are numbers (percentages) unless stated otherwise**

|                                                                      | Before matching       |                       |                         | After matching        |                       |                         |
|----------------------------------------------------------------------|-----------------------|-----------------------|-------------------------|-----------------------|-----------------------|-------------------------|
| Characteristics                                                      | BPCI hospitals        | Non-BPCI hospitals    | Standardized difference | BPCI hospitals        | Non-BPCI hospitals    | Standardized difference |
| No of hospitals                                                      | 238                   | 1415                  | NA                      | 226                   | 700                   | NA                      |
| No of beneficiaries                                                  | 329 221               | 587 295               | NA                      | 301 413               | 244 165               | NA                      |
| Total No of episodes                                                 | 387 963               | 695 249               | NA                      | 354 725               | 289 851               | NA                      |
| No of hospitals:                                                     |                       |                       |                         |                       |                       |                         |
| Acute myocardial infarction                                          | 96 (40)               | 1148 (81)             | NA                      | 89 (39)               | 257 (37)              | NA                      |
| COPD                                                                 | 136 (57)              | 1398 (99)             | NA                      | 119 (53)              | 342 (49)              | NA                      |
| Congestive heart failure                                             | 185 (78)              | 1386 (98)             | NA                      | 171 (76)              | 479 (68)              | NA                      |
| Pneumonia                                                            | 144 (61)              | 1411 (100)            | NA                      | 135 (60)              | 382 (55)              | NA                      |
| No of beneficiaries with condition:                                  |                       |                       |                         |                       |                       |                         |
| Acute myocardial infarction                                          | 28 228 (9)            | 65 815 (11)           | NA                      | 26 260 (9)            | 21 305 (9)            | NA                      |
| COPD                                                                 | 75 313 (23)           | 157 825 (27)          | NA                      | 64 580 (21)           | 57 743 (24)           | NA                      |
| Congestive heart failure                                             | 119 041 (36)          | 160 600 (27)          | NA                      | 110 261 (37)          | 85 347 (35)           | NA                      |
| Pneumonia                                                            | 106 639 (32)          | 203 055 (35)          | NA                      | 100 312 (33)          | 79 770 (33)           | NA                      |
| No of episodes:                                                      |                       |                       |                         |                       |                       |                         |
| Acute myocardial infarction                                          | 29 477 (8)            | 68 736 (10)           | NA                      | 27 448 (8)            | 22 307 (8)            | NA                      |
| COPD                                                                 | 96 602 (25)           | 204 267 (29)          | NA                      | 82 660 (23)           | 74 479 (26)           | NA                      |
| Congestive heart failure                                             | 144 262 (37)          | 195 253 (28)          | NA                      | 133 935 (38)          | 104 469 (36)          | NA                      |
| Pneumonia                                                            | 117 622 (30)          | 226 993 (33)          | NA                      | 110 682 (31)          | 88 596 (31)           | NA                      |
| Hospital characteristics                                             |                       |                       |                         |                       |                       |                         |
| Ownership:                                                           |                       |                       |                         |                       |                       |                         |
| For profit                                                           | 119 (21.3)            | 1072 (20.3)           | 0.5                     | 94 (18.3)             | 256 (17.5)            | <0.001                  |
| Government                                                           | 15 (2.7)              | 927 (17.5)            |                         | 15 (2.9)              | 50 (3.4)              |                         |
| Not-for-profit                                                       | 424 (76.0)            | 3295 (62.2)           |                         | 405 (78.8)            | 1154 (79.0)           |                         |
| Urban status                                                         | 520 (93.2)            | 3433 (64.9)           | 0.7                     | 476 (92.6)            | 1335 (91.4)           | 0.04                    |
| Geographic distribution*:                                            |                       |                       |                         |                       |                       |                         |
| Midwest                                                              | 41 (17.2)             | 402 (28.4)            | 0.7                     | 41 (18.1)             | 222 (31.7)            | 0.6                     |
| North east                                                           | 77 (32.3)             | 103 (7.3)             |                         | 74 (32.7)             | 69 (9.9)              |                         |
| South                                                                | 79 (33.2)             | 618 (43.7)            |                         | 74 (32.7)             | 230 (32.9)            |                         |
| West                                                                 | 41 (17.2)             | 292 (20.6)            |                         | 37 (16.4)             | 179 (25.9)            |                         |
| Teaching hospital status†:                                           |                       |                       |                         |                       |                       |                         |
| Major teaching                                                       | 59 (10.6)             | 354 (6.7)             | 0.3                     | 58 (11.3)             | 155 (10.6)            | 0.02                    |
| Minor teaching                                                       | 225 (40.3)            | 1541 (29.1)           |                         | 199 (38.7)            | 559 (38.3)            |                         |
| Non-teaching                                                         | 274 (49.1)            | 3399 (64.2)           |                         | 257 (50.0)            | 746 (51.1)            |                         |
| Mean (SD) ratio of medical and dental residents to total No of beds‡ | 8.5 (19.7)            | 4.9 (16.1)            | 0.2                     | 8.7 (19.8)            | 8.2 (20.9)            | 0.03                    |
| Mean (SD) disproportionate share (\$)§                               | 5 436 588 (6 898 391) | 3 087 284 (4 802 784) | 0.4                     | 5 303 249 (6 836 720) | 4 642 904 (6 359 659) | 0.1                     |
| Mean (SD) Medicare days as % of total patient days                   | 27.6 (7.4)            | 24.8 (9.0)            | 0.3                     | 27.4 (7.5)            | 27.0 (7.9)            | 0.06                    |
| Mean (SD) total No of hospital beds                                  | 321 (242)             | 205 (179)             | 0.5                     | 316 (238)             | 292 (220)             | 0.1                     |
| Hospital market characteristics¶                                     |                       |                       |                         |                       |                       |                         |
| Mean (SD) population                                                 | 2 211 711 (1925970)   | 1 145 138 (900 691)   | 0.7                     | 1840333 (1 386 232)   | 1 627 170 (1 078 766) | 0.2                     |
| Mean (SD) low income status (%)                                      | 18.6 (13.7)           | 27.0 (19.0)           | -0.5                    | 18.9 (14.1)           | 19.6 (15.6)           | -0.04                   |
| Mean (SD) Medicare Advantage penetration (%)                         | 25.7 (11.7)           | 24.1 (14.2)           | 0.1                     | 25.6 (11.8)           | 25.5 (14.5)           | 0.01                    |
| Mean (SD) SNF beds per 10 000 patients                               | 10 665 (9448)         | 6320 (4777)           | 0.6                     | 9178 (7917)           | 7986 (5297)           | 0.2                     |
| Mean (SD) hospital market share (%)                                  | 40.6 (43.8)           | 36.9 (46.2)           | 0.1                     | 43.1 (44.5)           | 43.7 (47.8)           | -0.01                   |
| Mean (SD) hospital HHI**                                             | 1587.8 (1423.2)       | 2099.9 (1542.4)       | -0.3                    | 1690.3 (1440.5)       | 1773.8 (1601.0)       | -0.05                   |
| Hospital use characteristics                                         |                       |                       |                         |                       |                       |                         |
| Mean (SD) BPCI related hospital discharges (%)††                     | 34.8 (6.2)            | 37.4 (8.2)            | -0.4                    | 34.8 (6.3)            | 35.3 (6.9)            | -0.07                   |
| Mean (SD) proportion of discharges to highest volume SNF (%)         | 27.8 (15.2)           | 38.0 (22.4)           | -0.5                    | 28.0 (14.8)           | 29.4 (17.3)           | -0.1                    |
| Mean (SD) proportion of discharges to highest volume IRF (%)         | 56.9 (43.4)           | 47.1 (47.0)           | 0.2                     | 56.4 (43.6)           | 54.2 (46.2)           | 0.05                    |

\$ 1.00 (£0.74; €0.88).

BPCI=bundled payments for care improvement initiative; COPD=chronic obstructive pulmonary disease; HHI=Herfindahl-Hirschman index; IRF=inpatient rehabilitation facility; SNF=skilled nursing facility; NA=not applicable.

Hospital, beneficiary, and episode counts are shown, along with characteristics used for propensity score matching (except geographic distribution added during review process). Hospital and episode counts are total counts for study period (2011-16). Hospital, hospital market, and hospital use characteristics used 2011 data because 2011, the first year of the study period, was used to match hospitals. BPCI hospitals were matched with non-BPCI hospitals in markets without BPCI hospital participants (non-BPCI markets) for acute myocardial infarction, congestive heart failure, COPD, or pneumonia. BPCI hospitals were propensity score matched by condition with up to three non-BPCI hospitals, using 0.2 of the log odds propensity score. Standardized differences between BPCI hospitals and non-BPCI hospitals were all less than -0.2 except for geographic distribution.

\*Not used in propensity score matching.

†Major teaching hospital is one with Council of Teaching Hospitals designation. Minor teaching refers to non-council hospitals with approved residency training programmes.

‡Measure of size of teaching programme used in Medicare's teaching adjustment. Means are shown because standardized differences were calculated with means.

§Amount paid to hospital under the disproportionate share programme reflecting the indigent population served.

¶By hospital referral region according to the Dartmouth Atlas.<sup>27</sup>

\*\*Measure of hospital market concentration.

††Proportion of annual admissions for 10 highest volume BPCI conditions (by total hospital and physician group episodes).

**Table 2 | Characteristics of patients and markets based on admission to US hospitals participating versus not participating in bundled payments for care improvement for four medical conditions, 2011-16. Values are numbers (percentages) unless stated otherwise**

| Characteristics                   | BPCI hospitals    |                  | Non-BPCI hospitals |                  | Difference-in-differences |          |
|-----------------------------------|-------------------|------------------|--------------------|------------------|---------------------------|----------|
|                                   | Baseline period   | Treatment period | Baseline period    | Treatment period | Estimate*                 | P value† |
| <b>Hospital characteristics</b>   |                   |                  |                    |                  |                           |          |
| No of hospitals                   | 226               | 226              | 700                | 691              | NA                        | NA       |
| No of beneficiaries               | 213 853           | 82 842           | 166 884            | 68 871           | NA                        | NA       |
| Total No of episodes              | 261 163           | 93 562           | 211 208            | 78 643           | NA                        | NA       |
| No of conditions:                 |                   |                  |                    |                  |                           |          |
| Acute myocardial infarction       | 19 720            | 7728             | 15 970             | 6337             | NA                        | NA       |
| COPD                              | 62 071            | 20 589           | 55 413             | 19 066           | NA                        | NA       |
| Congestive heart failure          | 94 470            | 39 465           | 72 442             | 32 027           | NA                        | NA       |
| Pneumonia                         | 84 902            | 25 780           | 67 383             | 21 213           | NA                        | NA       |
| <b>Patient characteristics</b>    |                   |                  |                    |                  |                           |          |
| Mean (SD) age (years)             | 77.2 (12.3)       | 77.4 (12.2)      | 75.7 (12.7)        | 76.2 (12.4)      | -0.3                      | <0.001   |
| Men                               | 115 853 (44.4)    | 41 443 (44.3)    | 94 069 (44.5)      | 34 981 (44.5)    | -0.01                     | 0.98     |
| Race/ethnicity:                   |                   |                  |                    |                  |                           |          |
| White                             | 216 355 (82.8)    | 76 223 (81.5)    | 169 623 (80.3)     | 63 066 (80.2)    | -1.3                      | <0.001   |
| Black                             | 30 979 (11.9)     | 11 976 (12.8)    | 32 475 (15.4)      | 12 115 (15.4)    | 0.9                       | <0.001   |
| Other                             | 13 829 (5.3)      | 5363 (5.7)       | 9110 (4.3)         | 3462 (4.4)       | 0.4                       | 0.005    |
| Mean (SD) Elixhauser index score‡ | 19.4 (13.8)       | 19.5 (13.8)      | 19.4 (13.8)        | 19.6 (13.8)      | -0.1                      | 0.27     |
| Dual eligibility§                 | 69 540 (26.6)     | 23 989 (25.6)    | 55 786 (26.4)      | 19 644 (25.0)    | 0.5                       | 0.07     |
| Disabled                          | 33 122 (12.7)     | 11 431 (12.2)    | 32 850 (15.6)      | 11 183 (14.2)    | 0.9                       | <0.001   |
| <b>Market characteristics¶</b>    |                   |                  |                    |                  |                           |          |
| Mean (SD) ACO penetration (%)*    | 7.4 (11.6)        | 24.5 (13.7)      | 6.8 (11.1)         | 20.9 (13.8)      | 3.1                       | <0.001   |
| Mean (SD) MA penetration (%)**    | 26.3 (11.7)       | 30.8 (11.3)      | 26.1 (13.2)        | 30.0 (13.2)      | 0.5                       | <0.001   |
| Mean (SD) beneficiary population  | 139 519 (106 320) | 136 390 (99 444) | 113 137 (64 271)   | 114 743 (64 118) | -4735                     | <0.001   |

BPCI=bundled payments for care improvement initiative; ACO=Accountable Care Organization; COPD=chronic obstructive pulmonary disease; MA=Medicare Advantage; NA=not applicable. Unadjusted episode level patient and market characteristics used in models are shown along with unadjusted differential changes between BPCI hospital group and non-BPCI hospital group from baseline to treatment periods. The baseline period spanned 1 January 2011 to 30 September 2013. The treatment period varied by hospital condition based on the date of entry so as to maintain consistency with the analytic models. The earliest possible start was 1 October 2013. Differential changes occurred in patient age and race, and market characteristics differed.

\*Difference-in-differences for categorical outcomes are shown as percentage points.

†P values for unadjusted difference-in-differences are based on unadjusted models with a BPCI group indicator, before or after period indicator, and the interaction, without additional covariates.

‡Measure of the likelihood of in-hospital death (range -19 to 89).

§Eligible for both Medicaid and Medicare.

¶Market is defined at the hospital referral region.

\*Determined from a random 20% sample of fee-for-service beneficiaries and the Centers for Medicare and Medicaid Services ACO provider level research identifiable file.

\*\*Determined by Medicare Beneficiary Summary File and computing at a market quarter level the proportion of Medicare beneficiaries in Medicare Advantage any time during that quarter.

change -0.7 percentage points). In adjusted analysis, BPCI participation was not associated with a change in 90 day mortality (-0.1 percentage points, 95% confidence interval -0.5 to 0.2 percentage points; fig 2).

### Secondary outcomes

In adjusted difference-in-differences analyses, spending on skilled nursing decreased (-6.3%, 95% confidence interval -10.0% to -2.5%) in favor of increased spending on home health services (4.4%, 1.4% to 7.5%; fig 1). A decrease also occurred in outpatient professional fees (-10.4%, -19.5% to -0.3%).

Among secondary use outcomes, BPCI participation was associated with a differential decrease in total skilled nursing days during the episode (-6.2%, -9.8% to -2.6%). No changes occurred in 90 day readmissions, emergency department visits, or discharge from initial hospital admission to skilled nursing or discharge with home health services.

### Outcomes by individual medical condition

Hospital participation in BPCI was associated with a differential decrease in total episode spending for pneumonia (-2.0%, 95% confidence interval -3.2% to -0.8%; appendix figure 4) but not for other individual conditions. No changes occurred in 90 day

mortality (appendix figure 5). For secondary outcomes, decreases occurred in outpatient professional fees for acute myocardial infarction, congestive heart failure, and COPD and a decrease in skilled nursing care spending for COPD. Secondary use outcomes did not change.

### Sensitivity analyses

The drop-out rate across all conditions was 31.6%. The decrease in total spending for hospitals that remained in the programme of -1.8% (95% confidence interval -2.6 to -1.0) was similar to that of the primary analysis results (appendix table 5). Removing the year 2013 did not change the direction or statistical significance of any outcomes (appendix table 6). Primary outcome results were also similar after modifying the comparison group to include non-BPCI hospitals from BPCI markets (appendix table 7). Savings in total spending were larger after two and three years of hospital participation compared with the first year (appendix table 8). Boxplots and scatterplots did not suggest any residual correlation between errors across groups after clustering standard errors (appendix figure 6). Constructing episodes with BPCI precedence in the treatment period yielded similar changes in spending on skilled nursing and home health services (appendix figure 7). However, there were also increases in spending on readmissions (3.6%, 0.9% to 6.3%)

**Table 3 | Unadjusted spending, mortality, and use for US hospitals participating versus not participating in bundled payments for care improvement for four medical conditions, 2011-16. Values are numbers (percentages) unless stated otherwise**

| Variables                              | BPCI hospitals  |                  | Non-BPCI hospitals |                  | Difference-in-differences |                |          |
|----------------------------------------|-----------------|------------------|--------------------|------------------|---------------------------|----------------|----------|
|                                        | Baseline period | Treatment period | Baseline period    | Treatment period | Estimate*                 | Percent change | P value† |
| Mean (SD) total episode spending (\$)‡ | 18 995 (16 155) | 18 771 (15 689)  | 18 849 (16 178)    | 18 797 (15 916)  | -172                      | -0.9           | 0.06     |
| 90 day mortality§                      | 32 547 (12.5)   | 11 218 (12.0)    | 23 587 (11.2)      | 8975 (11.4)      | -0.7                      | -5.8           | <0.001   |
| 90 day readmission¶                    | 82 087 (31.4)   | 27 202 (29.1)    | 68 927 (32.6)      | 23 759 (30.2)    | 0.1                       | 0.2            | 0.80     |
| 90 day ED visit**                      | 54 721 (21.0)   | 19 370 (20.7)    | 47 207 (22.4)      | 17 737 (22.6)    | -0.5                      | -2.2           | 0.05     |
| Any institutional PAC use              | 68 054 (26.1)   | 24 524 (26.2)    | 49 457 (23.4)      | 18 357 (23.3)    | 0.2                       | 0.9            | 0.35     |
| Index discharge to SNF                 | 52 582 (20.1)   | 18 910 (20.2)    | 36 760 (17.4)      | 13 773 (17.5)    | -0.03                     | -0.2           | 0.88     |
| Mean (SD) total No of SNF days         | 8.8 (19.7)      | 7.7 (17.3)       | 7.9 (18.9)         | 7.3 (17.4)       | -0.5                      | -5.8           | <0.001   |
| Mean (SD) spending by category (\$)††: |                 |                  |                    |                  |                           |                |          |
| Index hospital admission               | 7611 (2572)     | 7869 (2531)      | 7641 (2626)        | 7930 (2609)      | -31                       | -0.4           | 0.03     |
| Readmissions                           | 4371 (9545)     | 4174 (9380)      | 4632 (9854)        | 4424 (9810)      | 11                        | 0.3            | 0.84     |
| All institutional PAC care‡‡           | 4491 (10 004)   | 4061 (9064)      | 4022 (9530)        | 3814 (9038)      | -222                      | -4.9           | <0.001   |
| Outpatient professional fees§§         | 18 (341)        | 14 (182)         | 20 (245)           | 14 (169)         | 2                         | 13.0           | 0.14     |
| Durable medical equipment              | 190 (1197)      | 222 (1477)       | 217 (1452)         | 235 (1491)       | 13                        | 7              | 0.08     |
| SNF care                               | 4277 (9827)     | 3837 (8870)      | 3803 (9352)        | 3600 (8858)      | -237                      | -5.5           | <0.001   |
| Home health services                   | 1263 (1954)     | 1287 (2030)      | 1153 (1931)        | 1140 (1992)      | 37                        | 2.9            | 0.001    |
| Outpatient hospital care               | 1051 (3720)     | 1143 (4000)      | 1164 (3937)        | 1240 (4179)      | 17                        | 1.6            | 0.45     |

\$ 1.00 (£0.74; €0.88).

BPCI=bundled payments for care improvement initiative; ED=emergency department; PAC=post-acute institutional care; SNF=skilled nursing facility.

\*Difference-in-differences for categorical outcomes are shown as percentage points.

†P values for unadjusted difference-in-difference are based on unadjusted models with a BPCI group indicator, before or after period indicator, and the interaction, without additional covariates.

‡After removing the 5% of episodes with the highest total spending, the mean unadjusted spending for BPCI is \$16 378 versus \$16 359 for non-BPCI in the baseline period and \$16 255 versus \$16 284 in the treatment period.

§Percentage who died within 90 days after discharge alive from index hospital admission.

¶At least one acute care readmission during the 90 day episode.

\*\*At least one visit to an emergency department without readmission to hospital during the 90 day episode.

††Spending for index hospital admission, readmissions, and all institutional PAC care included professional fees for services provided during the stay as well as non-professional spending.

Although some spending outcomes have a right tail skew, means and standard deviations are shown because the models estimated means (with adjustment for skew). Spending by category does not sum to total episode spending because both SNF and all institutional PAC care include SNF facility spending.

‡‡Included skilled nursing facility, long term acute care hospital admission, and institutional rehabilitation as well as professional fees.

§§Professional fees for outpatient hospital care not associated with acute inpatient care or institutional PAC care. Primary outcomes included total episode spending and 90 day post-discharge mortality. Secondary outcomes included spending by category and use. The baseline period spanned 1 January 2011 to 30 September 2013. The treatment period varied by hospital condition based on date of entry so as to maintain consistency with the analytic models. The earliest possible start was 1 October 2013. There were unadjusted differential increases from baseline period to treatment period in the BPCI hospital group relative to non-BPCI group in home health, and differential decreases in 90 day mortality rate, total SNF days, spending on all institutional PAC care, and spending on SNF care.

and the 90 day readmission rate (1.4 percentage points, 95% confidence interval 0.9 to 2.0 percentage points; appendix figure 8). As a result, in contrast to primary analysis findings, total episode spending did not change.

## Discussion

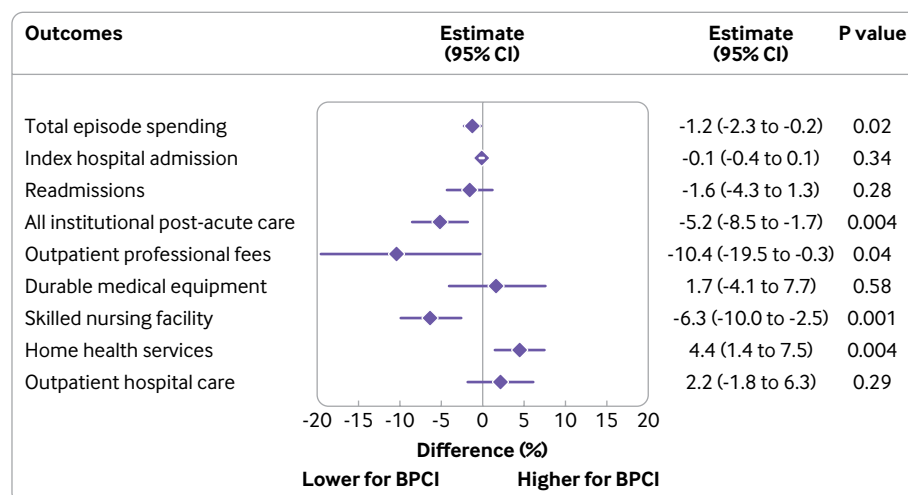
This independent, peer reviewed evaluation of long term changes in outcomes for medical condition episodes in a national bundled payment programme in the US found that hospital participation in these bundles was associated with a small decrease in total episode spending with no change in mortality. We observed a shift in spending for institutional post-acute care generally, with shifts away from skilled nursing care and toward home health services. No discernible changes in quality of care were observed.

## Policy implications

This study has five main implications. First, the association between medical bundles and long term savings is reassuring to policy makers in both the US and internationally who are pursuing bundles as a cost containment strategy.<sup>10</sup> In particular, decision makers should be encouraged by the 1-2% savings for episodes achieved under medical bundles, which complement the 2-4% savings for episodes observed under surgical bundles.<sup>8 11 13 15 17-20 36 37</sup>

Second, the fact that savings grow after one year of participation suggests that medical bundles require time to produce benefits. Our findings contrast with those from an earlier study that did not find practice changes or episode savings after an average of seven months of participation.<sup>16</sup> The differing results are likely related to time: we examined long term outcomes over an average follow-up period of nearly two years, whereas previous work evaluated outcomes over an average of seven months. Our analysis also found that the magnitude of savings was larger during the second and third years after the start of participation compared with the first year. Taken together, this evidence suggests that the type of practice redesign needed to succeed under medical bundles requires learning and implementation over time. Clinical intuition supports this possibility, given the greater complexity of patients cared for under medical bundles compared with surgical bundles.

Third, our analysis highlights that reducing the duration of skilled nursing care after its initiation is a mechanism for achieving savings in medical bundles. This suggests that practice changes in response to medical bundles might use a complementary mechanism to those under surgical bundles, in which the key driver of savings was reducing discharge to skilled nursing facilities (ie, decreasing the number of patients discharged to skilled nursing facilities in



**Fig 1 | Risk adjusted changes in spending associated with hospital participation in bundled payments for care improvement for four medical conditions, 2011-16.** The associations between participation in the bundled payments for care improvement initiative (BPCI) and changes in spending were estimated with separate difference-in-differences models, using a hospital condition specific indicator of entry to the BPCI programme, patient and time varying market characteristics, and quarterly time and hospital fixed effects. All models were generalized linear models with a log-link function and gamma distribution. Low volume hospitals with fewer than 10 episodes for each condition were treated as a single hospital to enable clustered models to converge for the skilled nursing facility care and outpatient professional fees outcomes. BPCI hospital participation was associated with a decrease in all institutional post-acute care spending and skilled nursing facility care spending specifically, as well as outpatient professional fees, combined with increases in spending for home health services

favor of discharge home). The differing foci of practice changes between medical and surgical bundles might not be surprising given the differences in the types of post-discharge care needed (for example, post-surgical physical therapy and rehabilitation can be more easily provided at home, whereas condition based intensive nursing care and drug management might require and be optimized through facility care). More broadly, our findings reinforce skilled nursing facility care as a widely acknowledged source of episode savings, and newly identify its role in medical bundles.

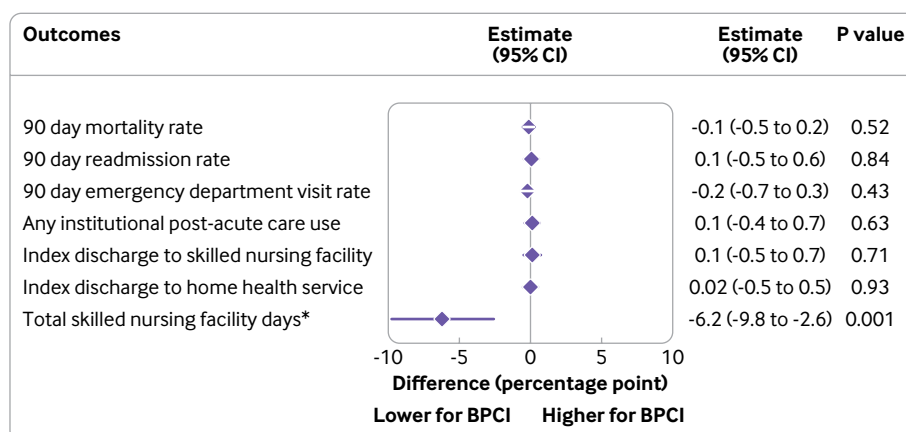
Fourth, our results reinforce existing evidence<sup>15 17-20 36</sup> that in achieving cost savings, participation in bundled payments does not appear to worsen quality. Although these findings are reassuring, more work is needed to fully understand the impact of practice changes in post-acute care use, including when unintended consequences such as increased readmissions might occur.

Finally, our analysis underscores the potential impact that policy design can have on measured performance. In the case of hospital admission triggered episodes in the US bundled payment programmes, medical episode patients might be more likely than surgical episode patients to experience multiple hospital admissions within a short period, creating the potential for overlapping episodes and the need to distinguish between initial hospital admissions and readmissions. The US BPCI and BPCI Advanced programmes apply a series of precedence rules to reconcile overlapping episodes and assign them to participants who should be held accountable for episode quality and costs.

Policy must be designed to deal with episode overlap and assignment to ensure that bundled payment programmes achieve their intended purposes. The use of precedence rules in our analysis would have created bias against BPCI hospitals by preferentially assigning episodes occurring during a series of hospital admissions within short intervals to them. This would have artificially inflated readmission rates in hospitals participating in bundled payments (appendix figure 9 and appendix table 9). Such policy design problems are of high relevance given both their potential importance to participant performance and the lack of attention they have received thus far in peer reviewed literature.

### Limitations of this study

Our study has limitations. First, its quasi-experimental design does not eliminate the possibility of residual confounding. However, our methodological approach, including use of time and hospital fixed effects as well as a time varying treatment variable, helped mitigate this risk. Second, although our results reflect an average of nearly two years of participation in the programme, it is possible that performance continued to change over longer periods. Third, our analysis did not capture all aspects of quality, including patient centered measures of symptom management and functional status. Fourth, the generalizability of our results is limited by our focus on one model in a US bundled payment programme, although BPCI model 2 has served as the basis for all subsequent US programmes. Fifth, our hospital results might not extend to physician group participation. Sixth, we used robust standard errors clustered at the hospital level following previous



**Fig 2 | Risk adjusted changes in mortality and use associated with hospital participation in bundled payments for care improvement for four medical conditions, 2011-16.** The associations between participation in the bundled payments for care improvement initiative (BPCI) and changes in mortality and use were estimated with separate difference-in-differences models, using a hospital condition specific indicator of entry to the BPCI programme, patient and time varying market characteristics, and quarterly time and hospital fixed effects. All models were ordinary least squares except for total number of days for skilled nursing facility care we used a generalized linear model with log-link function and negative binomial distribution. BPCI hospital participation was not associated with a change in the primary outcome of 90 day mortality. BPCI hospital participation was associated with a differential decrease in total days for skilled nursing facility care. \*Reported estimate is a percent (not percentage point) difference

literature, rather than at the market level, although clustering at the market level might have given more conservative estimates. Finally, we used claims rather than clinical data for risk adjustment. Though claims might not account for certain clinical differences between groups and might be influenced by changes in coding practice, adjustment of claims risk is nonetheless often comparable with adjustment of clinical risk.<sup>45-47</sup>

## Conclusion

In this long term study of hospitals that bundled care for four common medical conditions in a prominent national programme, participation in bundled payments was associated with 1-2% episode savings generated from reductions in the duration of skilled nursing care, without changes in mortality.

## AUTHOR AFFILIATIONS

<sup>1</sup>Corporal Michael J Crescenzo VA Medical Center, Philadelphia, PA, USA

<sup>2</sup>Division of General Internal Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

<sup>3</sup>National Clinician Scholars Program, Philadelphia, PA, USA

<sup>4</sup>University of Washington School of Medicine, Seattle, WA USA

<sup>5</sup>Leonard Davis Institute of Health Economics, Philadelphia, PA, USA

<sup>6</sup>Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, 423 Guardian Drive, Philadelphia, PA 19104, USA

**Contributors:** JAR, JML, EJE, XM, JZ, and ASN developed the plan for the study. XM, EZS, EW, and QH performed the data analysis. JAR wrote the first draft of the paper. ASN, JML, and EJE obtained funding. All authors contributed to the development of the study and interpretation of results. JAR, JML, EJE, XM, JZ, CD, DC, and ASN contributed to the critical revision of the manuscript. JAR and ASN are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. Note: The author order has been changed from that of the original submission to reflect QH's

contributions in revision, including leading data analysis for the updated results.

**Funding:** This work was funded by the Commonwealth Fund (grant No 20170816).

**Competing interests:** All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: support from the Commonwealth Fund for the submitted work; JAR reported consulting fees from Tuple Health. ASN reported receiving grants from Hawaii Medical Service Association, Anthem Public Policy Institute, the Healthcare Research and Education Trust, Cigna, Ochsner Health System, United Healthcare Blue Cross Blue Shield of NC, and Oscar Health; personal fees from Navvis Healthcare, Navigant, National University Health System – Singapore, Singapore Ministry of Health, the Medicare Payment Advisory Commission, and Agathos; personal fees and equity from NavaHealth; equity from Embedded Healthcare; speaking fees from the Cleveland Clinic; serving as a board member of Integrated Services. without compensation, and an honorarium from Elsevier Press, none of which are related to this manuscript. Since 2016, EJE reports speaker's fees from Tanner Healthcare System, Mid-Atlantic Permanente Group, American College of Radiology, Marcus Evans, Loyola University, Oncology Society of New Jersey, Good Shepherd Community Care, Remedy Partners, Medzel, Kaiser Permanente Virtual Medicine, Wallace H Coulter Foundation, Lake Nona Institute, Allocation, Partners Chicago, Pepperdine University, Huron, American Case Management Association, Philadelphia Chamber of Commerce, Blue Cross Blue Shield Minnesota, UnitedHealth Group, Futures Without Violence, CHOP, Washington State Hospital Association, Association of Academic Health Centers, Blue Cross Blue Shield of Massachusetts, American Academy of Ophthalmology, Lumeris, Roivant Sciences, Medical Specialties Distributors, Vizient University Healthcare System, Center for Neuro-Degenerative Research, Colorado State University, Genentech Oncology, Council of Insurance Agents and Brokers, Grifols Foundation, America's Health Insurance Plans, Montefiore Physician Leadership Academy, Greenwall Foundation, Medical Home Network, HFMA, Ecumenical Center – UT Health, American Association of Optometry, Associação Nacional de Hospitais Privados, National Alliance of Healthcare Purchaser Coalitions, Optum Laboratories, Massachusetts Association of Health Plans, District of Columbia Hospital Association, Washington University. Since January 2017, EJE has been a venture partner with Oak HC/FT, a firm that invests in health services but not pharmaceuticals or devices. EJE is on the board of two start-ups: Village MD and Oncology Analytics. EJE has investments in Gilead, Allergan, Amgen, Baxter, and UnitedHealth Group.

**Ethical approval:** This study was approved by the University of Pennsylvania's institutional review board (No 824812), with a waiver of informed consent.

**Data sharing:** The data used in this study are available by application from the Department of Health and Human Services or the American Hospital Association.

The manuscript's guarantors (JAR and ASN) affirm that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

**Dissemination to participants and related patient and public communities:** Because this study used deidentified secondary analysis of administrative claims, results could not be disseminated to research participants.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- White RS, Sastow DL, Gaber-Baylis LK, Tangel V, Fisher AD, Turnbull ZA. Readmission rates and diagnoses following total hip replacement in relation to insurance payer status, race and ethnicity, and income status. *J Racial Ethn Health Disparities* 2018;5:1202-14. doi:10.1007/s40615-018-0467-0
- Ellis RP, Chen T, Luscombe CE. Comparisons of health insurance systems in developed countries. Boston University Department of Economics working paper series. 2014;1.
- Struijs JN, Baan CA. Integrating care through bundled payments-lessons from The Netherlands. *N Engl J Med* 2011;364:990-1. doi:10.1056/NEJMp1011849
- de Bakker DH, Struijs JN, Baan CB, et al. Early results from adoption of bundled payment for diabetes care in the Netherlands show improvement in care coordination. *Health Aff (Millwood)* 2012;31:426-33. doi:10.1377/hlthaff.2011.0912
- Pross C, Geissler A, Busse R. Measuring, reporting, and rewarding quality of care in 5 nations: 5 policy levers to enhance hospital quality accountability. *Milbank Q* 2017;95:136-83. doi:10.1111/1468-0009.12248
- Porter ME, Marks CM, Landman ZC. OrthoChoice: bundled payments in the county of Stockholm (B). Harvard Business School Supplement. 2014;(suppl):714-515.
- Centers for Medicare & Medicaid Services. Bundled payments for care improvement (BPCI) initiative: general information. 2020. <https://innovation.cms.gov/innovation-models/bundled-payments>.
- Centers for Medicare & Medicaid Services. Bundled payments for care improvement (BPCI) initiative: public toolkit. 2019. <https://innovation.cms.gov/Files/x/bpci-toolkit.pdf>.
- Meyer H. Joint replacement bundled payments losing their appeal in BPCI Advanced. Modern Healthcare. <https://www.modernhealthcare.com/finance/joint-replacement-bundled-payments-losing-their-appeal-bpci-advanced> (accessed 24 January 2020).
- Centers for Medicare & Medicaid Services. BPCI-Advanced. 2020. <https://innovation.cms.gov/innovation-models/bpci-advanced>.
- Dummit LA, Kahvecioglu D, Marrufo G, 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;316:1267-78. doi:10.1001/jama.2016.12717
- Navathe AS, Troxel AB, Liao JM, et al. Cost of joint replacement using bundled payment models. *JAMA Intern Med* 2017;177:214-22. doi:10.1001/jamainternmed.2016.8263
- Navathe AS, Liao JM, Dykstra SE, 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;320:901-10. doi:10.1001/jama.2018.12345
- Centers for Medicare & Medicaid Services. BPCI Analytic File 2018\_Q3. 2019. <https://innovation.cms.gov/Files/x/bpcianalyticfile.xlsx>.
- Lewin Group. CMS bundled payments for care improvement (BPCI) initiative models 2-4: year 5 evaluation & monitoring annual report. 2018. <https://downloads.cms.gov/files/cmimi/bpci-models2-4-yr5evalrpt.pdf>.
- Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. Evaluation of Medicare's bundled payments initiative for medical conditions. *N Engl J Med* 2018;379:260-9. doi:10.1056/NEJMs1801569
- Lewin Group. CMS bundled payments for care improvement (BPCI) initiative models 2-4: year 1 evaluation & monitoring annual report. 2015. <https://innovation.cms.gov/Files/reports/BPCI-EvalRpt1.pdf>.
- Lewin Group. CMS bundled payments for care improvement (BPCI) initiative models 2-4: year 2 evaluation & monitoring annual report. 2016. <https://innovation.cms.gov/Files/reports/bpci-models2-4-yr2evalrpt.pdf>.
- Lewin Group. CMS bundled payments for care improvement (BPCI) initiative models 2-4: year 3 evaluation & monitoring annual report. 2018. <https://downloads.cms.gov/files/cmimi/bpci-models2-4yr3evalrpt.pdf>.
- Lewin Group. CMS bundled payments for care improvement (BPCI) initiative models 2-4: year 4 evaluation & monitoring annual report. 2018. <https://innovation.cms.gov/Files/reports/bpci-models2-4-yr4evalrpt.pdf>.
- Navathe AS, Shan E, Liao JM. What have we learned about bundling medical conditions? Health Affairs. 2018. <https://www.healthaffairs.org/doi/10.1377/hblog20180828.844613/full/>.
- Lewis VA, Fisher ES, Colla CH. Explaining sluggish savings under accountable care. *N Engl J Med* 2017;377:1809-11. doi:10.1056/NEJMp1709197
- 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;379:1139-49. doi:10.1056/NEJMs1803388
- Gillam SJ, Siriwardena AN, Steel N. Pay-for-performance in the United Kingdom: impact of the quality and outcomes framework: a systematic review. *Ann Fam Med* 2012;10:461-8. doi:10.1370/afm.1377
- Kahvecioglu DC. Medicare's bundled payments initiative for medical conditions. *N Engl J Med* 2018;379:e33.
- Centers for Medicare & Medicaid Services. Bioanalytical files. 2019. <https://innovation.cms.gov/Files/x/bpcianalyticfile.xlsx>.
- Dartmouth University. Dartmouth Atlas. 2019. <https://www.dartmouthatlas.org/>.
- American Hospital Association. Fast facts on US hospitals. 2019. <https://www.aha.org/statistics/fast-facts-us-hospitals>.
- Centers for Medicare & Medicaid Services. Research statistics data. 2020. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/index.html>.
- Donzé J, Lipsitz S, Bates DW, Schnipper JL. Causes and patterns of readmissions in patients with common comorbidities: retrospective cohort study. *BMJ* 2013;347:f7171. doi:10.1136/bmj.f7171
- Chamberlain AM, Dunlay SM, Gerber Y, et al. Burden and timing of hospitalizations in heart failure: a community study. *Mayo Clin Proc* 2017;92:184-92.
- Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharm Stat* 2011;10:150-61. doi:10.1002/pst.433
- Jha AK. To fix the hospital readmissions program, prioritize what matters. *JAMA* 2018;319:431-3. doi:10.1001/jama.2017.21623
- Navathe AS, Liao JM, Polsky D, et al. Comparison of hospitals participating in Medicare's voluntary and mandatory orthopedic bundle programs. *Health Aff (Millwood)* 2018;37:854-63. doi:10.1377/hlthaff.2017.1358
- Liao JM, Emanuel EJ, Polsky DE, et al. National representativeness of hospitals and markets in Medicare's mandatory bundled payment program. *Health Aff (Millwood)* 2019;38:44-53. doi:10.1377/hlthaff.2018.05177
- Navathe AS, Emanuel EJ, Venkataramani AS, et al. Spending and quality after three years of Medicare's voluntary bundled payment for joint replacement surgery: the spending and quality effects of Medicare's bundled payments for care improvement initiative among patients undergoing lower extremity joint-replacement. *Health Aff* 2020;39:58-66. doi:10.1377/hlthaff.2019.00466
- Liao JM, Emanuel EJ, Venkataramani AS, et al. Association of bundled payments for joint replacement surgery and patient outcomes with simultaneous hospital participation in accountable care organizations. *JAMA Netw Open* 2019;2:e1912270.
- Navathe AS, Liao JM, Dykstra SE, 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;320:901-10. doi:10.1001/jama.2018.12345
- 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;47:626-33. doi:10.1097/MLR.0b013e31819432e5
- Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS. SAS Global Forum, 2012;335:1-6.
- The National Bureau of Economic Research. Difference-in-differences with variation in treatment timing. 2018. <https://www.nber.org/papers/w25018>.
- Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995;310:170. doi:10.1136/bmj.310.6973.170
- Abdi H. Holm's sequential Bonferroni procedure. *Encyclopedia Of Research Design*. 2010;1:1-8.

- 44 Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci* 2010;25:1-21. doi:10.1214/09-STS313
- 45 Krumholz HM, Lin Z, Drye EE, et al. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circ Cardiovasc Qual Outcomes* 2011;4:243-52. doi:10.1161/CIRCOUTCOMES.110.957498
- 46 Bratzler DW, Normand SL, Wang Y, et al. An administrative claims model for profiling hospital 30-day mortality rates for pneumonia patients. *PLoS One* 2011;6:e17401. doi:10.1371/journal.pone.0017401
- 47 Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation* 2006;113:1683-92. doi:10.1161/CIRCULATIONAHA.105.611186

#### Web appendix: Supplementary appendix