Supplementary Online Content

Finkelstein A, Ji Y, Mahoney N, Skinner J. Implementation of a Mandatory Medicare Bundled Payment Program for Lower Extremity Joint Replacement Episodes and Discharge to Institutional Postacute Care in an Interim Analysis of the First Performance Year of a 5-Year Randomized Trial. *JAMA.* doi:10.1001/jama.2018.12346

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Supplemental Methods and Results

Table of Contents

1. CJR Eligibility Criteria – MSAs, Hospitals, Patients, Expenditures

MSA eligibility

Original Exclusion Criteria: In July 2015, CMS publicly announced the exclusion criteria, along with the resulting 196 (out of about 388) MSAs deemed eligible for randomization. The exclusion criteria were designed to exclude MSAs with low volume of LEJR, and to exclude MSAs where there was high take-up of BPCI, the voluntary bundled payment initiative that was launched in 2013 and applies to a broad set of 48 inpatient episodes[1]. BPCI was taken up by only 385 acute care hospitals (out of more than 5,500 total) and thus did not have a major effect on eligibility.[2]

More specifically, as detailed in the Federal Register[1], 196 MSAs were deemed eligible for randomization after applying the following exclusion criteria:

1. Excluded if MSA had fewer than 400 LEJR episodes between July 1, 2013 and June 30, 2014.
2. Excluded if MSA had fewer than 400 non-BPCI LEJR episodes between July 1, 2013 and June 30, 2014. The number of BPCI LEJR episodes is the number of potentially eligible LEJR episodes at hospitals participating in BPCI Model 1, and Phase 2 of BPCI Models 2 or 4 as of July 1, 2015.
3. Excluded if MSA had more than 50% of potentially CJR-eligible LEJR episodes in relevant BPCI models between July 1, 2013 and June 30, 2014. This exclusion criterion was applied in two steps. First, MSAs were excluded if more than 50% of potentially CJR-eligible LEJR episodes were in hospitals participating in Phase 2 of BPCI Models 2 or 4 as of July 1, 2015. Second, MSAs were excluded if more than 50% of LEJR referrals to SNF or HHA services were made up by SNF and HHA participating in BPCI Model 3 as of July 1, 2015.
4. Excluded if MSA had more than 50% otherwise CJR-eligible LEJR episodes in Maryland hospitals between July 1, 2013 and June 30, 2014.

Using the 2014 Medicare claims data, we estimate that there were about 486,000 patients with Medicare who underwent LEJR in 2014, of which 430,000 were in an MSA, and 345,000 were in an eligible MSA. About 140,000 were in (original) treatment MSAs and 200,000 in control MSAs. Thus, the LEJR patients in the MSAs included in the experiment represent about 70% of all fee-for-service Medicare LEJR patients, and about 80% of all fee-for-service Medicare-covered LEJR patients in an MSA.

At the same time that CMS published the exclusion criteria in July 2015, it also published the randomization procedure (including strata and treatment probabilities within strata), and the resulting 75 treatment MSAs.[1] Specifically, the 196 eligible MSAs were divided into eight strata, based on the full interaction of (1) average wage-adjusted historical LEJR episode payment, grouped into quartiles, and (2) MSA population size, grouped into above and below median. Treatment probabilities varied within the payment quartiles, and were set so that treatment probabilities were higher in MSAs with higher historical LEJR spending. Specifically, they were 30% in the first quartile of spending, 35% in the second, 40% in the third, and 45% in the fourth. Supplement Table 1 provides more detail.

SAS Enterprise Guide 7.1 was used to select MSAs from each stratum. The statement "PROC SURVEYSELECT" was used with "METHOD=SRS". The seeds for the strata were 907, 414, 525, 621, 1223, 827, 428, and 524[3].

Supplement Figure 1 shows the set of selected, not selected, and ineligible MSAs for the mainland U.S.

Simulating CMS' randomization protocol

We verified via simulation that we could replicate the randomization procedure described by CMS to within statistical sampling error. Specifically, we performed 1,000 simulations of the stratified randomization process. In each simulation, we drew 75 treatment MSAs from the 196 eligible MSAs based on the treatment probabilities in each stratum as shown in Supplement Table 1. We then compared the distribution of characteristics of the simulated treatment MSAs with those of the actual treatment MSAs. If CMS successfully implemented the randomization procedure, we would expect that the characteristics of the actual treatment MSAs to fall within the distribution of the simulated treatment MSAs.

The results of the simulation are shown in Supplement Table 2. For each variable, column (1) reports the mean among actual treatment MSAs, columns (2) and (3) report the mean and standard deviation of the means among simulated treatment MSAs, across 1,000 simulations, and column (4) reports the difference between the mean of actual treatment MSAs and the mean across the 1,000 means of simulated treatment MSAs, divided by the standard deviation of the means of simulated treatment MSAs. The means of the actual treatment MSAs fell within two standard deviations of the simulated means for all but two measures, suggesting that the randomization process was successfully implemented. This approach to verifying the randomization procedure that a public entity has announced it conducted is similar to the approach used in academic work to verify the randomization approach used by the state of Oregon in its Medicaid lottery.[18,19]

Revised Exclusion Criteria

After announcing in July 2015 the initial 75 treatment MSAs in a proposed rule, in November 2015 CMS modified its MSA exclusion criteria based on comments it received to the proposed rule. Specifically, CMS received comments that their original algorithm did not take into account providers who entered into Phase 2 BPCI by October 1, 2015. (October 1, 2015 was the final quarter a Phase 1 BPCI could transition into Phase 2.) CMS thus revised the second and third exclusion criteria. The revised exclusion criteria (compared to the original ones listed above) are as follows[3]:

1. Not changed.
2. The number of BPCI LEJR episodes is now computed as the number of potentially eligible LEJR episodes associated with hospitals participating in BPCI as of October 1, 2015, instead

of as July 1, 2015, and also episodes associated with a physician who was in a Physician Group Practice (PGP) in Phase 2 of BPCI Model 2 as of October 1, 2015.
3. As in 2, the number of BPCI LEJR episodes is now computed based on the list of BPCI-participating hospitals, SNF, and HHA as of October 1, 2015.
4. Not changed.

With the revised eligibility criteria, eight of the initially selected 75 treatment MSAs were no longer eligible for treatment. This results in the final list of 67 treatment MSAs, which was publicly listed.[3] Two MSAs that were originally not eligible for selection became eligible based on the revised eligibility criteria. However, their actual treatment probability remained zero as CMS did not consider them for treatment.

Within-MSA eligibility

Hospital Eligibility: All hospitals paid under Medicare's inpatient prospective payment system that are in the treatment MSAs are required to participate in CJR unless they participate in Model 1 of BPCI or in Phase 2 of Models 2 or 4 of BPCI, in which case the episode is excluded from CJR.[3] The hospital is required to participate the date these conditions cease to hold. We estimate that in 2016, among eligible MSAs, approximately 14% of all Medicare LEJR episodes are excluded due to the hospital eligibility criteria.

Patient Eligibility: CJR covers LEJR episodes for eligible hospitals in treatment MSAs for patients who meet the following criteria for the entire episode: They must have both Medicare Parts A and B coverage, they must not be in a Medicare Advantage Plan, they must not be in a United Mine Workers of America plan, and Medicare must be the primary payer. In addition, the episode is cancelled if the patient is readmitted during the episode to an acute care hospital for LEJR (in this case, the original episode is cancelled and the readmission triggers a new episode), starts an LERJ episode under any of the BPCI models during the episode, or dies during the episode.[3] Among LEJR episodes that pass the hospital eligibility criteria, we estimate that in 2016, among eligible MSAs, about 13% of patients are excluded due to the patient eligibility criteria.

Thus, combining the hospital and patient exclusions, in 2016, about 75% of Medicare fee-for-service patients admitted for one of the two covered DRGs would be eligible for CJR.

Covered by the Bundle: Acute care hospitals in the treatment MSAs are held financially responsible for Medicare FFS spending and quality of care during an episode. An episode begins with a hospital stay that results in a discharge in one of the two included DRGs and ends 90 days after discharge from the hospital.

The bundled payment covers the vast majority of Medicare spending during the episode. There are several categories of spending that are excluded from the bundle, but most of these excluded categories involve negligible dollar amounts.[3] The only consequential exception is the exclusion of spending on hospital admissions during the episode for reasons unrelated to LEJR – e.g., spending for hospital admissions for spending on oncology, trauma, or surgery for chronic or acute diseases. Based on the list of excluded services, we found that excluded spending averaged

to about $147 per episode, or about 0.6% relative to the mean episode claims of $24,160 (Table 2)[4].

2. Analysis

Data

Our main data source is Medicare fee-for-service claims data over the episode of care. Recall that an "episode" is defined as an acute care hospital stay that results in a discharge in one of the two included DRGs and ends 90 days after discharge from the acute care hospital. We use data on 100% of enrollees from 2012-2014, and 2016. We choose not to analyze 2015 data since the assignment of MSAs to treatment or control status was made mid-way through 2015, making it possible that there would be anticipatory responses in 2015.[1]

We draw on two additional data sets. First, the targeted quality measure (Composite Quality Score, or CQS) is constructed, in part, using data from the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS). Second, our estimate of *net* Medicare spending (which adjusts *gross* Medicare fee-for-service spending for any reconciliation payments – i.e., shared savings – made to hospitals at the end of the performance year) combines Medicare claims data with publicly available data published by CMS on hospital-specific reconciliation payments in the first performance year.[5]

We analyze the first performance year of CJR, which includes episodes that begin on or after April 1, 2016, and end no later than December 31, 2016. Specifically, our analysis includes episodes starting between April 1 2016 and September 15, 2016. The end point of September 15, 2016 is chosen so that virtually every episode (defined by the start of a hospital admission and ending 90 days post discharge) falls entirely within the performance year, given average length of stay for an LEJR admission of about 3 days[6].

Based on historical data of a 20% sample of Medicare beneficiaries in years 2012-2014, we estimated that our IV analysis of the effect of CJR had power to detect a 2 percentage point (5 percent) reduction in our primary outcome: share discharged to institutional post-acute care (2-sided size=0.05, power = 0.8). [7]

Outcome measures

Health care utilization: Our primary outcome is the share of patients discharged to institutional post-acute care (PAC). This includes discharges to skilled nursing facilities, long-term care facilities, and inpatient rehabilitation facilities, but excludes discharges to home health care (i.e. non-institutional post-acute care). About one-third of LEJR patients are discharged to institutional PAC.

We also analyze the share of patients discharged to each of the other major discharge locations: home health care (about one-third), home without home health care (about one-third), and other (the remaining approximately 2 percent of discharges); "other" includes Medicare-approved

swing beds, inpatient care at another acute care hospital, an intermediate care facility, and other less common destinations such as psychiatric hospitals, hospices, and federal hospitals.

Finally, we analyze the number of days in institutional PAC. This is defined as the sum of length of stay in SNF, LTCH and IRF during the episode. The number of days is measured across all episodes, not conditional on institutional PAC use.

Medicare spending: We analyze Medicare fee-for-service spending during the episode. This is defined as the total amount of Medicare Part A and Part B fee-for-service spending during the episode. It does not include any reconciliation payments. We exclude the negligible share of Medicare spending that, as described above, is excluded from the bundle. We also look separately at Medicare spending for institutional PAC during the episode.

Medicare net spending: This is calculated by adding any end-of-year reconciliation payment made to a treatment hospital under CJR to FFS spending.

Targeted quality measure (Composite Quality Score): To be eligible for reconciliation payments, the hospital must achieve an "acceptable" level of quality. This means a score of 5 or greater on the Composite Quality Score (CQS), which is capped at 20 points. Unlike all the other outcomes which we measure over the episode, the CQS measure is "backward looking". We therefore expect little movement, particularly in the first performance year that we analyze here.

In addition, unlike all the other outcomes we analyze, CQS is not measured in the Medicare claims data but through Hospital Compare, which reports it only for treatment hospitals. Since the Hospital Compare data on the Composite Quality Score is only available for treatment hospitals, we constructed our own CQS measure for treatment and control hospitals, which captures the two main components of CQS: the total hip arthroplasty/total knee arthroplasty (THA/TKA) complication measure (up to 10 points) and the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) patient experience measure (up to 8 points). Together these two components can together provide up to 18 points.

The THA/TKA complication measure is a facility-level risk-standardized 90-day complication rate, where a complication is defined as one of eight conditions occurring during or within a given time period of the admission. The eight conditions are heart attack, pneumonia, or sepsis/septicemia/shock within seven days of admission; surgical site bleeding, pulmonary embolism, or death within 30 days of admission; and mechanical complications or periprosthetic joint/wound infection within 90 days of admission. This measure is a three-year rolling average. Between 0 to 10 points are assigned to a hospital based on their quality performance percentile for this measure. For the first performance year, the three-year rolling average covers April 1, 2013 to March 31, 2016, a period prior to the implementation of CJR for the treatment hospitals. We downloaded this measure from Hospital Compare (where it is reported for both treatment and control hospitals).

The HCAHPS is a standardized national patient experience survey. A total of 11 measures, including seven composite measures and five single-item measures, are scored in the range of zero to 100. The HCAHPS Survey Measure is a hospital-level linear mean roll-up (HLMR) score

combining 10 of the 11 scores from the survey. These 10 measures include six composite measures (Communication with Nurses, Communication with Doctors, Responsiveness of Hospital Staff, Communication about Medicines, Discharge Information, and Care Transition) and four single-item measures (Cleanliness, Quietness, Overall Hospital rating, and Recommend the Hospital). The one composite measure excluded from the HLMR score is Pain Management. Between 0 to 8 points are assigned to a hospital based on their quality performance percentile for this measure. Note that the HCAHPS measure is based on a hospital-wide survey of patients, (and not limited to LEJR patients). In addition, for the first performance year, it is measured from July 1, 2015 through June 30, 2016,[3] three-quarters of which is prior to the implementation of CJR for the treatment hospitals. We therefore expected very little movement on this hospital-wide, mostly backward-looking measure[7]. We downloaded this measure from Hospital Compare, where it is reported for both treatment and control hospitals.

Finally, hospitals can receive 2 points from successfully submitting their patient-reported outcomes (PRO) data and the risk variable data; we did not include this in our targeted quality measure as this measure does not apply to the control MSAs. (In addition, hospitals can receive up to 1.8 points if their quality performance for either or both of the above measures improves from the previous performance year by two deciles or more. Needless to say, this is not relevant for the first performance year that we study.)

Non-targeted quality measures: We analyze three non-targeted quality measures. The first is a version of the THA/TKA complication rate, which, as noted above is an input into targeted quality. The targeted quality measure uses a three-year moving average of the facility-level risk-standardized 90-day complication rate for total hip and total knee arthroplasty. We instead measure the complication rate for episodes in the first performance year. Specifically, we code the underlying eight "complications" defined above and analyze the share of discharges with any complication, which we refer to as the "complication rate".

Second, we analyze the share of episodes with an ER visit – in other words, the 90-day ER visit rate. This was a measure of quality of care used in analysis of voluntary bundled payment models for LEJR in the Medicare population through BPCI.[8,9]

Finally, we analyze the 90-day all-cause readmission rate. This is a standard measure of quality of care, and has been used specifically for LEJR in the Medicare population,[10] and to study the effect of voluntary bundled payment models for LEJR in the Medicare population through BPCI.[9] We include under "all cause readmissions" the small number of patients readmitted or CJR, who would otherwise be excluded ex-post from CJR eligibility.

All 90-day measures examine the 90 days post discharge from the initiating acute care hospital.

Number and complexity of covered procedures: By making the hospital the residual claimant on all costs for the episode, CJR may create incentives for hospitals to avoid LEJR patients, or to avoid certain types of LEJR patients. Any such extensive margin effects would be of interest in their own right, as they would suggest a strategic response by providers to CJR which could have direct effects on patient health and healthcare use. In addition, any such strategic responses would need to be accounted

for in interpreting estimates of the effect of CJR on healthcare use and health care quality among LEJR admitted patients.

We therefore analyze the number of LEJR admissions per 1,000 FFS Medicare enrollee and the number of CJR-eligible LEJR admissions per 1,000 FFS Medicare enrollees; the latter is about 70 percent of the former, for reasons described in Section 1 above.

Because the target price on which incentives are based varies within hospital based on DRG and whether the patient has a hip fracture, we analyze the number of LEJR patients and the number of CJR-eligible patients admitted in each of the four different target price categories. Specifically, the target price on which CJR incentives are based is a function of historical episode spending calculated separately by DRG (470 vs 469) and within DRG by whether there was a hip fracture.

In addition, we analyze whether CJR affects the type of patients treated. Our main measure is the Elixhauser Comorbidity Measure, which has been used to analyze the effect of bundled payment for LEJR Medicare patients on patient composition in prior work.[11] The measure is defined as the sum of 31 different comorbidity indicators[12,13], and thus indicates the number of different comorbidities a patient has.

We also perform additional analyses on other measures of patient composition: average age, share age 80 and above, share white, share female, share disabled, share also on Medicaid. We also analyze the number of Charlson comorbidities. Charlson comorbidities are a set are of 17 comorbidities based on ICD-9 CM or ICD-10 diagnosis.[13,14] Our measure is the number of Charlson Comorbidities present in the episode.

Finally, we analyze two measures of predicted post-discharge Medicare spending during the episode. To do this, we use data from 2014 to generate predicted episode Medicare spending (exclusive of the index admission where spending is primarily a function of DRG) on CJR-eligible LEJR patients as a function of demographic conditions. Specifically, we regress Medicare spending for the episode (exclusive of the index admission) on a full set of interactions for age (in five-year age bins), race (white or not) and gender, as well as an indicator variable for Medicaid coverage, and an indicator variable for original Medicare eligibility due to disability; the regression also includes MSA fixed effects. A regression with MSA fixed effects alone has an R^2 of 0.04. Adding the age-by-race-by-gender interactions brings the R^2 up to 0.14 and adding Medicaid and disability status raises it further to 0.16. Adding indicators of Charlson Comorbidities raises the R^2 to 0.21, but we were concerned about potential endogeneity of diagnosis coding to incentives. We therefore report two different predicted post-discharge spending measures – one that includes the comborbidities in the prediction algorithm and one that does not.

Our prediction sets the MSA to be MSA 14460 (Boston-Cambridge-Newton, MA-NH) and uses the full set of age-by-race-by sex interactions as well as indicators for Medicaid coverage and original Medicare eligibility due to disability. With these demographics alone, we are able to generate a standard deviation of predicted Medicare spending (exclusive of the index admission) of $4,232, or about one-third of the standard deviation of actual post-discharge Medicare

spending in that MSA. Our prediction therefore seems to capture meaningful variation in likely patient spending. When we include the comorbidities and the DRG indicator the standard deviation of predicted Medicare spending (exclusive of the index admission) rises to $5,458.

Estimating equations

We take advantage of the random assignment of CJR to some MSAs and not others to examine the effect of CJR. Our analysis is at the MSA level. Our outcomes are measured in the first performance year for patients who would be CJR-eligible if they were in a treatment MSA.

Let j denote MSA and let t denote calendar year, where *t=1* is the first performance year (i.e. part of 2016), t=0 is 2014, and t=-1 is 2013. Let y_{jt} denote an average outcome in MSA j in year t.

We denote by z_j a binary variable which takes the value of 1 for the 75 MSAs that were assigned to the treatment group under the *initial* eligibility criteria. As mentioned above, after CMS published the list of 75 MSAs initially assigned to treatment (with the remaining 121 assigned to control) in July 2015,[1] CMS updated the exclusion criteria in November 2015 in response to comments on the proposed rule, dropping 8 (publicly listed) MSAs from the treatment group[3]. We were unable to determine exactly which control group MSAs would have been excluded under the updated eligibility criteria. We therefore follow an instrumental variable[15] strategy for analyzing the effect of CJR, in which we use initial random assignment to treatment (z_j) as an instrument for whether the MSA was actually treated (d_j). Specifically, d_j is a binary variable that takes on a value of 1 for the 67 MSAs that were actually treated. This approach thus adjusts for noncompliance using a contamination-adjusted ITT[16], also known as a Wald estimator[17]. More concretely, the IV estimates are simply the ITT estimates divided by the compliance rate (67/75).

Using initial random assignment to the initial treatment group as an instrument for actual treatment meets the criteria for a valid instrumental variable. First, the instrument, initial treatment group assignment, is correlated with the endogenous variable, actual treatment. Indeed, because nearly all of the MSAs that were initially assigned to treatment ended up being treated and none of the MSAs that were initially assigned to control ended up being treated, the correlation is close to 1. Second, the instrument is independent of unmeasured factors that may influence the outcome. This is trivially satisfied in this context because, conditional on strata fixed effects, initial assignment to treatment was determined at random. Third, the instrumental only effects outcomes through the treatment. This is satisfied because the only effect of treatment was being paid via the bundled payments model.

For the IV analysis, our first stage equation is given by

$$d_j = \alpha_0 + \alpha_1 z_j + \alpha_2\, y_{j0} + \alpha_3 y_{j,-1} + \alpha_{4,s(j)} + \epsilon_j \quad (1)$$

and the second stage by

$$y_{j1} = \beta_0 + \beta_1 d_j + \beta_2\, y_{j0} + \beta_3 y_{j,-1} + \beta_{4,s(j)} + \epsilon_{jt} \quad (2)$$

Here, $\alpha_{4,s(j)}$ and $\beta_{4,s(j)}$ are fixed effects for strata $s(j)$. The coefficient of interest is β_1 and captures the average effect of CJR on outcome y. We control for strata fixed effects because treatment probabilities varied across strata; since treatment is correlated with strata – and outcomes may well vary be strata – controlling for them is necessary for consistent estimation of β_1. Controlling for two years of lags of the dependent variable (y_{j0}, and $y_{j,-1}$) is not needed for consistent estimation of $\beta_1$but, we determined, substantially improves our power. [7]

Of course it is possible that the 8 originally selected MSAs were affected by the treatment; we therefore also estimate intent to treat (ITT) models, in which we compare outcomes for the original treatment MSAs to the control MSAs:

$$y_{j1} = \gamma_0 + \gamma_1 z_j + \gamma_2 y_{j0} + \gamma_3 y_{j,-1} + \gamma_{4,s(j)} + \epsilon_j \quad (3)$$

where the coefficient of interest is now γ_1- which captures the average difference in outcomes between originally treated MSAs and control MSAs - and $\gamma_{4,s(j)}$ are fixed effects for strata $s(j)$ and all other variables are as defined previously.

Additional results

Balance on MSA characteristics

In the main text we examined balance of many of our outcome variables across control and treatment MSAs. In Supplement Table 3, we additionally examine balance across a range of other MSA characteristics. There are no statistically significant differences for any of the characteristics we examine. An F-test of the joint hypothesis that all of the differences are zero yields a p-value of 0.43.

Means by strata

Randomization occurred within 8 strata (based on historical payments and population size of the MSAs). Within each strata the probability an MSA was treated (i.e. selected for CJR) varied from 30 percent to 45 percent, with the treatment probability higher for strata with higher historical payments. Thus, treatment probability is only random conditional on strata. Because the probability of treatment is correlated with strata and strata (by design) differ in their outcomes, both intent-to-treat and instrumental variable analysis must control for strata indicators. In addition, as we pre-specified, we control for lags of the dependent variable to improve statistical precision. The main analysis in the text therefore reports the control means and the intent-to-treat estimates based on equation (3) which include controls for strata and for lags of the dependent variable.

For completeness, we also report the raw control and treatment means separately for each stratum. These results are presented in Supplement Table 4 (analog of balance Table 2, by strata),

Supplement Table 5 (analog of balance Supplement Table 3, by strata) Supplement Table 11 (analog of healthcare use and spending Supplement Table 9, by strata), and Supplement Table 12 (analogue of quality and composition Supplement Table 10, by strata).

Incremental effect of controlling for lags of dependent variable

To examine the incremental effects of controlling for lags of the dependent variable, Supplement Table 8 shows instrumental variable (IV) estimates from equations (1) and (2) without and with these controls for all of the outcomes considered in Tables 3 and 4 of the main text. Column (1) reports the mean across the 121 control MSAs. Column (2) reports the coefficient from an two stage least squares IV regression of the row variable on a dummy for treatment, instrumented by a dummy for original treatment assignment, controlling for strata fixed effect. Column (3) reports the heteroskedasticity robust p-value from the IV regression in column (2). Column (4) reports the coefficient from our baseline analysis (equations 1 and 2): an IV regression of the row variable on a dummy for treatment, instrumented by a dummy for original treatment assignment, controlling for strata fixed effect and lagged outcomes from two prior years. Column (5) reports the heteroskedasticity robust p-value from the IV regression in column (4). As expected[7], controlling for lags of the dependent variable increases precision but does not generally affect the magnitude of the estimates.

Intent-to-treat analysis

Supplement Tables 9 and 10 show the intent-to-treat effects of CJR based on equation (3). They are laid out in parallel fashion to the instrumental variable estimates of the same outcomes based on equations (1) and (2) shown in Tables 3 and 4 in the main text.

Additional eligibility analysis

In the main text, in Table 4, we analyzed the effect on LEJR admissions and CJR-eligible LEJR admissions. Supplement Table 6 examines this in more detail. Hospitals and patients can be exempted from CJR based on certain exclusion criteria. Some of these exclusion criteria are potentially endogenous (such as being readmitted for LEJR during the episode or dying during the episode). Panel A shows no effect on any of the exclusion criteria. Panel B shows no effect on the number of LEJR patients in any of the four LEJR categories with different target prices and Panel C shows no effect on the number of CJR-eligible LEJR patients in any of the four categories.

Additional patient composition analysis

In the main text, in Table 4, we analyzed the effect on patient severity as measured by the Elixhauser Comorbidity Measure. Supplement Table 7 performs additional analyses on other measures of patient composition: average age, share age 80 and above, share white, share female, share disabled, share also on Medicaid, number of Charlson Comorbidities, and our two measures of predicted post-discharge Medicare spending during the episode. There is no evidence of an effect on any of these alternative measures.

References cited

1. 80 Federal Register 134. 2015.

2. CMS. Bundled Payments for Care Improvement Initiative (BPCI) [Internet]. 2016 Apr. Available from: https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2016-Fact-sheets-items/2016-04-18.html

3. 80 Federal Register 226. 2015.

4. Centers for Medicare and Medicaid Services. Episode exclusions. Comprehensive Care for Joint Replacement Model. [Internet]. [cited 2017 Oct 4]. Available from: https://innovation.cms.gov/Files/worksheets/ccjr-exclusions.xlsx.

5. CMS. Comprehensive Care for Joint Replacement Model [Internet]. 2018. Available from: https://innovation.cms.gov/initiatives/cjr

6. 81 Federal Register 162. 2016.

7. Finkelstein A, Ji Y, Mahoney N, Skinner J. The Impact of Medicare Bundled Payments: Pre-Analysis Plan [Internet]. [cited 2018 Mar 29]. Available from: https://clinicaltrials.gov/ct2/show/NCT03407885

8. Dummit LA, Kahvecioglu D, Marrufo G, Rajkumar R, Marshall J, Tan E, et al. Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. JAMA. 2016 Sep 27;316(12):1267–78.

9. Navathe AS, Troxel AB, Liao JM, Nan N, Zhu J, Zhong W, et al. Cost of Joint Replacement Using Bundled Payment Models. JAMA Intern Med. 2017 Feb 1;177(2):214–22.

10. Colla C, Bynum J, Austin A, Skinner J. Hospital Competition, Quality, and Expenditures in the U.S. Medicare Population. NBER Working Paper #22826. 2016 Nov;

11. Navathe AS, Troxel AB, Liao JM, Nan N, Zhu J, Zhong W, et al. Cost of Joint Replacement Using Bundled Payment Models. JAMA Intern Med. 2017 Feb 1;177(2):214–22.

12. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998 Jan;36(1):8–27.

13. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005 Nov;43(11):1130–9.

14. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–83.

15. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. Journal of the American statistical Association. 1996;91(434):444–455.

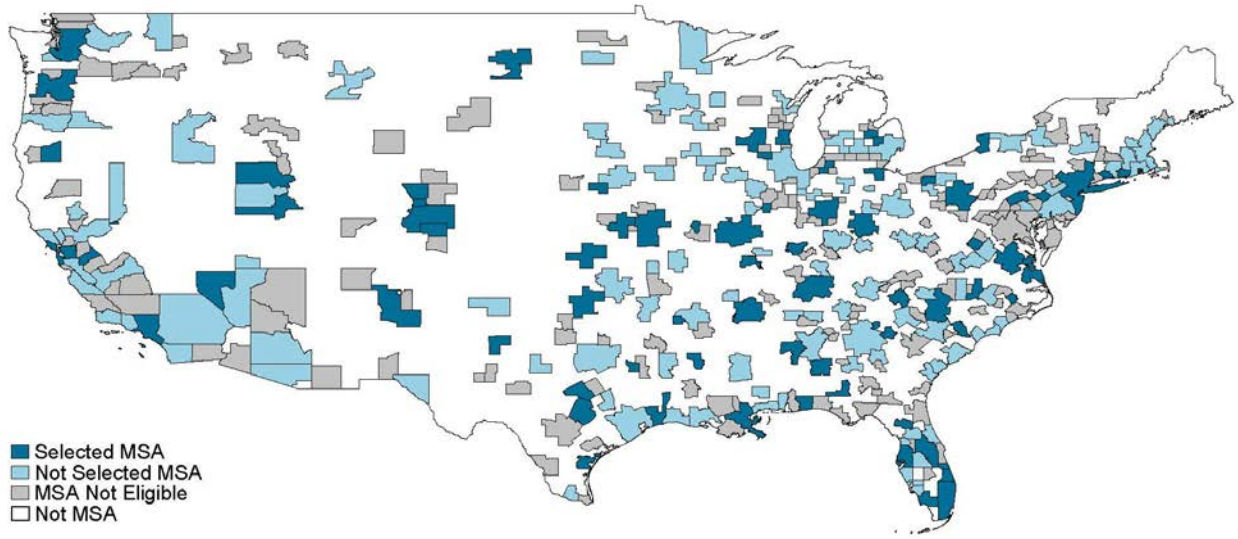
16. Sussman JB, Hayward RA. An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials. BMJ. 2010 May 4;340:c2073.

17. Baiocchi M, Cheng J, Small DS. Tutorial in Biostatistics: Instrumental Variable Methods for Causal Inference*. Stat Med. 2014 Jun 15;33(13):2297–340.

18. Finkelstein A, Taubman S, Wright B, Bernstein M, Gruber J, Newhouse JP, et al. The Oregon Health Insurance Experiment: Evidence from the First Year*. Q J Econ. 2012 Aug 1;127(3):1057–106.

19. Baicker K, Taubman SL, Allen HL, Bernstein M, Gruber JH, Newhouse JP, et al. The Oregon Experiment — Effects of Medicaid on Clinical Outcomes. New England Journal of Medicine. 2013 May 2;368(18):1713–22.

eFigure. Map of MSAs



Note: Figure shows originally selected, not selected, and not eligible MSAs in the mainland US. In addition, there is one not selected and one not eligible MSA in Alaska, one not selected and one not eligible MSA in Hawaii, and one not selected and six not eligible MSAs in Puerto Rico. Of the 75 originally selected MSAs, the following 8 were excluded from the final treatment: Colorado Springs, CO, Evansville, IN-KY, Fort Collins, CO, Las Vegas-Henderson-Paradise, NV, Medford, OR, Richmond, VA, Rockford, IL, and Virginia Beach-Norfolk-Newport News, VA-NC.

eTable 1: Summary of Strata

Stratum	2010 Census Population (1)	Wage-adjusted Episode Payments (2)	Treatment Probability (3)	Number of Eligible MSAs (4)	Number of Original Treatment MSAs (5)	Number of Final Treatment MSAs (6)	Examples of MSAs in Stratum (7)
1	Below Median	1st Quartile	30%	33	10	8	Anchorage AK, Boulder, CO, Flint, MI
2	Below Median	2nd Quartile	35%	19	7	6	Carson City, NV, Columbus, GA-AL, Springfield, IL
3	Below Median	3rd Quartile	40%	22	9	8	Ashville, NC, Lincoln, NE, Huntsville, AL
4	Below Median	4th Quartile	45%	24	11	11	Barnstable Town, MA, Bowling Green, KY, Mobile, AL
5	Above Median	1st Quartile	30%	16	5	5	Minneapolis-St. Paul-Bloomington, MN-WI, Portland-South Portland, ME, San Francisco-Oakland-Hayward, CA
6	Above Median	2nd Quartile	35%	30	11	10	Boston-Cambridge-Newton, MA-NH, Charleston-North Charleston, SC, Columbus, OH
7	Above Median	3rd Quartile	40%	27	11	9	Kansas City, MO-KS, Los Angeles-Long Beach-Anaheim, CA, Philadelphia-Camden-Wilmington, PA-NJ-DE-MD
8	Above Median	4th Quartile	45%	25	11	10	Chicago-Naperville-Elgin, IL-IN-WI, Miami-Fort Lauderdale-West Palm Beach, FL, New York-Newark-Jersey City, NY-NJ-PA
			Total	196	75	67	

Notes: Table reports the probability of treatment, number of MSAs eligible for selection, number of MSAs originally selected for treatment, number of MSAs eventually receiving treatment, and examples of MSAs, for each of the eight strata. Table based on Table 3 of 80 Federal Register 226, 2015 and https://innovation.cms.gov/Files/worksheets/ccjr-populationpayment.xlsx

eTable 2: Simulation

Outcome	Mean Value in Treatment MSAs (1)	Mean of Simulated Treatment Means (2)	SD of Simulated Treatment Means (3)	Number of SD Difference (4)
Population	1,410,979	1,187,231	176,880	1.26
Median Household Income, $	53,689	53,416	851	0.32
High School Graduation Rate, %	88.54	88.19	0.40	0.88
Hospital Beds per 1000 Residents, No	4.3	4.2	0.17	0.83
Share of For-profit Hospital Beds, %	17.60	20.75	2.1	1.54
Number of Medicare Enrollees, No	124,439	108,676	15,328	1.03
Share of Medicare Enrollees also on Medicaid, %	17.53	17.47	0.48	0.12
Share of Black Medicare Enrollees, %	10.71	9.34	0.72	1.91
Average Age of Medicare Enrollees, y	70.1	70.1	0.11	0.01
Share Discharged to Institutional Post Acute Care, %	43.28	42.94	0.98	0.35
Share Discharged to Home Health Agency, %	37.39	34.35	1.43	2.12
Share Discharged Home, %	17.39	20.36	1.33	2.24
Share Discharged to Other, %	1.94	2.35	0.37	1.11
Number of Days in Institutional PAC, d	10.9	10.8	0.22	0.58
Medicare Episode Spending in Institutional PAC, $	4,932	4,888	111	0.40
Total Medicare Episode Spending, $	24,570	24,401	276	0.61
90-day Complication Rate, %	1.45	1.43	0.07	0.42
90-day ER Visit Rate, %	20.92	20.46	0.28	1.66
90-day All Cause Readmission Rate, %	11.23	11.14	0.22	0.44
LEJR Admissions (per 1,000 enrollees), No	8.2	8.1	0.32	0.45
CJR-eligible LEJR Admissions (per 1,000 enrollees), No	6.9	6.8	0.25	0.51
Elixhauser Comorbidity Index, No	2.4	2.4	0.03	1.08

Notes: Table reports MSA-level average characteristics for the actual and simulated treatment MSAs. Column (1) reports the average across the 75 original treatment MSAs selected by CMS. Columns (2) and (3) report the results from our simulated treatment assignment, which mimics the original procedure of randomly selecting 75 treatment MSAs from the 196 eligible MSAs through a stratified randomization process. We performed 1000 simulations of the stratified random assignment. In each simulation, we randomly drew 75 treatment MSAs and recorded the mean of each characteristic across the MSAs selected for treatment. Column (2) reports the mean of these mean values across the 1000 simulations; column (3) reports the standard deviation of these mean values across the 1000 simulations. Column (4) reports the difference between the mean of the actual treatment MSAs in column (1) and the mean of the simulated means in column (2), divided by the standard deviation of the simulated means from column (3). All measures are based on the 2014 Medicare claims data; episodes starting between April 1, 2014 and September 15, 2014 are included. N = 196 MSAs. See Table 2 and Supplement Table 3 for variable definitions.

eTable 3: Additional Balance

Outcome	Mean Value in Control MSAs (1)	Original Treatment-Control Difference (2)	P-value (3)
Population[§]	997,056	357,361	0.26
Median Household Income, $[§§]	53,944	426	0.75
High School Graduation Rate, %[§§]	88.2	0.57	0.35
Hospital Beds per 1,000 Residents, No[§§§]	4.0	0.24	0.37
Percent of Hospital Beds that are For-profit, %[¥]	20.8	-5.0	0.12
Number of Medicare Enrollees, No[¥¥]	93,793	24,900	0.36
Percent of Medicare Enrollees also on Medicaid, %[¥¥]	17.4	0.10	0.91
Percent of Medicare Enrollees who are Black, %[¥¥]	7.8	2.2	0.07
Average Age of Medicare Enrollees, y[¥¥]	70.1	0.001	0.996

Notes: Table reports MSA-level average characteristics. Column (1) reports means for the control MSAs, column (2) reports the estimated difference between the 121 control MSAs and the original 75 treatment MSAs from an ordinary least squares regression of the outcome on an indicator variable for original treatment status, and controls for strata fixed effect, column (3) reports the p-value of the difference, with heteroskedasticity robust standard errors. All outcomes are measured in 2015. An f-test of the joint hypothesis that all of the outcomes are jointly zero yields an f-statistic of 1.12 and a p-value of 0.43.

[§] Population is obtained from the Census Bureau's 2015 population estimates.

[§§] Median household income and high school graduation rate (among individuals aged 25 and above) are obtained from the 2015 American Community Survey (ACS) 1-Year Summary File.

[§§§] Hospital beds per 1,000 residents is the total number of acute care hospital beds in an MSA, as recorded in the 2015 Medicare Provider of Services (POS) file, divided by the MSA population (in thousands).

[¥] Share of for-profit hospital beds is the share of all hospital beds in the MSA that are for-profit, as opposed to non-profit or government-owned. Facility ownership information is obtained from the 2015 Healthcare Cost Report Information System (HCRIS) and the number of beds at each facility is obtained from the 2015 POS file.

[¥¥] These measures are computed from the 2015 Medicare claims data. Number of Medicare Enrollees is the total number of non-ESRD Medicare enrollees in 2015 who were not in an HMO and on Parts A and B for all months that they were eligible.

eTable 4: Balance by Strata

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Original Treatment-Control Difference	P-value
	(1)	(2)	(3)	(4)	(5)
Share Discharged to Institutional Post Acute Care, %	1	29.1	26.4	-2.69	0.38
	2	38.0	37.1	-0.85	0.89
	3	45.0	42.1	-2.83	0.48
	4	48.8	51.3	2.50	0.71
	5	34.2	41.4	7.17	0.17
	6	41.7	45.3	3.55	0.41
	7	49.1	47.2	-1.84	0.66
	8	47.5	50.3	2.84	0.51
Share Discharged to Home Health Agency, %	1	14.4	31.3	16.86	0.06
	2	31.6	44.4	12.76	0.22
	3	36.4	44.6	8.23	0.35
	4	32.9	37.4	4.51	0.50
	5	26.2	30.9	4.79	0.56
	6	40.3	34.6	-5.73	0.21
	7	38.0	37.1	-0.89	0.86
	8	36.5	38.6	2.07	0.65
Share Discharged Home, %	1	53.3	40.1	-13.23	0.13
	2	26.7	16.6	-10.09	0.25
	3	16.3	10.9	-5.40	0.39
	4	12.5	7.7	-4.86	0.25
	5	38.6	26.8	-11.84	0.26
	6	16.1	18.9	2.78	0.50
	7	12.0	14.4	2.38	0.58
	8	14.2	9.5	-4.67	0.19
Share Discharged to Other, %	1	3.1	2.2	-0.94	0.46
	2	3.7	1.9	-1.82	0.15
	3	2.3	2.3	0.00	1.00
	4	5.8	3.7	-2.15	0.52
	5	1.0	0.9	-0.11	0.81
	6	1.8	1.2	-0.59	0.47
	7	1.0	1.3	0.35	0.50
	8	1.7	1.5	-0.24	0.80

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Original Treatment-Control Difference	P-value
	(1)	(2)	(3)	(4)	(5)
Number of Days in Institutional PAC, d	1	7.6	8.3	0.71	0.41
	2	10.9	9.3	-1.56	0.22
	3	11.8	10.7	-1.12	0.27
	4	12.1	12.3	0.22	0.86
	5	8.1	10.6	2.47	0.10
	6	10.5	11.1	0.56	0.40
	7	11.9	12.0	0.05	0.96
	8	11.5	12.1	0.64	0.58
Medicare Spending in Institutional PAC, $	1	3,569	3,304	-265	0.49
	2	4,475	3,859	-616	0.31
	3	4,991	4,815	-176	0.72
	4	5,313	5,672	359	0.52
	5	4,497	5,376	879	0.24
	6	4,689	4,962	274	0.48
	7	5,338	5,290	-47	0.93
	8	5,525	5,864	339	0.56
Total Medicare Episode Spending, $	1	22,567	21,384	-1,183	0.21
	2	22,849	22,857	8	1.00
	3	23,556	23,751	194	0.74
	4	23,880	24,926	1,046	0.35
	5	24,942	28,104	3,163	0.18
	6	25,239	24,182	-1,057	0.36
	7	25,225	25,288	63	0.97
	8	25,429	26,938	1,509	0.22
90-day Complication Rate, %	1	1.1	1.1	0.01	0.98
	2	1.3	1.8	0.47	0.21
	3	1.5	1.4	-0.05	0.89
	4	1.4	1.6	0.26	0.51
	5	1.3	1.2	-0.07	0.81
	6	1.4	1.3	-0.16	0.55
	7	1.6	1.2	-0.32	0.20
	8	1.6	1.9	0.34	0.09

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Original Treatment-Control Difference	P-value
	(1)	(2)	(3)	(4)	(5)
90-day ER Visit Rate, %	1	19.6	20.7	1.10	0.36
	2	22.4	20.0	-2.39	0.10
	3	20.6	20.2	-0.46	0.81
	4	21.5	23.5	1.97	0.12
	5	19.7	21.0	1.33	0.52
	6	19.0	19.6	0.56	0.47
	7	18.6	22.0	3.34	0.01
	8	20.5	20.0	-0.54	0.55
90-day All Cause Readmission Rate, %	1	10.0	10.0	0.00	1.00
	2	10.8	10.3	-0.46	0.65
	3	10.4	10.8	0.37	0.71
	4	11.7	13.9	2.24	0.14
	5	11.2	10.3	-0.88	0.50
	6	11.1	9.7	-1.37	0.11
	7	10.6	11.9	1.24	0.13
	8	12.3	11.9	-0.41	0.66
LEJR Admissions (per 1,000 enrollees), No	1	11.9	10.8	-1.14	0.57
	2	10.5	13.2	2.70	0.40
	3	8.5	8.5	-0.03	0.98
	4	7.2	8.2	1.07	0.46
	5	7.2	6.4	-0.79	0.55
	6	6.9	7.4	0.49	0.43
	7	6.6	6.9	0.26	0.71
	8	5.8	5.6	-0.18	0.72
CJR-eligible LEJR Admissions (per 1,000 enrollees), No	1	9.3	9.4	0.05	0.98
	2	8.8	12.0	3.28	0.25
	3	7.6	6.9	-0.75	0.38
	4	6.4	7.3	0.93	0.46
	5	5.6	5.0	-0.63	0.44
	6	5.7	6.1	0.38	0.51
	7	5.7	5.0	-0.76	0.27
	8	5.0	4.7	-0.34	0.48

Notes: Table replicates Table 2 (Balance) in the paper separately for each strata, without controlling for lags. N = {33, 19, 22, 24, 16, 30, 27, 25} MSAs for stratum 1 through stratum 8, respectively.

eTable 5: Additional Balance by Strata

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Original Treatment-Control Difference	P-Value
	(1)	(2)	(3)	(4)	(5)
Population	1	276,496	235,782	-40,714	0.33
	2	275,576	190,268	-85,308	0.06
	3	254,156	315,770	61,614	0.22
	4	238,699	287,585	48,886	0.31
	5	1,250,330	2,386,026	1,135,696	0.20
	6	1,812,714	1,470,683	-342,030	0.46
	7	1,385,217	2,401,964	1,016,746	0.39
	8	2,443,692	3,781,740	1,338,048	0.49
Median Household Income, $	1	56,830	53,983	-2,847	0.44
	2	49,339	48,437	-903	0.81
	3	49,894	49,513	-381	0.93
	4	46,800	45,391	-1,409	0.59
	5	56,431	68,847	12,417	0.12
	6	60,940	56,784	-4,156	0.26
	7	55,886	57,226	1,339	0.63
	8	49,880	54,956	5,076	0.15
High School Graduation Rate, %	1	89.1	89.8	0.68	0.73
	2	89.6	88.7	-0.90	0.56
	3	89.2	88.4	-0.77	0.54
	4	87.1	85.3	-1.8	0.23
	5	88.0	88.6	0.59	0.88
	6	89.8	90.1	0.37	0.65
	7	87.3	89.1	1.7	0.18
	8	84.5	88.4	4.0	0.09
Hospital Beds per 1000 Residents, No	1	4.8	4.0	-0.77	0.42
	2	4.6	5.4	0.83	0.44
	3	4.5	4.1	-0.34	0.65
	4	4.1	6.2	2.0	0.02
	5	2.9	2.9	0.01	0.97
	6	3.2	3.7	0.53	0.31
	7	3.8	4.1	0.38	0.46
	8	4.3	3.6	-0.64	0.26

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Original Treatment-Control Difference	P-Value
	(1)	(2)	(3)	(4)	(5)
Share of For-profit Hospital Beds, %	1	13.9	2.7	-11.2	0.06
	2	20.1	7.0	-13.1	0.16
	3	23.5	13.2	-10.2	0.40
	4	41.7	25.8	-15.9	0.16
	5	7.0	15.4	8.4	0.57
	6	13.1	19.0	5.9	0.40
	7	22.6	19.8	-2.8	0.71
	8	30.0	30.7	0.66	0.95
Number of Medicare Enrollees, No	1	34,400	29,649	-4,751	0.39
	2	34,592	26,230	-8,362	0.16
	3	38,150	44,308	6,158	0.41
	4	34,510	35,949	1,438	0.79
	5	86,967	177,073	90,106	0.16
	6	161,104	131,162	-29,942	0.48
	7	136,824	183,478	46,654	0.49
	8	213,659	337,472	123,813	0.51
Share of Medicare Enrollees also on Medicaid, %	1	17.1	15.4	-1.7	0.23
	2	16.8	17.7	0.83	0.77
	3	17.6	16.6	-1.0	0.70
	4	17.9	18.8	0.85	0.72
	5	18.4	24.3	5.8	0.29
	6	17.8	17.0	-0.83	0.61
	7	16.3	15.9	-0.49	0.84
	8	17.8	18.1	0.35	0.89
Share of Black Medicare Enrollees, %	1	2.9	5.8	2.9	0.22
	2	6.8	8.3	1.5	0.70
	3	12.1	12.0	-0.02	1.00
	4	8.2	15.8	7.6	0.06
	5	3.8	4.7	0.91	0.57
	6	8.5	10.5	2.0	0.47
	7	9.1	11.8	2.6	0.33
	8	13.3	12.6	-0.72	0.87

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Original Treatment-Control Difference	P-Value
	(1)	(2)	(3)	(4)	(5)
Average Age of Medicare Enrollees, y	1	70.2	70.3	0.15	0.70
	2	69.9	70.1	0.18	0.78
	3	70.1	70.3	0.18	0.84
	4	70.1	69.9	-0.19	0.75
	5	70.3	69.8	-0.50	0.49
	6	69.7	69.5	-0.14	0.77
	7	70.2	70.5	0.26	0.52
	8	70.4	70.3	-0.11	0.82

Notes: Table replicates Supplement Table 3 (Additional Balance Table) separately for each strata, without controlling for lags. N = {33, 19, 22, 24, 16, 30, 27, 25} MSAs for stratum 1 through stratum 8, respectively.

eTable 6: Eligibility Analysis

Outcome	Mean Value in Control MSAs (1)			Change with Bundled Payment (2)						P-value (3)
Panel A: Exclusions based on eligibility criteria:[§]										
Share in non-PPS hospital	0.08	+/-	0.22	-0.009	(	-0.032	to	0.142	)	0.44
Share in BPCI	0.54	+/-	1.04	0.051	(	-0.267	to	0.153	)	0.76
Share not on Part AB for entire episode	0.55	+/-	0.29	-0.013	(	-0.064	to	0.076	)	0.61
Share ESRD	0.004	+/-	0.009	0.000	(	-0.003	to	0.026	)	0.996
Share Medicare not primary payer	0.46	+/-	0.30	0.015	(	-0.031	to	0.022	)	0.52
Share Readmitted for LEJR	0.10	+/-	0.09	-0.014	(	-0.030	to	0.001	)	0.08
Share Died	0.17	+/-	0.10	-0.012	(	-0.037	to	0.013	)	0.36
Panel B: Share with LEJR:[§§]	9.02	+/-	4.76	-0.137	(	-0.416	to	0.142	)	0.34
Share with LEJR (DRG 470 no fracture)	7.78	+/-	4.38	-0.121	(	-0.395	to	0.153	)	0.39
Share with LEJR (DRG 470 fracture)	0.80	+/-	0.35	0.019	(	-0.038	to	0.076	)	0.52
Share with LEJR (DRG 469 no fracture)	0.20	+/-	0.15	-0.007	(	-0.040	to	0.026	)	0.67
Share with LEJR (DRG 469 fracture)	0.24	+/-	0.13	-0.004	(	-0.030	to	0.022	)	0.75
Panel C: Share in CJR:[§§§]	7.20	+/-	3.54	0.049	(	-0.321	to	0.419	)	0.80
Share in CJR (DRG 470 no fracture)	6.28	+/-	3.26	0.018	(	-0.321	to	0.357	)	0.92
Share in CJR (DRG 470 fracture)	0.62	+/-	0.31	0.049	(	-0.011	to	0.109	)	0.12
Share in CJR (DRG 469 no fracture)	0.15	+/-	0.12	-0.006	(	-0.035	to	0.023	)	0.69
Share in CJR (DRG 469 fracture)	0.14	+/-	0.09	0.012	(	-0.011	to	0.035	)	0.31

Notes: Table reports change with bundled payments on eligibility criteria (Panel A) and number of LERJ and CJR admissions (Panels B and C). In column (1), plus-minus values are means +/- SD. Change with bundled payment was estimated with the use of two-stage least-squares instrumental variable regression (See equations 1 and 2). All regressions include indicators for strata, and control for two lags of the dependent variable; p-values and confidence intervals are based on heteroskedasticity-robust standard errors. All outcomes measured during the episode. All measures are based on the 2016 Medicare claims data. N = 196 MSAs.

[§] The number of LEJR admissions per 1,000 fee-for-service Medicare beneficiaries excluded from CJR due to based on each eligibility criterion. The categories are not mutually exclusive and sum up to more than the difference between the share with LEJR and the share in CJR.

[§§] The number of LEJR admissions per 1,000 fee-for-service Medicare beneficiaries.

[§§§] The number of CJR-eligible LEJR admission per 1,000 fee-for-service Medicare beneficiaries.

eTable 7: Additional Patient Composition

Outcome	Mean Value in Control MSAs (1)			Change with Bundled Payment (2)						P-value (3)
Average Age, y	72.5	+/-	0.9	0.03	(	-0.14	to	0.20	)	0.71
Share Age 80 and Above, %	19.9	+/-	3.3	-0.13	(	-0.99	to	0.73	)	0.77
Share White, %	90.6	+/-	7.5	-0.20	(	-0.87	to	0.47	)	0.56
Share Female, %	64.8	+/-	3.4	-0.42	(	-1.21	to	0.38	)	0.31
Share Disabled, %	17.3	+/-	5.8	-0.05	(	-0.80	to	0.71	)	0.91
Share Also on Medicaid, %	11.7	+/-	6.0	-0.06	(	-0.84	to	0.73	)	0.89
Number of Charlson Comorbidities, No[§]	0.75	+/-	0.11	0.01	(	-0.02	to	0.04	)	0.43
Predicted Medicare Spending Post-discharge, $[§§]	14,677	+/-	521	-6.5	(	-106.9	to	93.9	)	0.90
Predicted Medicare Spending Post-discharge (incl. comorbidities), $[§§§]	14,984	+/-	608	31.7	(	-111.1	to	174.4	)	0.67

Notes: Plus-minus values are means +/- SD. Change with bundled payment was estimated with the use of two-stage least-squares instrumental variable regression (See equations 1 and 2). All regressions include indicators for strata, and control for two lags of the dependent variable; p-values and confidence intervals are based on heteroskedasticity-robust standard errors. All outcomes measured during the episode. All measures are based on the 2016 Medicare claims data; episodes starting between April 1, 2016 and September 15, 2016 are included. N = 196 MSAs.

[§] The Charlson Comorbidities are a set of 17 comorbidities based on ICD-9 CM or ICD-10 diagnosis codes. This measure is the number of Charlson Comorbidities present in the episode. 0 in this measure means no comorbidity was present and a higher number means more comorbidities.

[§§] We generate predicted Medicare spending during the episode (excluding the index admission) based on coefficients from a regression using pre intervention 2014 data of episode spending (excluding the index admission) on fully interacted five-year age bin, race, and sex, Medicaid status, and disability status, as well as MSA fixed effects. For our projection we set all MSAs to MSA 14,460 (Boston-Cambridge-Newton, MA-NH). Actual spending for CJR-eligible admissions for this MSA had a standard deviation of $13,021; our predicted spending measure is able to capture about one-third of that variation, with a standard deviation of $4,232.

[§§§] Same as [§§] except that we additionally include indicators for each of the 17 Charlson comorbidities when generating the predicted Medicare episode spending.

eTable 8: Sensitivity

Outcome	Mean Value in Control MSAs (1)	IV (w/o lags) (2)	p-Value (3)	IV (w/ lags) (4)	P-value (5)
Share Discharged to Institutional Post Acute Care, %	33.7	-2.4	0.13	-2.9	0.005
Share Discharged to Home Health Agency, %	32.2	6.6	0.03	1.31	0.40
Share Discharged Home, %	32.2	-3.4	0.27	2.56	0.14
Share Discharged to Other, %	1.9	-0.80	0.04	-0.40	0.08
Number of Days in Institutional PAC, d	8.2	-0.33	0.36	-0.48	0.12
Medicare Spending in Institutional PAC, $	3,871	-252	0.18	-307	0.04
Total Medicare Episode Spending, $	22,872	-96	0.85	-453	0.06
Net Medicare Payment, $	22,872	590	0.24	234	0.31
Targeted Quality Measure: CQS	9.3	0.72	0.1	0.72	0.10
90-day Complication Rate, %	1.2	0.04	0.73	-0.002	0.99
90-day ER Visit Rate, %	20.1	0.58	0.17	0.25	0.48
90-day All Cause Readmission Rate, %	10.1	0.15	0.64	0.01	0.97
LEJR Admissions (per 1,000 enrollees), No	9.0	0.12	0.84	-0.14	0.34
CJR-eligible LEJR Admissions (per 1,000 enrollees), No	7.2	0.27	0.58	0.05	0.80
Elixhauser Comorbidity Index, No	2.3	0.03	0.51	-0.01	0.73

Notes: For each row variable, column (1) reports the mean across the 121 control MSAs. Column (2) reports the coefficient from a two-stage least-squares instrumental variable regression, including indicators for strata but not lags of the outcome variable. Column (3) reports the heteroskedasticity robust p-value from the IV regression in column (3). Column (4) reports the coefficient from our baseline two-stage least-squares instrumental variable regression (see equations 1 and 2); this includes indicators for strata and lagged outcomes from two prior years. Column (5) reports the heteroskedasticity robust p-value from the IV regression in column (4). The sample includes all CJR-eligible LEJR episodes that began between April 1 and September 15, 2016. The lagged outcomes in two prior years come from the corresponding period in 2013 and 2014. N = 196 MSAs. See Table 2 for variable definitions.

eTable 9: Healthcare Use and Spending (ITT)

Outcome	Mean Value in Control MSAs (1)			Intent-to-treat Estimate (2)						P-Value (3)
Discharge Destination, %										
Discharged to Institutional Post Acute Care	33.7	+/-	11.2	-2.6	(	-4.5	to	-0.7	)	0.007
Discharged to Home Health Agency^	32.2	+/-	19.4	1.2	(	-1.6	to	3.9	)	0.42
Discharged Home (w/o Home Health Agency)^	32.2	+/-	23.3	2.3	(	-0.8	to	5.3	)	0.15
Discharged to Other^	1.9	+/-	3.0	-0.4	(	-0.8	to	0.1	)	0.09
Number of Days in Institutional PAC, d	8.2	+/-	2.7	-0.4	(	-1.0	to	0.1	)	0.13
Medicare Episode Spending in Institutional PAC, $	3,871	+/-	1,394	-274	(	-531	to	-17	)	0.04
Total Medicare Episode Spending, $	22,872	+/-	3,619	-402	(	-822	to	18	)	0.07
Medicare Spending Net of Reconciliation Payment, $*	22,872	+/-	3,619	209	(	-205	to	623	)	0.33

Notes: Plus-minus values are means +/- SD. Intent to treat in column 2 was estimated from a regression of the outcome variable on an indicator variable for original treatment status, controls for strata, and two lags of the dependent variable (equation 3); p-values and confidence intervals are based on heteroskedasticity-robust standard errors. All outcomes measured during the episode. All measures are based on the 2016 Medicare claims data unless otherwise noted; episodes starting between April 1, 2016 and September 15, 2016 are included; lags of the dependent variable are based on the corresponding period in 2013 and 2014. N = 196 MSAs. See Table 2 (balance) for variable definitions.

^ In our pre-analysis plan (available at https://clinicaltrials.gov/ct2/show/NCT03407885) we specified that we would combine discharges to Home Health Agency with Institutional PAC rather than analyzing Home Health Agency separately, and we did not specify looking at the remaining categories. The changes are non-substantive and designed for ease of exposition.

* Medicare spending Net of Reconciliation Payment equals Total Medicare Episode Spending (previous row) plus any reconciliation payments made by Medicare. Data on reconciliation payment was downloaded from https://innovation.cms.gov/Files/x/cjr-py1reconpym.xlsx. Since reconciliation payment did not exist prior to 2016, we control for lags of total Medicare Episode Spending instead.

eTable 10: Quality and Patient Composition (ITT)

Outcome	Mean Value in Control MSAs (1)			Intent-to-treat Estimate (2)						P-Value (3)
Panel A: Quality Measures										
Targeted Measure: Composite Quality Score[§]	9.3	+/-	2.8	0.64	(	-0.13	to	1.41	)	0.11
Non-targeted Quality Measures, 90-day %										
Complication Rate	1.2	+/-	0.71	-0.002	(	-0.19	to	0.19	)	0.99
ER Visit Rate	20.1	+/-	2.9	0.22	(	-0.41	to	0.86	)	0.50
All Cause Readmission Rate	10.1	+/-	2.0	0.01	(	-0.51	to	0.54	)	0.97
Panel B: Admissions and Patient Composition										
LEJR Admissions (per 1,000 enrollees), No	9.0	+/-	4.8	-0.12	(	-0.38	to	0.14	)	0.36
CJR-eligible LEJR Admissions (per 1,000 enrollees), No	7.2	+/-	3.5	0.04	(	-0.30	to	0.39	)	0.81
Elixhauser Comorbidity Index, No[^]	2.3	+/-	0.27	-0.01	(	-0.06	to	0.04	)	0.73

Notes: Plus-minus values are means +/- SD. Intent to treat in column 2 was estimated from a regression of the outcome variable on an indicator variable for original treatment status, controls for strata, and two lags of the dependent variable (equation 3); p-values and confidence intervals are based on heteroskedasticity-robust standard errors. All measures are based on the 2016 Medicare claims data unless otherwise noted; episodes starting between April 1, 2016 and September 15, 2016 are included, lags of the dependent variable are based on the corresponding period in 2013 and 2014. N=196 MSAs. See Table 2 (balance) for variable definitions.

[§] This measure ranges from 0 to 18, with higher numbers indicate higher quality. The intent to treat estimate does not include lags of the dependent variable since it involves 3-year moving averages.

[^] Outcome not pre-specified in analysis plan.

eTable 11: Healthcare Use and Spending by Strata (ITT)

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Intent-to-treat Estimate	P-value
	(1)	(2)	(3)	(4)	(5)
Share Discharged to Institutional Post Acute Care, %	1	24.6	23.8	-0.82	0.80
	2	32.2	27.6	-4.6	0.37
	3	39.9	32.1	-7.9	0.08
	4	40.5	38.6	-1.9	0.73
	5	28.3	27.1	-1.2	0.65
	6	31.6	35.2	3.5	0.39
	7	38.8	32.4	-6.4	0.06
	8	38.8	39.2	0.40	0.93
Share Discharged to Home Health Agency, %	1	13.6	28.0	14.4	0.08
	2	31.1	38.2	7.0	0.53
	3	34.9	48.5	13.6	0.13
	4	38.6	42.4	3.8	0.63
	5	24.7	35.0	10.3	0.33
	6	40.7	36.2	-4.5	0.38
	7	40.8	43.6	2.8	0.68
	8	40.1	43.2	3.1	0.60
Share Discharged Home, %	1	59.2	47.3	-11.9	0.20
	2	34.4	32.7	-1.7	0.89
	3	22.6	18.2	-4.4	0.62
	4	17.4	17.1	-0.34	0.97
	5	46.5	37.2	-9.3	0.40
	6	26.4	27.3	0.8	0.88
	7	19.6	23.2	3.6	0.63
	8	19.8	16.6	-3.1	0.60
Share Discharged to Other, %	1	2.6	0.9	-1.7	0.02
	2	2.2	1.5	-0.71	0.52
	3	2.6	1.2	-1.4	0.41
	4	3.5	1.9	-1.6	0.29
	5	0.5	0.7	0.2	0.37
	6	1.2	1.3	0.1	0.76
	7	0.8	0.8	0.02	0.95
	8	1.4	1.0	-0.38	0.68

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Intent-to-treat Estimate	P-value
	(1)	(2)	(3)	(4)	(5)
Number of Days in Institutional PAC, d	1	6.1	6.7	0.5	0.50
	2	8.5	6.8	-1.7	0.19
	3	9.8	8.0	-1.8	0.07
	4	10.3	9.9	-0.4	0.76
	5	6.4	7.1	0.7	0.31
	6	7.9	8.6	0.7	0.34
	7	8.9	7.9	-1.0	0.17
	8	9.0	9.2	0.2	0.76
Medicare Episode Spending in Institutional PAC, %	1	2,975	2,824	-151	0.65
	2	3,901	3,035	-866	0.15
	3	4,257	3,480	-777	0.11
	4	4,573	4,483	-90	0.87
	5	3,704	3,831	127	0.86
	6	3,694	3,863	169	0.67
	7	4,051	3,675	-375	0.40
	8	4,475	4,493	18	0.98
Total Medicare Episode Spending, $	1	21,819	20,405	-1,414	0.15
	2	22,012	22,000	-12	1.00
	3	22,565	22,064	-500	0.60
	4	22,690	22,983	294	0.79
	5	23,758	25,918	2,160	0.38
	6	23,750	22,578	-1,172	0.35
	7	23,372	23,034	-337	0.82
	8	23,334	24,930	1,596	0.26

Notes: Table replicates Supplement Table 9 (Health Care Use and Spending) separately for each strata, without controlling for lags. N = {33, 19, 22, 24, 16, 30, 27, 25} MSAs for stratum 1 through stratum 8, respectively.

eTable 12: Quality and Patient Composition, by Strata (ITT)

Outcome	Strata	Mean Value in Control MSAs	Mean Value in Treatment MSAs	Intent-to-treat Estimate	P-Value
	(1)	(2)	(3)	(4)	(5)
Targeted Quality Measure: CQS	1	8.8	8.8	0.03	0.99
	2	9.8	12.0	2.2	0.05
	3	8.5	9.4	0.90	0.57
	4	10.2	9.4	-0.79	0.60
	5	10.5	9.7	-0.85	0.39
	6	9.8	10.7	0.85	0.22
	7	8.7	10.6	1.8	0.04
	8	8.8	9.5	0.72	0.45
90-day Complication Rate, %	1	1.2	1.5	0.29	0.42
	2	1.3	1.2	-0.05	0.93
	3	1.1	1.1	0.00	1.00
	4	1.4	1.6	0.19	0.61
	5	1.3	1.4	0.17	0.44
	6	1.3	1.0	-0.33	0.09
	7	1.4	1.1	-0.25	0.21
	8	1.1	1.4	0.32	0.04
90-day ER Visit Rate, %	1	19.9	20.2	0.31	0.77
	2	22.6	22.3	-0.33	0.80
	3	20.0	20.1	0.08	0.96
	4	22.7	22.9	0.20	0.89
	5	18.4	19.8	1.5	0.18
	6	19.1	20.3	1.3	0.13
	7	18.8	20.3	1.5	0.15
	8	20.3	19.9	-0.39	0.61
90-day All Cause Readmission Rate, %	1	9.0	9.1	0.13	0.89
	2	10.8	10.7	-0.15	0.91
	3	9.7	9.7	0.04	0.95
	4	10.6	11.6	0.99	0.25
	5	10.5	8.8	-1.7	0.05
	6	10.3	10.4	0.08	0.92
	7	10.3	10.3	0.03	0.97
	8	10.7	11.5	0.79	0.32

Outcome	Strata (1)	Mean Value in Control MSAs (2)	Mean Value in Treatment MSAs (3)	Intent-to-treat Estimate (4)	P-Value (5)
LEJR Admissions (per 1,000 enrollees), No	1	13.0	11.0	-2.0	0.33
	2	11.1	13.4	2.2	0.47
	3	9.4	9.2	-0.23	0.85
	4	7.5	8.8	1.4	0.39
	5	8.0	6.9	-1.1	0.46
	6	7.6	8.1	0.52	0.44
	7	7.2	8.0	0.75	0.34
	8	6.6	6.1	-0.45	0.46
CJR-eligible LEJR Admissions (per 1,000 enrollees), No	1	10.2	9.6	-0.51	0.73
	2	9.1	12.2	3.1	0.25
	3	8.2	7.6	-0.55	0.60
	4	6.5	8.0	1.5	0.33
	5	5.6	5.1	-0.50	0.50
	6	6.0	6.2	0.14	0.85
	7	5.4	5.5	0.08	0.92
	8	5.3	4.5	-0.78	0.28
Elixhauser Comorbidity Index, No	1	2.3	2.3	0.01	0.98
	2	2.5	2.3	-0.20	0.28
	3	2.4	2.5	0.11	0.31
	4	2.3	2.6	0.30	0.01
	5	2.2	2.3	0.08	0.50
	6	2.4	2.5	0.09	0.48
	7	2.4	2.2	-0.15	0.06
	8	2.3	2.3	-0.03	0.74

Notes: Table replicates Supplement Table 10 (Quality and Patient Composition) separately for each strata, without controlling for lags. N = {33, 19, 22, 24, 16, 30, 27, 25} MSAs for stratum 1 through stratum 8, respectively.

eAppendix 2. Pre-specified Analysis Plan

(also available on clinicaltrials.gov: Finkelstein A, Ji Y, Mahoney N, Skinner J. The Impact of Medicare Bundled Payments: Pre-Analysis Plan. https://clinicaltrials.gov/ct2/show/NCT03407885).

The Impact of Medicare Bundled Payments: Evidence from a Nationwide Randomized Evaluation of Lower Extremity Joint Replacement
Date: October 10, 2017

Amy Finkelstein[1], Yunan Ji[2], Neale Mahoney[3], and Jonathan Skinner[4]

Abstract

This analysis plan pre-specifies our planned approach and main analyses of a nationwide randomized-controlled trial (RCT) of bundled payments in Medicare for knee and hip replacements. The RCT was designed by CMS and launched in April 2016. Randomization was conducted at the Metropolitan Statistical Area (MSA) level with 67 MSAs and about 800 hospitals assigned to the treatment group. We propose to study the impact of this intervention in the first two performance years (i.e. from launch through the end of 2017) on health care utilization and quality.

DEPARTMENT(S):

1. Department of Economics, Massachusetts Institute of Technology and Abdul Latif Jameel Poverty Action Lab, Massachusetts Institute of Technology
2. Graduate School of Arts and Sciences, Harvard University
3. Booth School of Business, University of Chicago
4. Department of Economics, Dartmouth College

Introduction

The shift towards alternative payment models (APMs) is one of the most important trends in U.S. health care. In 2015, the Obama Administration announced the goal of shifting 30% of Traditional Medicare reimbursement from fee-for-service (FFS) to APMs by the end of 2016 and 50% by the end of 2018. In March 2016, CMS announced they had reached the 30% goal, almost 9 months ahead of schedule.[1] While the change in Executive Branch control has created some uncertainty, APMs have bipartisan support[2] and industry observers expect the trend to continue.[3]

One of the main APMs is the bundled payment (BP) model. Under bundled payments, Medicare makes a single payment for all services related to a specific treatment (e.g., hip replacement), including the initial acute care hospital stay and post-acute care (PAC) during the recovery period. BP is viewed as a middle ground between FFS, which may result in the over-provision of care because providers are paid for each service, and capitation, which may result in the under-provision of care because providers are paid a fixed amount irrespective of the volume of services. BP may also encourage coordination of care by holding multiple providers jointly accountable for quality and costs.

We propose to study a randomized-controlled trial (RCT) of bundled payments for knee and hip replacement called Comprehensive Care for Joint Replacement (CJR). The RCT was designed by CMS, and launched in April 2016, with the intention of lasting for 5 years. Randomization was conducted at the Metropolitan Statistical Area (MSA) level, with 67 MSAs and about 800 hospitals assigned to the treatment group. Participation was mandatory for all covered hospitals.

The RCT attracted early controversy due to the mandatory participation requirement. Then-House Ways and Means Chairman Tom Price wrote a letter to CMS on September 29, 2016, signed by 179 members of the House objecting to the fact that participation was not voluntary.[4] Subsequently, in August 2017, HHS, led by then-Secretary Price, proposed that participation in the BP model be made voluntary in 33 of the 67 treated MSAs; comments on the proposed rule are due by October 17, 2017.[5] At the time of this writing it is unclear whether and when this rule would go into effect.

---

[1] See https://innovation.cms.gov/Files/x/ffs-apm-goalmemo.pdf

[2] For instance, President Bush's HHS secretary, Michael Leavitt, writes, "APMs are one of few broadly supported ideas for moving healthcare forward." http://thehill.com/blogs/pundits-blog/healthcare/315999-alternative-payment-models-in-healthcare-are-a-must

[3] For instance, observers have noted that the House and Senate versions of the "ACA repeal" bills left in place the delivery reforms enacted under the ACA: http://www.modernhealthcare.com/article/20170707/NEWS/170709971

[4] See http://www.ascrs.org/legislative-and-regulatory/www/article/tom-price-along-other-ways-means-committee-members-spearhead-letter-cms-insisting-cmmi-stop

[5] See https://s3.amazonaws.com/public-inspection.federalregister.gov/2017-17446.pdf and http://www.modernhealthcare.com/article/20170815/NEWS/170819935. For a negative reaction to the proposed

This analysis plan covers the first two performance years of the potentially five-year intervention (from launch in April 2016 through the end of 2017). Depending on what the outcome of the proposed rule is, we might study longer-term effects of a temporary policy, extend the current planned analyses through more performance years, separately study voluntary selection into and out of the bundled payment treatment, or we might otherwise modify or abandon any future analyses.

CMS has contracted with the Lewin Group to be the official evaluator of the CJR program. We have discussed our interest in analyzing the CJR program with the Lewin group and learned, among other things, that any work product done with or by the Lewin Group is owned by CMS, who may choose whether and what to disclose. We therefore felt that there would be value in our independent analysis of the intervention.

The study is of interest for several reasons. First, and most narrowly, it will quantify the impact of bundled payment for knee and hip replacements. Knee and hip replacements are common and expensive medical procedures, accounting for nearly 0.5 million (or 5.1%) of hospital discharges and $6.0 billion (or 4.4%) in FFS inpatient spending in 2014 (authors' calculations based on 2014 Medicare claims data). Second, the findings should also be relevant for other BP models currently being considered by CMS. For instance, in 2013 CMS launched the Bundled Payments for Care Improvement Initiative (BPCI) to test bundles for 48 different types of episodes, including for orthopedic and cardiac care. However, since participation in BPCI was voluntary, separating treatment from selection effects is difficult (Gronniger et al., 2017).

Third, we plan to undertake heterogeneity analysis to shed light on the economic factors that influence the impact of bundled payments. Better understanding of these factors should help with the design of payment models and our understanding of how integrating financial incentives across multiple providers affects the delivery of medical care. We anticipate that these "reduced form" results will serve as inputs into subsequent (beyond the scope of the current analysis plan) structural estimation of a model of the response of providers to the incentives to "internalize the externalities" created by bundled payments.

Finally, this study will be, to our knowledge, the first study of a nationwide payment reform RCT conducted in the US healthcare system. System-wide policies such as payment reform are typically held up as a prime example of the limited utility of RCTs in US healthcare policy: the important questions are system-wide, the argument goes, and system-wide RCTs either cannot or will not be done. Our study and its results will highlight the feasibility of such RCTs.

2. CJR Model and Experimental Design

The Comprehensive Care for Joint Replacement (CJR) model is a Medicare bundled payment model for lower-extremity joint replacement (LEJR) that holds acute care hospitals (ACHs) financially responsible for the spending and quality of the entire episode of care. The CJR model

---

rule change see e.g. https://www.wsj.com/articles/a-health-care-fix-that-works-now-being-rolled-back-1503258369

targets two types of hospital admissions: major joint replacement or reattachment of lower extremity with and without major complications or comorbidities (MS-DRG 469 and 470). An episode begins with an ACH stay that results in a discharge in one of the two included DRGs, and ends 90 days after ACH discharge.

2.1 Incentives

Before each performance year (PY) begins, hospitals receive their target prices from CMS, determined by historical hospital and regional episode expenditures and a 3% discount factor (DF) to reflect Medicare's portion of savings from CJR. Hospitals are eligible for reconciliation payment from CMS if they spend less than the target prices for an episode, provided that they met an "acceptable" quality standard. Conversely, they are responsible for paying the difference if they spend more than the target prices.

The quality standard is based on a composite quality score (CQS); CQS is capped at 20 points and hospitals must earn at least 5 points ("acceptable quality") to be eligible for CJR reconciliation payments.[6] They can earn up to 10 points based on their quality performance percentile on the total hip arthroplasty / total knee arthroplasty (THA/TKA) complication measure; it is a facility-level risk-standardized 90-day complication rate, where a complication is defined as one of eight conditions occurring during or within a given time period of the admission.[7] They can earn up to 8 points based on the Hospital Consumer Assessment for Health Plan Survey (HCAHPS) score; this is a standardized national patient experience survey in which patients self-report their experience on various dimensions (e.g. communication with nurses, and cleanliness); it is based on all patients at the hospital (not just those receiving the covered procedures). They can earn up to 1.8 points for improvement in either of the prior two measures relative to the previous PY, and finally, they can earn up to 2 points for submitting the patient-reported outcomes and risk variable data.[8]

The CJR program began on April 1, 2016, and, as discussed above, was intended to last for five years. The upside and downside risks of payments increase over time as CMS relaxes the stop gain and stop loss limits. Specifically, in PY1 the maximum payment gain (as a percentage of the aggregate target price) is 5%, and there is no downside risk. In PY 2 the maximum gain remains at 5% but there is now also a maximum 5% loss (downside risk). The upside and downside risks

[6] https://innovation.cms.gov/Files/x/cjr-qualsup.pdf

[7] The eight conditions are heart attack, pneumonia, or sepsis/septicemia/shock within seven days of admission, surgical site bleeding, pulmonary embolism, or death within 30 days of admission, and mechanical complications or periprosthetic joint/wound infection within 90 days of admission. See https://www.medicare.gov/hospitalcompare/Data/Surgical-Complications-Hip-Knee.html

[8] Patient-reported outcomes (PRO) are collected from pre- and post-operative outcome surveys of CJR patients. Additional risk variables are collected with the pre-operation survey. Data collection includes questions on physical and mental health, pain, and activities of daily living, among others. Risk variables include pre-operative use of narcotics, BMI, back pain, among others. For more information about the data, see Table 2 of https://innovation.cms.gov/Files/x/cjr-faq.pdf

increase to 10% in PY3, to 20% in PY4, and to 20% in PY5.[9] Assuming a target price of $26,000 (average cost of LEJR episodes) and "acceptable" quality, we calculate that payments at stake rise from +$1,261 (+5% of baseline spending) in year 1 to +/- $5,044 (+/-20% of baseline spending) in years 4 and 5 of the program.

2.2 Randomization.

In July 2015, CMS publicly announced its exclusion criteria for eligible MSAs in its proposed rule, and posted the list of 196 eligible MSAs on the CJR website.[10] CMS restricted the set of eligible MSAs to exclude MSAs with low volume of LEJR, and to exclude MSAs where there was high take-up of BPCI, the voluntary BP initiative that was launched in 2013 and applies to a broad set of 48 inpatient episodes (80 Federal Register 226).[11]

The resulting 196 MSAs (out of a total of about 380 potential MSAs) were then divided into eight strata, based on the full interaction of the following dimensions: (1) Average wage-adjusted historical LEJR episode payment, grouped into quartiles; (2) MSA population size, grouped into above and below median. Treatment probabilities varied within the wage-adjusted episode payment quartiles: they were 30% in the first quartile, 35% in the second, 40% in the third, and 45% in the fourth. (80 Federal Register 226). CMS published the randomization procedure (including strata and treatment probabilities within strata), and the resulting 75 treatment MSAs (80 Federal Register 134, 2015).

Crucially, we verified that we can replicate CMS's randomization procedure. We confirmed via computer simulation that we could replicate their procedure to within sampling error, and that it produces treatment and control MSAs that are balanced on pre-randomization characteristics. This approach is similar to the steps taken to corroborate the fidelity of randomization by a government entity in the Oregon Health Insurance Experiment (Finkelstein et al., 2012).

In response to comments on the proposed rule, CMS updated the exclusion criteria in November 2015, dropping 8 (again publicly listed) MSAs from the treatment group (80 Federal Register 226, 2015). We have been unable to determine precisely which control group MSAs are excluded under the updated eligibility criteria. Therefore, we will analyze the RCT using the original assignment to treatment as an instrumental variable for being covered by BP (i.e. actual treatment).

[9] For more information see 80 Federal Register 226, 2015, and also Table 6 here: https://innovation.cms.gov/Files/x/cjr-qualsup.pdf

[10] https://innovation.cms.gov/Files/worksheets/ccjr-volumeexclusion.xlsx

[11] BPCI was taken up by only 385 acute care hospitals (out of more than 5,500 total) and thus did not have a major impact on eligibility. See https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2016-Fact-sheets-items/2016-04-18.html for information about BPCI.

3. Proposed analyses

3.1 Time frame of analysis

This analysis plan is for the first two performance years (PYs), or through the end of 2017. We will conduct our baseline analysis using Medicare claims data aggregated to the MSA-by-PY level.

Given the fact that data becomes available gradually over time, we plan to first analyze results from PY1 and to limit our analysis to outcomes that occur during the "episode". As more data become available we will analyze outcomes over a longer time horizon than the "episode" for admissions in PY1, as well as adding in episodes through PY2.

PY1 includes episodes started on or after April 1, 2016 and ended by the end of calendar year 2016. Our analysis of PY1 will therefore be based on episodes starting between April 1 2016 and September 15, 2016; the end point is chosen so that every episode (defined by the start of an ACH admission and ending 90 days post discharge from the ACH) likely falls entirely within the PY, given average length of stay for an LEJR admission of about 3 days (authors' calculations based on 2014 data). PY2 includes episodes that end in calendar year 2017.

As noted, the economic incentives scale up over time. This might lead us to expect that the impact of the BP reform would increase over time. However, to the extent that hospitals anticipate the ramp-up in incentives, evidence on "lumpy adjustment" (e.g., Cooper and Haltiwanger, 2006) suggests that hospitals may fully adjust to the new incentives fairly quickly.

3.2 Baseline Specification

Let j denote MSAs and let t denote performance years, which we normalize to take on a value of $t = 1$ in the performance year of program launch. Let y_{jt} denote an average outcome in MSA j in performance year t, z_j be a binary instrumental variable which takes on a value of 1 for MSAs that were assigned to the treatment group under the *initial* eligibility criteria, and let d_j be a binary treatment variable that takes on a value of 1 for MSAs that were actually treated. Our first stage equation is given by

$$d_j = \alpha_0 + \alpha_1 z_j + \alpha_2 y_{j0} + \alpha_3 y_{j,-1} + \alpha_{4,s(j)} + \epsilon_j$$

The second stage specification takes the form

$$y_{jt} = \beta_0 + \beta_1 d_j + \beta_2 y_{j0} + \beta_3 y_{j,-1} + \beta_{4,s(j)} + \epsilon_{jt}$$

Here, y_{j0} is the average outcome in the MSA in the year prior to launch (i.e. April 1, 2015-September 15 2015), $y_{j,-1}$ is the average outcome in the MSA two years prior to launch (i.e. April 1, 2014- September 15 2014), and $\alpha_{4,s(j)}$ and $\beta_{4,s(j)}$ are fixed effects for strata $s(j)$. The coefficient of interest is β_1 and captures the average effect of the treatment on outcome y. We control for strata fixed effects because treatment probabilities varied across strata. Controlling for the lagged dependent variable (we use the prior two years) substantially improves our power.

We are interested not only in average responses but also heterogeneity in responses across hospitals. We expect larger cost-savings for hospitals with higher baseline post-acute care (PAC)

spending. We therefore intend to examine heterogeneity in the estimated IV effects by the hospital's baseline level of rates of discharge to PAC and downstream PAC spending. We also expect savings may be greater among patients within hospitals with a higher-than-average predicted PAC use, and will examine this dimension as well. More broadly, we plan to use machine-learning techniques to explore factors that predict larger and smaller treatment effects.

3.3 Planned outcomes

All outcomes will be measured using Medicare fee-for-service claims data, with the exception of the targeted quality measure (Composite Quality Score), which comes from Hospital Compare.

It is useful to distinguish between four broad domains of outcomes we plan to study:

1. Health care utilization and spending during the episode
2. Health care quality during the episode
3. Number and complexity of covered procedures (i.e. a potential extensive margin response)
4. Health care use and quality beyond the initial episode

1. <u>Health care utilization and spending during the episode.</u>

Recall that an "episode" is defined as an ACH stay that results in a discharge in one of the two included DRGs, and ends 90 days after ACH discharge.

a) Share of LEJR admissions discharged to institutional post-acute care (PAC). This will be our *primary outcome.* It is defined as the share of index admissions discharged to institutional (i.e. non-home health care) post-acute care – this includes discharges to skilled nursing facilities, long-term care facilities, and inpatient rehabilitation facilities. We chose this as our primary outcome for two reasons. First, it is substantively important. Second, intuition, as well as prior work on bundled payments for LEJR (Dummit et al., 2016), suggests that reductions in discharges to institutional PAC are likely to be especially responsive.

b) Share of LEJR admissions discharged to any PAC. This is defined as the share of index admissions discharged to institutional PAC or a home health agency.

c) Number of days in Institutional PAC. This will be the sum of length of stay in SNF, LTCH and IRF during episode.

d) Total covered Medicare payments. Total covered Medicare payments are defined as the total amount of Medicare Part A and Part B FFS payments that are included in the bundle.[12] On average, the index admission is about half of the total covered Medicare

[12] All Medicare Part A and B payments and services that occur in the episode are included, except for the following: hemophilia clotting factors, new technology add-on payments, transitional pass-through payment for medical devices, payments from certain incentive programs, otherwise included payments that exceed two standard deviations of the regional mean, and services unrelated to the index admission as defined by CMS (including certain inpatient hospital stays, Part B services, and per beneficiary per month (PBPM) payments).The list of

payments for the episode. Note that, as defined, total covered Medicare payments are the payments that *would be made in the absence of BP* (i.e. payments that would occur under FFS Medicare). These are counterfactual for the treatment MSAs. If the data become available, we plan to also look at actual payments made during the episode (which would include any reconciliation payments or repayments to or from hospitals in the treatment MSAs). We also plan to look at price-adjusted payments (using the approach of Gottlieb et al. 2010) to purge the measure of regional price variation.

e) **Total covered Medicare payments for institutional PAC**
f) **Total covered Medicare payments for any PAC**
g) **Total beneficiary payments owed out of pocket. (**In practice, these payments may be covered by a Medigap plan or by employer-sponsored retiree health insurance, but this is not observable in the claims data that we plan to analyze.)

2. Health care quality during the episode

We will analyze two types of measures of health care quality during the episode: the one directly targeted by the BP incentive system ("targeted quality measure") and other measures ("non-targeted quality measures").

Targeted quality measure

As described in Section 2.1, to be eligible for reconciliation payments, the hospital must achieve a score of at least 5 on the Composite Quality Score (CQS), which is capped at 20 points. We are able to measure two components of this CQS, which can provide up to 18 of these 20 points: the THA/TKA complication measure and the HCAHPS patient experience measure. We refer to this outcome as "**targeted quality."**

We expect to find little impact on targeted quality, or either of its components. The THA/TKA measure will be a three-year rolling average. For PY 1 it will be based on a three-year average from April 1, 2013 to March 31, 2016, a period that does not overlap with the study period. In later performance years the measure will be based on a period that overlaps with the first three months of a performance year.[13] Given the backward looking nature of the measure, we expect little movement. The HCAHPS measure is based on a general patient survey at the hospital; because it is a hospital-wide measure (and not limited to LEJR patients) we expect little movement. For PY1 it will be measured using data from July 1, 2015 through June 30, 2016.[14]

Unlike all the other outcomes we will analyze, these targeted quality outcomes are measured not in the Medicare claims data but through Hospital Compare.

---

excluded services can be found at https://innovation.cms.gov/Files/worksheets/ccjr-exclusions.xlsx Most of the exclusions account for a trivial share of spending.

[13] The measure period will be the most current available period when CMS computes reconciliation payments.

[14] 80 Federal Register 226

Non-targeted quality measures.

a) **THA/TKA complication measure.** This facility-level risk-standardized 90-day complication rate for total hip and total knee arthroplasty is part of the targeted quality measure. However this component of the targeted quality measure is a three-year moving average. While we cannot necessarily replicate the measure, we will code the underlying the eight "complications", and analyze "any complication" and "number of complications" during our study period.
b) **Share of episodes with an ER visit.** This was a measure of quality of care used in analysis of the BPCI bundled payment demonstration (Dummit et al., 2016).
c) 90-day all-cause readmission rate. Readmissions are a standard measure of quality of care.

3. Number and complexity of covered procedures

We will examine whether the payment reform affects **LEJR admissions per capita** and measures of **LEJR patient mix** (such as projected spending based on patient characteristics). We will look not only at the number of LEJR admissions per capita but also separately for admissions that are excluded from the BP, such as those to non-PPS hospitals or those that are excluded ex-post because the patient died or was re-admitted for one of the two CJR DRGs during the episode; these are rare outcomes (1.3% and 2.1% of LEJR admissions respectively). Any such extensive margin effects are important in their own right; they will also affect interpretation of our estimates of the impact of BP incentives on spending and utilization.

4. Health care use and quality beyond the initial episode.

The bundled payment only applies to covered healthcare services in the initial episode. But the impact of BP may spill over beyond the episode. We will therefore examine:

a) **1 year total covered Medicare payments**
b) **1 year mortality**
c) **1 year all cause readmissions**
d) **1 year outpatient opioid use.** One way to keep payments within the bundle low during the episode could be to substitute pain medication for care in the bundle, such as days in institutional PAC; this could potentially create long-term opioid dependence. This will only be measurable for the sub-sample of LEJR patients who have Part D coverage.

4. Summary Statistics and Power Calculations.

Our summary statistics and power calculations focus on only the first domain of outcomes: health care utilization and spending during the episode. For these power calculations, we focus on the first performance year (PY). Eventually adding in data from PY 2 will triple our sample size of episodes, which will negligibly improve our power to detect mean effects, because the within-MSA cluster sample size is already fairly large relative to the intra-cluster correlation;

however, the additional sample size from the second year is likely useful for heterogeneity analysis on smaller subsamples.

We created summary statistics and conducted the power calculations for the seven outcomes in our first domain ("health care spending and utilization during the episode") using Medicare claims data for episodes starting between April 1, 2014 and September 15, 2014; the end point is chosen so that every episode (defined by the start of an ACH admission and ending 90 days post discharge from the ACH) likely falls entirely within the PY.

Table 1 shows the results. Rows show different outcome variables. The first row shows our primary outcome (share discharged to institutional post-acute care). Column 1 shows the mean outcome in control group MSAs, column 2 shows our estimated effect size using the IV specification described above, and columns 3 shows the MDE (defined as the effect size divided by the control mean).

The power calculations show that for utilization measures, we are able to detect fairly small effects, both from an economic perspective and relative to non-experimental estimates from Dummit et al (2016). Dummit et al (2016) analyzes the effects of BPCI in its first 21 months on cost and utilization for LJER episodes using a matched-control difference-in-differences design. An important caveat is that the BPCI incentives are the same as the CJR incentives in PY 4 and 5, which are about 4 times stronger than in PY 1 and 2 (American Hospital Association 2016). However, to the extent that hospitals anticipate the ramp-up in incentives, evidence on "lumpy adjustment" (e.g., Cooper and Haltiwanger, 2006) suggests that hospitals may fully adjust to the new incentives fairly quickly.

As noted, we suspect use of institutional PAC will be a primary margin of utilization response. Dummit et al. (2016) find a 9.3% decline in the probability of discharge to Institutional PAC conditional on discharge to any PAC. We are able to detect a 5% change in discharges to institutional PAC conditional on discharge to any PAC, which is nearly half the size of the Dummit et al. (2016) estimate.[15] For perspective, Newhouse et al. (2013) find that PAC spending accounts for 73% of the geographic variation in Medicare spending, suggesting that there is substantial scope for CJR to affect PAC utilization.

On total covered Medicare payments for the episode, we are able to detect a 4% change in total payments, which is about the effect size in Dummit et al (2016). A 4% reduction in spending is also small relative to historical 6% per year spending increases[16] and geographic variation in spending of more than two-fold (Skinner, 2011). As another useful benchmark, Gupta (2017)

[15] This outcome is not one we plan to analyze – since it conditions on a potentially endogenous outcome – and therefore it is not reported in Table 1. We report it here to give a sense of power relative to a reported existing estimate.

[16] See "Health expenditures by state of residence: summary tables, 1991-2014", https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsStateHealthAccountsResidence.html

found that fairly modest incentives of the Medicare program to reduce hospital readmissions had meaningful effects on readmissions and mortality, suggesting that even small financial incentives can impact hospital provider behavior. Finally, we are able to detect a 6% decline in covered Medicare payments for Institutional PAC during the episode, which we suspect is slightly smaller than what Dummit et al. (2016) would find.[17]

Table 1: Power Calculations

Outcome	Control Mean (1)	Effect Size (2)	MDE (3)
Share of LEJR Admissions Discharged to Institutional Post-Acute Care	0.41	0.02	0.052
Share of LEJR Admissions Discharged to Any Post-Acute Care	0.72	0.03	0.038
Number of Days in Institutional Post-Acute Care	10.38	0.72	0.070
Total Covered Medicare Payments	23,962	953.02	0.040
Total Covered Medicare Payments for Institutional PAC	4,738	283.23	0.060
Total Covered Medicare Payments for Any PAC	6,551	299.79	0.046
Total Beneficiary Payments Owed Out of Pocket	2,667	128.55	0.048

Notes: Significance level is 5%, power is 80%, total sample size is 196 MSAs, of which 75 are initially assigned to treatment and 67 end up actually treated. Analysis is based on IV estimates in text; first stage R-squared 0.84. Power calculations are based on 2014 "outcome" data meant to mimic the time period of the first performance year (episodes that begin between April 1 and September 15). Column 3 (MDE) reports the effect size (column 2) as a percent of the control mean.

5. Conclusion

This is an analysis plan of the impact of the two years of a major, nationwide payment reform in Medicare. We propose analyzing health care utilization and spending, health care quality, and

[17] They report a 9.5% decline in payments to skilled nursing facilities, and we calculate an MDE of about 8 percent here (not shown).

changes in volume and patient mix. We will focus primarily on outcomes during the episode of care covered by the payment reform, but also look at some longer-run outcomes beyond that window. We will focus on average effects but also explore heterogeneity along both pre-specified dimensions and the results of machine learning. We expect the results to serve as inputs into future work modeling how hospitals respond to these types of financial incentives.

References

American Hospital Association. 2016. "Medicare's Bundled Payment Initiatives: Consideration for Providers". Retrieved from: http://www.aha.org/content/16/issbrief-bundledpmt.pdf

Cooper, R.W., Haltiwanger, J.C. "On the nature of capital adjustment costs." *The Review of Economic Studies* 73.3 (2006): 611-633.

Dummit, L.A., Kahvecioglu, D., Marrufo, G., Rajkumar, R., Marshall, J., Tan, E., Press, M.J., Flood, S., Muldoon, L.D., Gu, Q., Hassol, A., Bott, D.M., Bassano, A., Conway, P.H. Association Between Hospital Participation in a Medicare Bundled Payment Initiative and Payments and Quality Outcomes for Lower Extremity Joint Replacement Episodes. *JAMA.* 2016;316(12):1267-1278. doi:10.1001/jama.2016.12717

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K. & Oregon Health Study Group. (2012). "The Oregon health insurance experiment: evidence from the first year." *The Quarterly journal of economics*, 127(3):1057-1106.

Gottlieb, Daniel J., Weiping Zhou, Yunjie Song, Kathryn G. Andrews, Jonathan S. Skinner, and Jason M. Sutherland. (2010) "Prices Don't Drive Regional Medicare Spending Variations." Health Affairs, 29(3): 537–543.

Gronniger, T., Fiedler, M., Patel, K., Adler, L., Ginsburg, P.B. (2017) "How Should The Trump Administration Handle Medicare's New Bundled Payment Programs?" http://healthaffairs.org/blog/2017/04/10/how-should-the-trump-administration-handle-medicares-new-bundled-payment-programs/#one

Gupta, A., "Impacts of performance pay for hospitals: The Readmissions Reduction Program" Working Paper available at http://www.web.stanford.edu/~atulg/Gupta_JMP.pdf

"Medicare Program; Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services", 80 Federal Register 134 (July 14, 2015), pp. 41198-41316.

"Medicare Program; Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services", 80 Federal Register 226 (November 24, 2015), pp. 73274-73554.

Newhouse, J., Garber, A., Graham, R., McCoy, M., Mancher, M., Kibria, A. (2013) “Variation in Health Care Spending: Target Decision Making, Not Geography.” Institute of Medicine. Available at http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2013/Geographic-Variation2/geovariation_rb.pdf

Skinner, Jonathan. (2011) “Causes and Consequences of Regional Variations in Health Care.” In Handbook of Health Economics. Volume 2, eds. Pauly, M.V., McGuire, T.G., Barros, P.P. 45–93. Elsevier.