

# Protocol

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

Protocol for: Barnett ML, Wilcock A, McWilliams JM, et al. Two-year evaluation of mandatory bundled payments for joint replacement. *N Engl J Med* 2019;380:252-62. DOI: 10.1056/NEJMsa1809010

**Comprehensive Care for Joint Replacement (CJR) Evaluation Analysis Protocol  
(Final Version)**

November 1, 2018

*The original protocol as posted prior to analysis of post-intervention CJR data (2016-2017) is included after this final version of the protocol.*

**Abbreviations**

BPCI	Bundled Payments for Care Improvement
CJR	Comprehensive Care for Joint Replacement Initiative
CMS	Center for Medicare and Medicaid Services
DRG	Diagnosis related group
ED	Emergency department
HHA	Home health agency
ICD-9 or ICD-10	International Statistical Classification of Diseases and Related Health Problems, revision 9 or 10
IRF	Inpatient rehabilitation facility
LEJR	Lower extremity joint replacement
MSA	Metropolitan statistical area
PAC	Post-acute care
SNF	Skilled nursing facility

## Addendum History

### Overview of changes:

We provide an overview of the changes to the protocol, because the changes may be hard to follow when presented in chronological order.

In March of 2018, we first posted the protocol. Before viewing the post-intervention data, we made a series of refinements of the protocol and our final pre-specified protocol was posted in May 2018 (this is the “original pre-analysis protocol” included below). After viewing the post-intervention data, we made a series of post-hoc changes that addressed both new issues we identified and reviewer feedback. The changes and their rationale are summarized in a table below. This updated protocol was posted in August of 2018.

In the revision process, reviewers requested that the primary results presented in the manuscript reflect the final pre-specified protocol with one exception. Given the nature of the randomization they felt it was important to include random effects for each MSA in our analytic model. In our final manuscript, we present these as primary results and we present the results from the updated protocol we posted in August of 2018 as a sensitivity analysis.

### Chronological description of changes:

- March 28, 2018: First version of protocol posted
- May 2, 2018: Modified version posted
  - Minor modifications to the definition of primary outcome #3 (predicted spending risk score)
    - Re-defined period for estimating risk prediction model from pre-intervention period (2014-2015) to an earlier period completely outside the data in the study (2012-2013)
    - Modified description of the pre-intervention period in section C.1. to reflect this
- May 9, 2018:
  - Analysis of main outcomes in the pre-period (2014-2015) showed diverging trends in the composite complications rate (primary outcome #2).
    - Trends show a decline in complication rate in treatment MSAs and stable complication rate in control MSAs
  - Added language to section D, primary outcome #2, describing how we will interpret changes in complication rates with our difference-in-differences models in light of these diverging trends.
- May 12, 2018: Modified version posted
  - Added additional clarifying language to the pre-intervention trend test description in section F
  - Changes to secondary outcomes
    - Added new outcome: rate of observation stays in 90d post-discharge period
    - Clarified denominator for secondary outcome #4, per capita LEJR rate
    - Clarified secondary outcome #11, length of stay in institution PAC

- Added additional sub-analyses
      - Hip vs. knee procedures
      - Non-profit vs. for-profit hospitals
- May 15, 2018: Modified version posted
  - Clarified that chronic condition codes for patients will be used from the beneficiary summary file from the prior calendar year, given the given of 12 months of prior continuous coverage for cohort inclusion.
  - We extended the post-intervention period to September 30, 2017 given new availability of all 4 quarters of 2017 data
  - Clarified that total spending outcomes will only be calculated for 2016 in the post-intervention period because of data availability (no Carrier file for 2017 available at the time of analysis)
  - Corrected model specification notation for equations (1) and (2) in Part F
- May 18, 2018: Modified version posted
  - Excluding index episodes where the initial claim was denied
  - Pre-intervention period shortened to 2015 alone due to data consistency issue
    - MedPAR file available 2014-2016, and Inpatient/SNF files available 2015-2017
    - To prioritize consistency in the source of outcomes data, will restrict study period to the time period when Inpatient and SNF files are available
    - If 2014 Inpatient/SNF data becomes available before submission or publication, will expand study period to originally proposed 2014-2015 pre-period
- May 24, 2018: minor revision with typo and formatting changes
- **June 2018 – Post-CJR implementation data available, analysis started**
- July 2, 2018
  - Made minor modifications to correct errors in timeline of revisions above
  - Correction: removed a note on total LEJR episode spending (secondary outcome #1) “Because of data availability, this will only be calculated for the post-intervention period in 2016,” which was no longer the case because 2015 data were available.
  - Made several post-hoc analytic changes, with a description of each change, rational and impact summarized in the table below:

July 2, 2018 Post-Hoc Change	Rationale	Impact on Results and Corresponding Table in Manuscript
Expanded time period for pre-period trend assessment from 2015 to 2011-2015	We initially planned to formally test for parallel trends using 2015 data where we had a 100% sample. To be more confident that the parallel trend tests were robust for a longer time period, we used the 20% Medicare sample for a full five years of data (2011-2015).	Pre-period trends assessed using either 2015 (with 100% sample) or 2011-2015 (with 20% sample) showed no significant deviation in trends violating the “parallel trends” assumption for difference-in-differences analysis
Examined differential changes in procedure characteristics	Changes in the prevalence of diagnosis related group (DRG) coding for 469 (with complications) vs. 470 (without complications) in led us to believe that coding changes deserved scrutiny as a possible effect of CJR implementation	We observed a non-significant differential increase in the prevalence of coding for DRG 469 (0.2 percentage points, 95% CI - 0.03-0.5)

Included an additional individual-level analysis of LEJR volume in addition to MSA-level analysis	Initially, we described analyses to assess for differences in volume of LEJR at the MSA level. We decided to add another model where we examine volume at the beneficiary level to enable adjustment for beneficiary characteristics at the individual level.	No significant change in LEJR procedure volume per 1,000 beneficiaries in either the individual-level or MSA-level analyses
Included multiple LEJR episodes per beneficiary	We recognized that limiting the initial analysis to the first recorded LEJR episode per beneficiary potentially biased the mix of first vs. repeat procedures across the time period. Early in the study period, it was more likely that procedures were repeat LEJR because we had less data to “look back” for prior procedures for each beneficiary. To limit the potential bias from a changing mix of initial vs. repeat procedures over the study period, we relaxed this restriction and included all LEJR episodes per beneficiary	No meaningful change in primary outcomes using multiple procedures vs. one procedure per beneficiary
Dropped DRG as a covariate from the main difference-in-differences models	Analyses showed modest differential changes in DRG 469 vs. 470 in the treatment group, which could reflect upcoding rather than true changes in the underlying patient population. To avoid this potential bias, we dropped DRG as a control variable.	No meaningful change in primary outcomes

- August 25, 2018
  - Minor clarifying edits
  - Updated the July 2 post-hoc changes with a clearer table to outline changes, rationale and impact
  - Added post-hoc outcome in response to reviewer comments:
    - Institutional post-acute care length of stay excluding IRFs, since IRFs are not paid based on length of stay, except in outlier cases.
- October 15, 2018
  - In response to reviewer feedback, we made the following changes:
    - The post-hoc changes described above in the July 2, 2018 that affect our primary difference-in-differences models (dropping DRG as a covariate and including all LEJR episodes) are now presented as a sensitivity analysis
    - The primary analysis reported in the manuscript will keep the original specification of using only the first LEJR episode per beneficiary and including DRG as a covariate in our difference-in-differences models
    - We are also changing the specification of our analytic models:
      - Originally pre-specified: hospital fixed effects with robust standard errors clustered at the MSA level
        - This is now reported as a sensitivity analysis
      - New primary analysis: hospital and MSA random effects
    - The significance level for the primary outcomes was redefined from  $P=0.05$  to a Bonferroni corrected  $P=0.05/3=0.0167$ .
  - The analysis plan below reflects these changes.

## **Study Team**

Project Leads: Michael L. Barnett, Andrew Wilcock

Co-Authors: J. Michael McWilliams, Arnold Epstein, Karen Joynt Maddox, E. John Orav, David Grabowski, Ateev Mehrotra

Department of Health Policy and Management, Harvard TH Chan School of Public Health,  
Boston MA (MLB, AE, EJO)

Department of Health Care Policy, Harvard Medical School, Boston, MA (AW, JMM, DG, AM)

Department of Medicine, Brigham and Women's Hospital, Boston, MA (MLB, JMM, AE, EJO)

Department of Internal Medicine, Washington University School of Medicine in St. Louis, St.  
Louis, MO (KJM)

## Part A: Introduction and Overview

Concerns about rising health care spending have led Medicare to pilot a number of episode-based payment programs, or “bundled” payments.” The hope is that these new designs will lower health care spending by incentivizing more efficient delivery of services across disconnected care settings. Early evidence suggests that the programs can be effective: bundled payment designs like the recent [Bundled Payments for Care Improvement](#) (BPCI) initiative have demonstrated that providers can reduce overall spending without lowering quality or adversely impacting patient outcomes. However, what we know from programs like the BPCI may be biased due to the fact that organizations volunteered as participants, making the generalizability of prior evaluations to widespread implementation unknown. In contrast, starting in 2016 the [Comprehensive Care for Joint Replacement](#) (CJR) program mandated hospitals in 67 randomly selected metropolitan statistical areas (MSAs) to adopt bundled payments for lower joint replacement procedures. In this pre-specified analytic protocol, we describe our plan to evaluate the impact of the CJR initiative over its first year.

### Overview

The CJR payment model bundles together the costs of a hospitalization for lower extremity joint replacement (LEJR) and all costs occurring 90 days post-discharge (with some exceptions noted below). The payment changes began on April 1<sup>st</sup>, 2016. CMS randomly selected 75 MSAs among a pool of 196 eligible MSAs. After this random selection, 8 MSAs were dropped due to post-randomization changes in eligibility made by CMS before program initiation (in short, updating the exclusion criteria to take account of hospitals newly enrolled in another voluntary bundled payment program, the Bundled Payments for Care Improvement initiative). No information was provided by CMS on which MSAs would have been dropped from the “control group” (non-selected MSAs) if applying the same criteria. However, to preserve the integrity of randomization, the main analytic approach in this evaluation will be intention-to-treat, considering the originally randomized 75 MSAs as the treatment group. In secondary analyses we will focus on the 67 MSAs where the payment change was implemented.

Our three primary endpoints are total institutional (i.e. non-Medicare Part B) spending per LEJR episode, rates of post-surgical complications, and risk profiles of patients receiving a LEJR (as measured by average predicted spending). Our secondary outcomes will focus on changes in utilization, other clinical outcomes (e.g. readmissions, mortality), and patterns of post-acute care use.

Due to the strategy used by CMS to randomly select MSAs for participation in CJR (more detail below), there remain some systematic differences between the treatment and control MSAs at baseline. Given these exogenous baseline differences, our strategy is to use difference-in-differences to estimate the differential change in outcomes for CJR treatment MSAs versus secular changes in control MSAs. We propose several sensitivity analyses to examine the robustness of our findings.

## Part B: Study Objectives

Key Objective 1: Evaluate the effect of CJR bundled payments on LEJR episode spending.

Hypothesis 1: *CJR will be associated with decreased facility/institutional spending per LEJR episode.*

Key Objective 2: Evaluate the effect of CJR bundled payments on patient outcomes including complications.

Hypothesis 2: *Spending decreases in CJR participating hospitals will not be associated with a clinically meaningful change in outcomes*

Key Objective 3: Examine whether CJR bundled payments led to a change in the selection of patients for LEJR procedures.

Hypothesis 3: *CJR participation will be associated with a decrease in the average predicted spending for patients receiving LEJR.*

### Secondary Objectives

- 1) Evaluate the impact of CJR on separate components of Medicare spending (e.g. inpatient, outpatient, SNF, HHA)
- 2) Evaluation the impact of CJR on other patient outcomes, including readmissions, ED visits, and mortality.
- 3) Evaluate the effect of CJR bundled payments on the use of institutional post-acute care after discharge.
- 4) Evaluate the effect of CJR on patient selection on specific attributes (e.g. age, chronic conditions, dual eligible status).

## Part C: Study Design

### C.1. Time Period Definitions

This evaluation will analyze CJR implementation as a pragmatic, government implemented cluster randomized trial comparing study endpoints for LEJR episodes in hospitals located in treatment MSAs (‘treatment’ or ‘intervention’ group) vs. those located in control MSAs (‘control’ group) in the post-implementation period vs. the pre-implementation period.

We will define the following time periods **based on admission date**:

2015												2016												2017											
1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Pre-Intervention												Excluded			Washout			Post-Intervention									Follow-Up								
												↑			↑																				
												Final Rule Published			CJR Start																				

- 1) Pre-intervention: January 1, 2015- December 31, 2015 (12 months)
  - Time period used for assessing patient, hospital and MSA characteristics
- 2) Exclusion and “washout” periods: January 1, 2016-June 30, 2016 (6 months)
  - We will exclude January 1, 2016-March 31, 2016 (3 months) because LEJR episodes beginning in this period will cross over between CJR implementation and non-implementation periods.
  - We will also drop the first 3 months after CJR implementation (April 1, 2016 – June 30, 2016) from the evaluation as a “washout” period for hospitals to adapt to new payment model. Because hospitals were mandated to enroll in CJR, hospitals may be poorly prepared initially to respond to payment incentives. We will also examine this period in sensitivity analyses.
- 3) Post-intervention: July 1, 2016-September 30, 2017 (15 months)
  - This 15-month period includes the full post-washout implementation period through the most recently available data at the time of analysis
  - Data will be available through December 31, 2017 to capture 90-day post discharge episode claims
  - Any episodes whose 90-day post discharge period ends after December 31, 2017 will be excluded.
    - It is possible that late 4<sup>th</sup> quarter 2017 (October-December) data may be partially incomplete due to the time needed for CMS to finalize claims. If we observe evidence of this, we may revise the definition of the post-period to accommodate these data limitations.

### C.2. Study Population:

The entire study population will consist of all eligible LEJR episodes, as defined below, in any of 196 CJR-eligible MSAs with an admission date from January 1, 2014 until September 30, 2017 (with the exception of admissions during the excluded period defined above from January 1, 2016-June 30, 2016).

The unit of randomization is the MSA, with 75 “treatment” MSAs (9 of which “dropped out” from the intervention prior to implementation due to rule changes, see below) exposed to bundled payments beginning in April 1, 2016, and 121 “control” MSAs with no payment changes. The unit of observation/analysis will be the LEJR episode, with only one LEJR episode included per Medicare beneficiary.

Treatment and Control Group Definitions:

- Treatment vs. Control group is defined by the MSA of a hospital where a LEJR procedure takes place (regardless of beneficiary residence)
- CMS identified 196 MSAs eligible for randomization based on the following inclusion criteria:
  1. MSA had more than 400 LEJR episodes from 7/1/2013 to 6/30/2014
  2. MSA had more than 400 non-Bundled Payments for Care Improvement (BPCI) LEJR episodes from 7/1/2013 to 6/30/2014
  3. MSA had less than 50% percent of LEJR episodes included in any BPCI model
  4. MSA had >50% of LEJR episodes paid through the inpatient prospective payment system (e.g. excluding Maryland, critical access hospitals, etc.)
- Out of these 196 MSAs:
  - 75 treatment MSAs selected in July 2015 interim final rule
  - 121 MSAs in control group
  - 196 MSAs were divided into 8 strata defined by quartiles of pre-period episode spending and above vs. below median population (Table below)
  - Probability of MSA selection depended on pre-period average LEJR 90-day episode spending as shown in the Table below
- The treatment group was reduced from 75 MSAs to 67 MSAs in November 2015 due to updating BPCI enrollment lists up to October 2015 (i.e. 8 MSAs fell out of inclusion criteria #2 and #3 above due to increased BPCI enrollment)
  - Because this change disrupts initial randomization, we define the treatment group as the initially randomized 75 MSAs with an “intent-to-treat” approach

**Table: MSA Selection by Payment Quartile and Above/Below Median Population Criteria, as Defined by CMS**

	Payment in lowest quarter	Payment in 2nd lowest quarter	Payment in 3rd lowest quarter	Payment in highest quarter	Total eligible MSAs
Selection Proportion .....	30%	35%	40%	45%	.....
Less Than Median Population (Group #) .....	(1)	(2)	(3)	(4)	.....
Number Eligible MSAs per Proposed Rule (80 FR 41198) .....	33	19	22	24	98
Proportion x Number .....	9.9	6.65	8.8	10.8	
Number initially selected from group .....	10	7	9	11	37
Number finally selected from group .....	8	6	8	11	33
More Than Median Population (Group #) .....	(5)	(6)	(7)	(8)	
Number Eligible MSAs per NPRM .....	16	30	27	25	98
Proportion x Number .....	4.8	10.5	10.8	11.25	
Number initially selected from group .....	5	11	11	11	38
Number finally selected from group .....	5	10	9	10	34
Total Eligible MSAs per Proposed Rule (80 FR 41198) .....	49	49	49	49	196
Number initially selected .....	15	18	20	22	75
Number finally selected from group .....	13	16	17	21	67

Source: Reproduced from CJR Final Rule, Federal Register 80 FR 73298 (<https://www.federalregister.gov/d/2015-29438/page-73298>)

Index Episode Definition:

- Only one LEJR episode included per Medicare beneficiary (the earliest)
  - o In a sensitivity analysis, we will examine all eligible LEJR episodes, including repeat episodes within beneficiary
- Episodes will be assigned to a time period defined in C.1. above by hospital admission for LEJR, which is the episode start date.
- As defined by CMS for the CJR program, LEJR episodes will be defined by discharge diagnosis related group [DRG]):
  - o DRG 469 (major joint replacement with major complications/comorbidities)
  - o DRG 470 (major joint replacement without major complications/comorbidities)
  - o The DRGs above are inclusive of hip and knee replacements, including procedures due to hip fractures.
  - o In the case of hospital transfers, the episode payment is based with the hospital discharging the patient with the above DRGs.

Other Episode Inclusion Criteria:

- Continuously enrolled in Medicare A+B for 12 months prior and throughout 90-day LEJR episode (Based on 2014 Medicare claims, roughly 6% of episodes are lost due to mandating 12 mo of prior coverage)
  - o We will include beneficiaries who die during the LEJR episode as long as they are continuously enrolled in Medicare A+B until the date of death

Episode Exclusion Criteria:

- o Per CJR regulations, if there are overlapping joint replacement episodes (e.g. a patient receives two knee replacements 80 days apart), both episodes will be excluded.

- Currently eligible for Medicare due to end-stage renal disease (ESRD)
- Hospital opened after 2015 (i.e., no pre-period)
- Index hospitalization claim denied (likely wrong DRG)
- BPCI Participation
- Hospital is an episode initiator for LEJR in BPCI Model 1, 2 or 4
- Patient discharged to BPCI Model 3 participating PAC provider for LEJR
  - If any portion of an episode overlaps with the onset of BPCI participation for LEJR episodes, it will be excluded
  - Note: We are unable to exclude patients in physician group practices that are BPCI Model 2 episode initiators due to lack of available information (i.e. National Provider Identifiers of participating physicians and attribution rules) to accurately assign patients to these physician groups

#### Other Study Population Notes:

The CJR program excludes any LEJR episodes with any mortality from the payment model. We will not make this exclusion for two reasons. First, within-episode mortality is an important secondary outcome in our evaluation as well as part of the composite primary outcome of complications. Second, even though episodes with mortality are excluded from bundled payments their total cost and patterns of care are still a reflection of hospitals' potentially shifting approach to LEJR care due to the CJR payment model. Therefore, we will include these episodes to capture a more complete picture of patient outcomes following CJR implementation.

### **Part D: Study Endpoints and Data Sources**

#### Data Sources

We will use the following data sources for this evaluation (data source purpose in parenthesis):

1. Publicly Available CJR Enrollment Files (treatment vs. control assignment)
2. Medicare Claims Files – 2015-2017 (primary and secondary endpoints)
  - a. 100% sample available for inpatient, post-acute care and hospital outpatient files
  - b. 20% sample available for Part B “Carrier” file
3. Medicare Denominator Files – 2014-2016 (patient characteristics, 2017 unavailable at the time of analysis)
4. Medicare Provider of Services File (hospital characteristics)
5. Area Health Resource File – 2015 (MSA characteristics)
6. BPCI Participation Database (exclusion of hospitals and episodes)

#### Primary Endpoints

##### Primary Outcome #1: Facility/Institutional Spending per Episode

- We will use the full unadjusted payments (aka, spending) to providers (including Medicare, patient and primary payer portions) that include common CMS adjustments such as wage index

- Rationale: we are interested in the pragmatic national impact of the program on spending, regardless of regional and service-specific adjustments that Medicare applies
- As secondary outcomes, we will also examine standardized institutional costs and total costs (see below)
- Any denied payments will be excluded from spending outcomes
- We will examine the potential influence of extreme outlier episode payments, and potentially Winsorize episodes with extreme payments if a small number of episodes (less than 0.5% of total episodes) exert a large influence on the final results. If we do Winsorize, then the analyses will be presented with and without Winsorizing.
- The payments for any services that cross over the end of the LEJR episode (e.g. SNF admission or HHA episode) will be pro-rated based on the percentage of days occurring within the episode
  - For example, if a HHA service begins 86 days post discharge (i.e. 5 days left in the episode) and lasts 20 days, 5/20, then 25% of the HHA payment will be attributed to the LEJR episode.
- In the episode reconciliation process, CMS removes LEJR “unrelated” costs from calculated episode costs (see CMS website for [more documentation](#)). We are not applying these exclusions to capture a broader picture of total institutional spending per episode.

#### Primary Outcome #2: TKA/THR complications rate

- Definition: we will capture all complication events as defined in NQF Measure #1550, including hip fracture patients
- We used Measure #1550 because it is endorsed by the NQF, employed by CMS for quality reporting, and it is used by CJR to set quality thresholds for receiving bonuses
- While we will capture complications as described by this measure, we will define the sample differently. We will include all episodes for the primary analysis as defined above in Section C.2.
  - We chose a different sample, because the measure defines its sample based on the presence of ICD-9 procedure codes 81.51 (total hip replacement) and 81.54 (total knee replacement), not DRGs 469 and 470 as above.
  - The NQF measure also makes a number of exclusions to focus on elective joint replacements, mainly excluding hip fracture patients
    - As a secondary outcome, we will additionally examine complication rates using the exact sample definition in the NQF defined measure
- This measure, as defined by CMS and used in CJR, captures a composite of the following outcomes:
  - Within 7 days of admission
    - Myocardial infarction
    - Pneumonia
    - Sepsis/shock
  - Within 30 days of admission

- Surgical site bleeding
- Pulmonary embolism
- All-cause mortality
- Within 90 days of admission
  - Implant mechanical complications
  - Joint/wound infection
- May 9, 2018 Update:
  - Pre-analysis of this outcome showed diverging trends between treatment and control MSAs, making interpretation of our difference-in-differences study design more difficult
  - Given this issue, we will only report a positive effect of the program on complications if the effect we observe is greater than what we expect from the pre-period AND an outlier on permutation tests (see section F below)
  - Additional sensitivity analyses may be necessary to examine the potential effect of CJR on complication rates beyond trends present prior to program start

Primary Outcome #3: Patient risk score, estimated as predicted LEJR episode spending

- The purpose of this outcome is to capture patient selection by hospitals subsequent to CJR implementation in a single, integrated measure
- If we find clinically meaningful evidence of patient selection in this outcome, it may substantially shift our interpretation of any findings in primary outcomes #1 and #2, given that we cannot exclude that any effects observed are due to unobserved changes in patient case mix.
- Individual patient characteristics (e.g., age) will also be examined as secondary outcomes, but as a primary outcome we will estimate patients' predicted spending
- Using data from the before the full study period (2013-2014) for a 20% sample of beneficiaries, we will estimate a linear regression model with the dependent variable of predicted total episode spending
  - We will use the 20% sample to be able to include estimated Part B ("carrier") costs, which are only available for a 20% sample.
  - Estimates from this model will be used to generate predicted spending for each episode in our study sample
- Patient covariates in the model will include:
  - Age (in 5 year bins up until 90+)
  - Gender
  - Race (white/black/other)
  - Original reason for Medicare enrollment: disability, age, ESRD
  - Medicaid status (any Medicaid enrollment in prior 12 months)
  - Prior inpatient admission in prior 12 months
  - Any institutional PAC (SNF, IRF, LTAC) admission in prior 12 months
  - 27 chronic conditions as captured in the Chronic Condition Warehouse file from prior calendar year
- If the model has adequate predictive power without the chronic illness indicators, we may forgo including them to avoid the possibility that treatment hospitals intentionally "up-code" their patients to get improving quality ratings.

- If needed to aid interpretation of this measure, we may use an alternate specification of this measure, such as classifying patients into quartiles of predicted spending or creating a risk score scaled from 0-100

## Secondary Outcomes

### *Spending and Utilization*

- 1) Total LEJR episode spending. Our data sample has only Part B claims for a 20% sample. We will include facility/institutional and Part B claims in this measure.
- 2) LEJR institutional spending for 100% sample in which we use price standardization to remove Medicare fee adjustments
  - Because of data availability, this analysis would only include 2015 in the pre-period
- 3) 90-day post-discharge spending in mutually exclusive categories (100% sample unless otherwise specified)
  - Index hospitalization
  - Repeat inpatient, ED and observation spending in 90-days
  - Post-acute care spending: SNF, IRF, and HHA
  - Other Medicare Part B spending (20% sample)
- 4) Number of LEJR episodes per 1,000 beneficiaries
  - LEJR episodes per 1,000 beneficiaries at the MSA level
  - *Post-hoc change*: Also LEJR episodes per 1,000 beneficiaries at the individual level
  - Denominator defined as all beneficiaries in an MSA-quarter with continuous FFS Medicare coverage for that quarter
- 5) 90-day post hospital discharge all-cause hospital readmission rates
  - Binary outcome, 1 = any inpatient admission in the 90-day window
  - Excluding certain planned admissions as defined in the CMS hospital-wide readmission measure (NQF #1789)
- 6) 90-day post hospital discharge all cause observation visit rates without admission
  - Binary outcome, 1 = any observation stay in the 90-day window without admission
- 7) 90-day post hospital discharge all cause emergency department rates without admission
  - Binary outcome, 1 = any emergency department visit in the 90-day window without admission
- 8) 90-day post hospital discharge all cause hospital visit of any kind (emergency department visit, observation stay or inpatient admission)
  - Binary outcome, 1 = any hospital visit in the 90-day window
- 9) 90-day all-cause mortality
- 10) LEJR complications rate as defined in primary outcome #2 with CMS-defined exclusion criteria (as opposed to using full study sample defined in C.2.)
- 11) Proportion of LEJR admissions discharged to home vs. different PAC settings:
  - Proportion discharged to institutional PAC
  - Proportion discharged home
  - Proportion discharged to home health agencies

- 12) Average length of stay among those admitted to institutional PAC settings during full episode among those with at least 1 day of institution PAC use
- *Post hoc change*: Institutional post-acute care length of stay excluding IRFs, since IRFs are not paid based on length of stay, except in outlier cases.

#### *Patient Selection*

Examine the proportion of patients in the treatment and control groups in the pre vs. post-periods with the following characteristics:

- 1) Sex
- 2) Age: 65-79, 80-89, 90+
- 3) Race
- 4) Urban/rural residence
- 5) Medicaid eligibility
- 6) Disability as original reason for enrollment
- 7) Prior inpatient admission in prior 12 months
- 8) Any institutional PAC (SNF, IRF, LTAC) admission in prior 12 months
- 9) *Post-hoc additions*:
  - a. Diagnosis related group 469 or 470
  - b. Presence of hip fracture
  - c. Type of procedure (defined by ICD-9 procedure code)

#### **Part E: Episode, Patient, Hospital, and MSA Characteristics**

We will examine the characteristics listed below for balance on observable characteristics across the treatment and control groups in the 24-month pre period. This list may be not inclusive of all characteristics examined in the final analysis. Other characteristics not specified below will be marked as post-hoc additions.

Episode characteristics:

- Diagnosis related group 469 or 470 (*excluded as a model covariate post-hoc in sensitivity analyses*)
- Presence of hip fracture
- Type of procedure (defined by ICD-9 procedure code)

Patient characteristics:

- Age (in 5 year bins up until 90+)
- Gender
- Race (white/black/other)
- Urban/rural residence
- Original reason for Medicare enrollment: disability, age (ESRD excluded)
- Medicaid status (any Medicaid enrollment in prior 12 months)
- CCW conditions (27 total, used from prior calendar year)
  - Number of conditions
  - Proportion with  $\geq 6$  or  $\geq 9$  conditions

Hospital characteristics:

- Ownership type (profit, government, non-profit)
- Size (<100, 100-249, >250 beds)
- Teaching status (non-teaching, minor, major or unknown)
- Census region (North, South, Midwest, West)

- DSH percent (in quartiles)
  - Medicare accountable care organization participation
  - Pre-implementation average 90-day LEJR spending
- MSA characteristics (averages of county-level measures within MSA):
- Median household income
  - % Medicare advantage penetration
  - % penetration of BPCI LEJR episodes
  - SNF total beds

## **Part F: Statistical Analysis**

We will follow an intent-to-treat (ITT) approach for all main analyses in which we include all 75 MSAs randomized to the intervention (8 of these MSAs were subsequently dropped). This is to best preserve the initial randomization of MSAs to treatment groups performed by CMS. Secondary analyses using instrumental variables approaches will be considered to calculate a treatment-on-treated (ToT) effect estimate on the primary outcomes.

### Procedures for Missing Data

Because the data for this project will primarily use administrative Medicare claims data, we anticipate little to no missing data, particularly for the primary outcomes. If missingness for a key covariate exceeds 1-2% and is deemed important for the analysis, we will use multiple imputation to account for the missing data.

### Statistical Significance Levels

Initially, the overall significance level for the study was set at  $P < 0.05$  for the three main outcomes. In response to reviewer feedback, we modified this threshold to a Bonferroni-corrected  $P < 0.0167$  ( $0.05$  divided by 3 primary outcomes). For the multiple secondary analyses, these tests will be considered as exploratory and we did not present P-values.

### Evaluation of Randomization/Balance Across Study Arms

We will evaluate balance across the episode, patient, hospital and MSA characteristics listed in Part E between the treatment and control groups using data from the pre-implementation period. We will use chi-squared and Student's t-tests as appropriate for two-sided tests of differences in means across the study arms. However, given the large sample size of episodes we anticipate in each arm, statistical significance will be only one factor in evaluating the adequacy of balance between the groups as clinically insignificant results may demonstrate statistically significant mean differences.

#### *Accounting for Variable Treatment Assignment Probability*

As shown in Part C.2. above, MSAs had different probabilities of being assigned to the treatment group depending on their quartile of average LEJR spending from July 1, 2013 to June 30, 2014 (as assessed by CMS). Because pre-period spending will be highly

correlated with post-period spending, one of our primary outcomes, this could create baseline imbalance between treatment and control groups in unadjusted comparisons. We will adjust for treatment probability stratum in the statistical models for adjudicating outcomes below, which will provide within-stratum estimates of the effect of treatment assignment.

However, for assessing randomization, these differing treatment probabilities could be misleading about the effectiveness of randomization. To account for the varying probabilities of treatment assignment, in addition to unweighted balance tables, we will also produce weighted balance tables that weight episodes, hospitals and MSAs such that the probability of the treatment or control MSAs being selected are equal within each stratum (i.e. analogous to “direct standardization”). The weights will be chosen to match each stratum to the treatment/control probability of the entire sample. The assignment probabilities and weights used are summarized in the Table below:

Stratum	Payment Quartile	Above/Below Median Population	Initial Treatment MSAs	Control MSAs	Treatment Probability	Control Probability	Treatment Weight (Overall Prob/Stratum Prob)	Control Weight (Overall Prob/Stratum Prob)
<b>Overall</b>	-	-	<b>75</b>	<b>121</b>	<b>0.383</b>	<b>0.617</b>	-	-
1	1 (low)	Below	10	23	0.303	0.697	1.263	0.886
2	1 (low)	Above	5	11	0.313	0.688	1.224	0.898
3	2	Below	7	12	0.368	0.632	1.039	0.977
4	2	Above	11	19	0.367	0.633	1.044	0.975
5	3	Below	9	13	0.409	0.591	0.935	1.045
6	3	Above	11	16	0.407	0.593	0.939	1.042
7	4 (high)	Below	11	13	0.458	0.542	0.835	1.140
8	4 (high)	Above	11	14	0.440	0.560	0.870	1.102

#### Unadjusted and Adjusted Statistical Analyses

The main model will use a “difference-in-differences” estimator to calculate the differential change in outcomes in the treatment vs. control groups in the post-period. The advantage of this model is adding an extra layer of robustness to reduce potential bias from imbalance in pre-period characteristics or outcomes between the treatment and control groups.

Our main pre-specified analysis to report the primary outcomes will be a regression model adjusted for patient characteristics and a hospital fixed effect. However, we will also compare these results to unadjusted statistical analyses (i.e., without the  $X_i$  term in equation (1) below), and regression models that use hospital and MSA characteristics in place of the hospital fixed effect.

For the unadjusted model, to test the hypotheses stated in Part B, we will estimate linear regression model with episodes ( $i$ ) nested within hospitals ( $j$ ) nested within MSAs ( $k$ ) as follows:

$$y_{ijk} = \beta_0 + \beta_1 \text{TREAT}_k * \text{POST} + \text{TREAT}_k + \text{POST} + X_i \beta_2 + \eta_j + \mu_k \tau_i + \varepsilon_{ijk} \quad (1)$$

In equation (1),  $y_{ijk}$  is a continuous or dichotomous outcome corresponding to one of the primary or secondary outcomes specified in Part D.  $\beta_0$  is the intercept and  $\beta_1$  is the key coefficient on the intent-to-treat differential effect of treatment assignment vs. control in the post period ( $TREAT_k*POST$ ) of an episode as the MSA level.  $POST$  is a dummy variable specifying whether an episode occurred in the post-implementation period and  $TREAT$  is an indicator of whether an episode occurred in a treatment vs. control MSA. In the model with fixed effects, there are no main effects for  $POST$  and  $TREAT$  included because these are incorporated into hospital and quarter fixed effects as below.

$X_i$  is a vector of episode/patient characteristics, all specified above in Part E (except  $DRG$ , which was excluded post-hoc).  $\tau_i$  is a set of quarter fixed effects,  $\eta_j$  is a set of hospital-level random effects and  $\mu_k$  is a set of MSA-level random effects. The addition of random effects controls for any time-invariant differences between hospitals and MSAs. Adding these effects adds statistical power to the analysis by accounting for random differences between the treatment and control groups that arise by chance in the randomization process.

We will also examine an alternate model specification that we described in the original pre-analysis version of this protocol. This specification matches the equation above with the following differences: (1) we used hospital fixed effects with clustered standard errors at the MSA level instead of hospital and MSA random effects, (2) the study sample included multiple LEJR procedure per beneficiary and (3)  $DRG$  was dropped as a covariate.

#### Testing for pre-intervention differences in trend

One of the basic assumptions behind difference-in-differences models is that in the absence of the treatment effect, both treatment and control groups would have stable and parallel trends in outcomes. To test this hypothesis, we will examine unadjusted models similar to equation (1) above, where a linear term for time (in quarters or years) will be used instead of the  $POST$  indicator. To capture a fuller pre-period trend using 2015 alone, we will use the 20% Medicare sample for 2011-2015 to assess these trends. Significant differences in the  $\beta_1$  coefficient will be used to evaluate the parallel trends assumption.

In prior work we have found that testing for statistical differences in pre-trend may not be sufficient given that even small *insignificant* differences can bias findings. The differences in pre-trend should also be assessed in the context of the size of the difference in outcomes.

If we observe any differences in pre-trends that could change our interpretation of the results we will employ a permutation-based approach for our estimates. More specifically, if the estimated impact of CJR observed in our primary and secondary outcomes are small enough that they could be reasonably impacted by differential changes in pre-intervention trends we will use permutation tests.

This involves generating 100-1,000 samples of the MSAs in our study and randomly assigning treatment and control classification based on the distribution in the **Table** in Section C.2. Using these randomly permuted combinations of treatment and control, we will generate a distribution of estimates for the effect of interest. If the effect we observe is due to CJR treatment, then the

effect we observe should be in the tail, ideally in the upper/lower 2.5<sup>th</sup> percentile or less, of the distribution of effect sizes from all the permuted samples. This permutation test may also be employed if the differences in the primary outcomes are marginal in terms of statistical significance.

The concept behind this approach is that instead of discarding the difference-in-differences estimator entirely, we will instead assess how much more extreme the effect we observe is vs. random allocation of treatment and control assignment.

### Instrumental Variables Analysis

Because not all MSAs initially assigned to participate in CJR bundled payments actually ended up in the new payment model (8 MSAs dropped due to CMS regulation changes), our estimates of  $\beta_1$  in Equation (1) above will be biased towards the null. To ask the question of the causal impact of CJR bundled payments for those MSAs that did participate, we will use instrumental variables (IV) analysis.

We will use assignment to the treatment group as our instrument for actual bundled payment implementation. Treatment group assignment is a valid instrument because:

- Study arm was randomly assigned and thus uncorrelated with outcome
- Study arm is assumed to have an effect on outcomes only through payment policy (i.e. study arm assignment meets the “exclusion criterion” for a valid instrument)
- Study arm is highly predictive of bundled payment implementation

The causal effect of bundled payment participation on the outcomes of interest in this analysis will be modeled along the lines of Equation (1) above, with the replacement of TREAT with an indicator for true exposure to bundled payments, BUNDLE:

$$y_{ijk} = \pi_0 + \pi_1 \text{BUNDLE}_k * \text{POST} + X_i \pi_2 + \eta_j + \tau_i + \varepsilon_{ijk} \quad (2)$$

We will estimate equation (2) by two stage least squares (2SLS), using the following first stage equation:

$$\text{BUNDLE}_k = \delta_0 + \delta_1 \text{TREAT}_k + \delta_3 \text{STRATUM}_k + \mu_{ijk} \quad (3)$$

In which the excluded instrument is the variable TREAT with the first stage coefficient of  $\delta_1$ .

Because the model is just identified through randomizations, the 2SLS estimate of  $\pi_1$  is given by the ratio of the reduced form in Equation (1) and first stage coefficients ( $\beta_1/\delta_1$ ).

### Planned Subgroup and Sensitivity Analyses

We plan to investigate other alternative specifications of our analyses above in subgroups of interest. Subgroups and sensitivity analyses may include but are not limited to the following. Any other further analyses conducted that are not specified below will be marked as post-hoc in the final analysis:

- Subgroup Analyses
  - Elective LEJR vs. hip fracture
  - Hip vs. knee replacement
  - Hospital subgroups:
    - ACO hospital participants
    - Teaching vs. non-teaching hospitals
    - For-profit vs. non-profit hospitals
    - Large (>250 beds) vs. smaller (<250 beds) hospitals
    - Top quartile of pre-period LEJR episode volume vs. lowest quartile
    - Top quartile of pre-period average episode costs vs. lowest quartile
- Sensitivity Analyses
  - Include the initial washout period in our main analyses
  - Generalized linear models with log-link function or other functional forms for spending outcomes
  - Logistic regression models for binary outcomes
  - Post-period treatment estimator alone without difference-in-differences if pre-period data are adequately balanced

*Original protocol posted online prior to analysis of CJR post-period data*

**Comprehensive Care for Joint Replacement (CJR) Evaluation Analysis Protocol**

**Version:** May 18, 2018

## Addendum History

- March 28, 2018: First version of protocol posted
- May 2, 2018: Modified version posted
  - Minor modifications to the definition of primary outcome #3 (predicted spending risk score)
    - Re-defined period for estimating risk prediction model from pre-intervention period (2014-2015) to an earlier period completely outside the data in the study (2012-2013)
    - Modified description of the pre-intervention period in section C.1. to reflect this
- May 9, 2018: Modified version posted
  - Analysis of main outcomes in the pre-period (2014-2015) showed diverging trends in the composite complications rate (primary outcome #2).
    - Trends show a decline in complication rate in treatment MSAs and stable complication rate in control MSAs
  - Added language to section D, primary outcome #2, describing how we will interpret changes in complication rates with our difference-in-differences models in light of these diverging trends.
- May 12, 2018: Modified version posted
  - Added additional clarifying language to the pre-intervention trend test description in section F
  - Changes to secondary outcomes
    - Added new outcome: rate of observation stays in 90d post-discharge period
    - Clarified denominator for secondary outcome #4, per capita LEJR rate
    - Clarified secondary outcome #11, length of stay in institution PAC
    - Added additional sub-analyses
      - Hip vs. knee procedures
      - Non-profit vs. for-profit hospitals
- May 15, 2018: Modified version posted
  - Clarified that chronic condition codes for patients will be used from the beneficiary summary file from the prior calendar year, given the given of 12 months of prior continuous coverage for cohort inclusion.
  - We extended the post-intervention period to September 30, 2017 given new availability of all 4 quarters of 2017 data
  - Clarified that total spending outcomes will only be calculated for 2016 in the post-intervention period because of data availability (no Carrier file for 2017 available at the time of analysis)
  - Corrected model specification notation for equations (1) and (2) in Part F
- May 18, 2018: Modified version posted
  - Excluding index episodes where the initial claim was denied
  - Pre-intervention period shortened to 2015 alone due to data consistency issue
    - MedPAR file available 2014-2016, and Inpatient/SNF files available 2015-2017
    - To prioritize consistency in the source of outcomes data, will restrict study period to the time period when Inpatient and SNF files are available

- If 2014 Inpatient/SNF data becomes available before submission or publication, will expand study period to originally proposed 2014-2015 pre-period
- May 22, 2018: minor revision with typo and formatting changes

## **Study Team**

Project Leads: Michael L. Barnett, Andrew Wilcock

Co-Authors: Arnold Epstein, J. Michael McWilliams, Karen Joynt Maddox, E. John Orav, David Grabowski, Ateev Mehrotra

Department of Health Policy and Management, Harvard TH Chan School of Public Health,  
Boston MA (MLB, AE)

Department of Health Care Policy, Harvard Medical School, Boston, MA (AW, JMM, DG, AM)

Department of Medicine, Brigham and Women's Hospital, Boston, MA (EJO)

Department of Internal Medicine, Washington University School of Medicine in St. Louis, St.  
Louis, MO (KJM)

**Abbreviations**

BPCI	Bundled Payments for Care Improvement
CJR	Comprehensive Care for Joint Replacement Initiative
CMS	Center for Medicare and Medicaid Services
DRG	Diagnosis related group
ED	Emergency department
HHA	Home health agency
ICD-9 or ICD-10	International Statistical Classification of Diseases and Related Health Problems, revision 9 or 10
IRF	Inpatient rehabilitation facility
LEJR	Lower extremity joint replacement
MSA	Metropolitan statistical area
PAC	Post-acute care
SNF	Skilled nursing facility

## Part A: Introduction and Overview

Concerns about rising health care spending have led Medicare to pilot a number of episode-based payment programs, or “bundled” payments.” The hope is that these new designs will lower health care spending by incentivizing more efficient delivery of services across disconnected care settings. Early evidence suggests that the programs can be effective: bundled payment designs like the recent [Bundled Payments for Care Improvement](#) (BPCI) initiative have demonstrated that providers can reduce overall spending without lowering quality or adversely impacting patient outcomes. However, what we know from programs like the BPCI may be biased due to the fact that organizations volunteered as participants, making the generalizability of prior evaluations to widespread implementation unknown. In contrast, starting in 2016 the [Comprehensive Care for Joint Replacement](#) (CJR) program mandated hospitals in 67 randomly selected metropolitan statistical areas (MSAs) to adopt bundled payments for lower joint replacement procedures. In this pre-specified analytic protocol, we describe our plan to evaluate the impact of the CJR initiative over its first year.

### Overview

The CJR payment model bundles together the costs of a hospitalization for lower extremity joint replacement (LEJR) and all costs occurring 90 days post-discharge (with some exceptions noted below). The payment changes began on April 1<sup>st</sup>, 2016. CMS randomly selected 75 MSAs among a pool of 196 eligible MSAs. After this random selection, 8 MSAs were dropped due to post-randomization changes in eligibility made by CMS before program initiation (in short, updating the exclusion criteria to take account of hospitals newly enrolled in another voluntary bundled payment program, the Bundled Payments for Care Improvement initiative). No information was provided by CMS on which MSAs would have been dropped from the “control group” (non-selected MSAs) if applying the same criteria. However, to preserve the integrity of randomization, the main analytic approach in this evaluation will be intention-to-treat, considering the originally randomized 75 MSAs as the treatment group. In secondary analyses we will focus on the 67 MSAs where the payment change was implemented.

Our three primary endpoints are total institutional (i.e. non-Medicare Part B) spending per LEJR episode, rates of post-surgical complications, and risk profiles of patients receiving a LEJR (as measured by average predicted spending). Our secondary outcomes will focus on changes in utilization, other clinical outcomes (e.g. readmissions, mortality), and patterns of post-acute care use.

Due to the strategy used by CMS to randomly select MSAs for participation in CJR (more detail below), there remain some systematic differences between the treatment and control MSAs at baseline. Given these exogenous baseline differences, our strategy is to use difference-in-differences to estimate the differential change in outcomes for CJR treatment MSAs versus secular changes in control MSAs. We propose several sensitivity analyses to examine the robustness of our findings.

## **Part B: Study Objectives**

Key Objective 1: Evaluate the effect of CJR bundled payments on LEJR episode spending.

Hypothesis 1: *CJR will be associated with decreased facility/institutional spending per LEJR episode.*

Key Objective 2: Evaluate the effect of CJR bundled payments on patient outcomes including complications.

Hypothesis 2: *Spending decreases in CJR participating hospitals will not be associated with a clinically meaningful change in outcomes*

Key Objective 3: Examine whether CJR bundled payments led to a change in the selection of patients for LEJR procedures.

Hypothesis 3: *CJR participation will be associated with a decrease in the average predicted spending for patients receiving LEJR.*

### Secondary Objectives

- 1) Evaluate the impact of CJR on separate components of Medicare spending (e.g. inpatient, outpatient, SNF, HHA)
- 2) Evaluation the impact of CJR on other patient outcomes, including readmissions, ED visits, and mortality.
- 3) Evaluate the effect of CJR bundled payments on the use of institutional post-acute care after discharge.
- 4) Evaluate the effect of CJR on patient selection on specific attributes (e.g. age, chronic conditions, dual eligible status).

## Part C: Study Design

### C.1. Time Period Definitions

This evaluation will analyze CJR implementation as a pragmatic, government implemented cluster randomized trial comparing study endpoints for LEJR episodes in hospitals located in treatment MSAs (‘treatment’ or ‘intervention’ group) vs. those located in control MSAs (‘control’ group) in the post-implementation period vs. the pre-implementation period.

We will define the following time periods **based on admission date**:

2015												2016												2017											
1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
<b>Pre-Intervention</b>												<b>Excluded</b>			<b>Washout</b>			<b>Post-Intervention</b>									<b>Follow-Up</b>								
												↑			↑																				
												Final Rule Published			CJR Start																				

- 1) Pre-intervention: January 1, 2015- December 31, 2015 (12 months)
  - Time period used for assessing patient, hospital and MSA characteristics
  
- 2) Exclusion and “washout” periods: January 1, 2016-June 30, 2016 (6 months)
  - We will exclude January 1, 2016-March 31, 2016 (3 months) because LEJR episodes beginning in this period will cross over between CJR implementation and non-implementation periods.
  - We will also drop the first 3 months after CJR implementation (April 1, 2016 – June 30, 2016) from the evaluation as a “washout” period for hospitals to adapt to new payment model. Because hospitals were mandated to enroll in CJR, hospitals may be poorly prepared initially to respond to payment incentives. We will also examine this period in sensitivity analyses.
  
- 3) Post-intervention: July 1, 2016-September 30, 2017 (15 months)
  - This 15-month period includes the full post-washout implementation period through the most recently available data at the time of analysis
  - Data will be available through December 31, 2017 to capture 90-day post discharge episode claims
  - Any episodes whose 90-day post discharge period ends after December 31, 2017 will be excluded.
    - It is possible that late 4<sup>th</sup> quarter 2017 (October-December) data may be partially incomplete due to the time needed for CMS to finalize claims. If we observe evidence of these, we may revise the definition of the post-period to accommodate these data limitations.

### C.2. Study Population:

The entire study population will consist of all eligible LEJR episodes, as defined below, in any of 196 CJR-eligible MSAs with an admission date from January 1, 2014 until September 30, 2017 (with the exception of admissions during the excluded period defined above from January 1, 2016-June 30, 2016).

The unit of randomization is the MSA, with 75 “treatment” MSAs (9 of which “dropped out” from the intervention prior to implementation due to rule changes, see below) exposed to bundled payments beginning in April 1, 2016, and 121 “control” MSAs with no payment changes. The unit of observation/analysis will be the LEJR episode, with only one LEJR episode included per Medicare beneficiary.

Treatment and Control Group Definitions:

- Treatment vs. Control group is defined by the MSA of a hospital where a LEJR procedure takes place (regardless of beneficiary residence)
- CMS identified 196 MSAs eligible for randomization based on the following inclusion criteria:
  5. MSA had more than 400 LEJR episodes from 7/1/2013 to 6/30/2014
  6. MSA had more than 400 non-Bundled Payments for Care Improvement (BPCI) LEJR episodes from 7/1/2013 to 6/30/2014
  7. MSA had less than 50% percent of LEJR episodes included in any BPCI model
  8. MSA had >50% of LEJR episodes paid through the inpatient prospective payment system (e.g. excluding Maryland, critical access hospitals, etc.)
- Out of these 196 MSAs:
  - 75 treatment MSAs selected in July 2015 interim final rule
  - 121 MSAs in control group
  - 196 MSAs were divided into 8 strata defined by quartiles of pre-period episode spending and above vs. below median population (Table below)
  - Probability of MSA selection depended on pre-period average LEJR 90-day episode spending as shown in the Table below
- The treatment group was reduced from 75 MSAs to 67 MSAs in November 2015 due to updating BPCI enrollment lists up to October 2015 (i.e. 8 MSAs fell out of inclusion criteria #2 and #3 above due to increased BPCI enrollment)
  - Because this change disrupts initial randomization, we define the treatment group as the initially randomized 75 MSAs with an “intent-to-treat” approach

**Table: MSA Selection by Payment Quartile and Above/Below Median Population Criteria, as Defined by CMS**

	Payment in lowest quarter	Payment in 2nd lowest quarter	Payment in 3rd lowest quarter	Payment in highest quarter	Total eligible MSAs
Selection Proportion .....	30%	35%	40%	45%	.....
Less Than Median Population (Group #) .....	(1)	(2)	(3)	(4)	.....
Number Eligible MSAs per Proposed Rule (80 FR 41198) .....	33	19	22	24	98
Proportion x Number .....	9.9	6.65	8.8	10.8	
Number initially selected from group .....	10	7	9	11	37
Number finally selected from group .....	8	6	8	11	33
More Than Median Population (Group #) .....	(5)	(6)	(7)	(8)	
Number Eligible MSAs per NPRM .....	16	30	27	25	98
Proportion x Number .....	4.8	10.5	10.8	11.25	
Number initially selected from group .....	5	11	11	11	38
Number finally selected from group .....	5	10	9	10	34
Total Eligible MSAs per Proposed Rule (80 FR 41198) .....	49	49	49	49	196
Number initially selected .....	15	18	20	22	75
Number finally selected from group .....	13	16	17	21	67

Source: Reproduced from CJR Final Rule, Federal Register 80 FR 73298  
(<https://www.federalregister.gov/d/2015-29438/page-73298>)

#### Index Episode Definition:

- Episodes will be assigned to a time period defined in C.1. above by hospital admission for LEJR, which is the episode start date.
- As defined by CMS for the CJR program, LEJR episodes will be defined by discharge diagnosis related group [DRG]):
  - o DRG 469 (major joint replacement with major complications/comorbidities)
  - o DRG 470 (major joint replacement without major complications/comorbidities)
  - o The DRGs above are inclusive of hip and knee replacements, including procedures due to hip fractures.
  - o In the case of hospital transfers, the episode payment is based with the hospital discharging the patient with the above DRGs.

#### Other Episode Inclusion Criteria:

- Continuously enrolled in Medicare A+B for 12 months prior and throughout 90-day LEJR episode (Based on 2014 Medicare claims, roughly 6% of episodes are lost due to mandating 12 mo of prior coverage)
  - o We will include beneficiaries who die during the LEJR episode as long as they are continuously enrolled in Medicare A+B until the date of death

#### Episode Exclusion Criteria:

- Repeat LEJR episode for beneficiaries
  - We are excluding repeat procedures to remove the potential bias from clustering of outcomes within patient
  - If there are overlapping joint replacement episodes (e.g. a patient receives two knee replacements 80 days apart), those patients will be excluded.
- Currently eligible for Medicare due to end-stage renal disease (ESRD)

- Hospital opened after 2015 (i.e., no pre-period)
- Index hospitalization claim denied (likely wrong DRG)
- BPCI Participation
  - Hospital is an episode initiator for LEJR in BPCI Model 1, 2 or 4
  - Patient discharged to BPCI Model 3 participating PAC provider for LEJR
    - If any portion of an episode overlaps with the onset of BPCI participation for LEJR episodes, it will be excluded
    - Note: We are unable to exclude patients in physician group practices that are BPCI Model 2 episode initiators due to lack of available information (i.e. National Provider Identifiers of participating physicians and attribution rules) to accurately assign patients to these physician groups

#### Other Study Population Notes:

The CJR program excludes any LEJR episodes with any mortality from the payment model. We will not make this exclusion for two reasons. First, within-episode mortality is an important secondary outcome in our evaluation as well as part of the composite primary outcome of complications. Second, even though episodes with mortality are excluded from bundled payments their total cost and patterns of care are still a reflection of hospitals' potentially shifting approach to LEJR care due to the CJR payment model. Therefore, we will include these episodes to capture a more complete picture of patient outcomes following CJR implementation.

### **Part D: Study Endpoints and Data Sources**

#### Data Sources

We will use the following data sources for this evaluation (data source purpose in parenthesis):

7. Publicly Available CJR Enrollment Files (treatment vs. control assignment)
8. Medicare Claims Files – 2015-2017 (primary and secondary endpoints)
  - a. 100% sample available for inpatient, post-acute care and hospital outpatient files
  - b. 20% sample available for Part B “Carrier” file
9. Medicare Denominator Files – 2014-2016 (patient characteristics, 2017 unavailable at the time of analysis)
10. Medicare Provider of Services File (hospital characteristics)
11. Area Health Resource File – 2015 (MSA characteristics)
12. BPCI Participation Database (exclusion of hospitals and episodes)

#### Primary Endpoints

##### Primary Outcome #1: Facility/Institutional Spending per Episode

- We will use the full unadjusted payments (aka, spending) to providers (including Medicare, patient and primary payer portions) that include common CMS adjustments such as wage index

- Rationale: we are interested in the pragmatic national impact of the program on spending, regardless of regional and service-specific adjustments that Medicare applies
- As secondary outcomes, we will also examine standardized institutional costs and total costs (see below)
- Any denied payments will be excluded from spending outcomes
- We will examine the potential influence of extreme outlier episode payments, and potentially Winsorize episodes with extreme payments if a small number of episodes (less than 0.5% of total episodes) exert a large influence on the final results. If we do Winsorize, then the analyses will be presented with and without Winsorizing.
- The payments for any services that cross over the end of the LEJR episode (e.g. SNF admission or HHA episode) will be pro-rated based on the percentage of days occurring within the episode
  - For example, if a HHA service begins 86 days post discharge (i.e. 5 days left in the episode) and lasts 20 days, 5/20, then 25% of the HHA payment will be attributed to the LEJR episode.
- In the episode reconciliation process, CMS removes LEJR “unrelated” costs from calculated episode costs (see CMS website for [more documentation](#)). We are not applying these exclusions to capture a broader picture of total institutional spending per episode.

Primary Outcome #2: TKA/THR complications rate

- Definition: we will capture all complication events as defined in NQF Measure #1550, including hip fracture patients
- We used Measure #1550 because it is endorsed by the NQF, employed by CMS for quality reporting, and it is used by CJR to set quality thresholds for receiving bonuses
- While we will capture complications as described by this measure, we will define the sample differently. We will include all episodes for the primary analysis as defined above in Section C.2.
  - We chose a different sample, because the measure defines its sample based on the presence of ICD-9 procedure codes 81.51 (total hip replacement) and 81.54 (total knee replacement), not DRGs 469 and 470 as above.
  - The NQF measure also makes a number of exclusions to focus on elective joint replacements, mainly excluding hip fracture patients
    - As a secondary outcome, we will additionally examine complication rates using the exact sample definition in the NQF defined measure
- This measure, as defined by CMS and used in CJR, captures a composite of the following outcomes:
  - Within 7 days of admission
    - Myocardial infarction
    - Pneumonia
    - Sepsis/shock
  - Within 30 days of admission

- Surgical site bleeding
- Pulmonary embolism
- All-cause mortality
- Within 90 days of admission
  - Implant mechanical complications
  - Joint/wound infection
- May 9, 2018 Update:
  - Pre-analysis of this outcome showed diverging trends between treatment and control MSAs, making interpretation of our difference-in-differences study design more difficult
  - Given this issue, we will only report a positive effect of the program on complications if the effect we observe is greater than what we expect from the pre-period AND an outlier on permutation tests (see section F below)
  - Additional sensitivity analyses may be necessary to examine the potential effect of CJR on complication rates beyond trends present prior to program start

Primary Outcome #3: Patient risk score, estimated as predicted LEJR episode spending

- The purpose of this outcome is to capture patient selection by hospitals subsequent to CJR implementation in a single, integrated measure
- If we find clinically meaningful evidence of patient selection in this outcome, it may substantially shift our interpretation of any findings in primary outcomes #1 and #2, given that we cannot exclude that any effects observed are due to unobserved changes in patient case mix.
- Individual patient characteristics (e.g., age) will also be examined as secondary outcomes, but as a primary outcome we will estimate patients' predicted spending
- Using data from the before the full study period (2013-2014) for a 20% sample of beneficiaries, we will estimate a linear regression model with the dependent variable of predicted total episode spending
  - We will use the 20% sample to be able to include estimated Part B (“carrier”) costs, which are only available for a 20% sample.
  - Estimates from this model will be used to generate predicted spending for each episode in our study sample
- Patient covariates in the model will include:
  - Age (in 5 year bins up until 90+)
  - Gender
  - Race (white/black/other)
  - Original reason for Medicare enrollment: disability, age, ESRD
  - Medicaid status (any Medicaid enrollment in prior 12 months)
  - Prior inpatient admission in prior 12 months
  - Any institutional PAC (SNF, IRF, LTAC) admission in prior 12 months
  - 27 chronic conditions as captured in the Chronic Condition Warehouse file from prior calendar year
- If the model has adequate predictive power without the chronic illness indicators, we may forgo including them to avoid the possibility that treatment hospitals intentionally “up-code” their patients to get improving quality ratings.

- If needed to aid interpretation of this measure, we may use an alternate specification of this measure, such as classifying patients into quartiles of predicted spending or creating a risk score scaled from 0-100

## Secondary Outcomes

### *Spending and Utilization*

- 1) Total LEJR episode spending. Our data sample has only Part B claims for a 20% sample. We will include facility/institutional and Part B claims in this measure.
  - Because of data availability, this will only be calculated for the post-intervention period in 2016
- 2) LEJR institutional spending for 100% sample in which we use price standardization to remove Medicare fee adjustments
  - Because of data availability, this analysis would only include 2015 in the pre-period
- 3) 90-day post-discharge spending in mutually exclusive categories (100% sample unless otherwise specified)
  - Index hospitalization
  - Repeat inpatient, ED and observation spending in 90-days
  - Post-acute care spending: SNF, IRF, and HHA
  - Other Medicare Part B spending (20% sample)
- 4) Number of LEJR episodes per 1,000 beneficiaries
  - LEJR episodes per 1,000 beneficiaries at the MSA level
  - Denominator defined as all beneficiaries in an MSA-quarter with continuous FFS Medicare coverage for that quarter
- 5) 90-day post hospital discharge all-cause hospital readmission rates
  - Binary outcome, 1 = any inpatient admission in the 90-day window
  - Excluding certain planned admissions as defined in the CMS hospital-wide readmission measure (NQF #1789)
- 6) 90-day post hospital discharge all cause observation visit rates without admission
  - Binary outcome, 1 = any observation stay in the 90-day window without admission
- 7) 90-day post hospital discharge all cause emergency department rates without admission
  - Binary outcome, 1 = any emergency department visit in the 90-day window without admission
- 8) 90-day post hospital discharge all cause hospital visit of any kind (emergency department visit, observation stay or inpatient admission)
  - Binary outcome, 1 = any hospital visit in the 90-day window
- 9) 90-day all-cause mortality
- 10) LEJR complications rate as defined in primary outcome #2 with CMS-defined exclusion criteria (as opposed to using full study sample defined in C.2.)
- 11) Proportion of LEJR admissions discharged to home vs. different PAC settings:
  - Primary outcome defined above for institutional PAC discharge
  - Proportion discharged home
  - Proportion discharged to home health agencies

- 12) Average length of stay among those admitted to institutional PAC settings during full episode among those with at least 1 day of institution PAC use

#### *Patient Selection*

Examine the proportion of patients in the treatment and control groups in the pre vs. post-periods with the following characteristics:

- 10) Sex
- 11) Age: 65-79, 80-89, 90+
- 12) Race
- 13) Urban/rural residence
- 14) Medicaid eligibility
- 15) Disability as original reason for enrollment
- 16) Prior inpatient admission in prior 12 months
- 17) Any institutional PAC (SNF, IRF, LTAC) admission in prior 12 months

#### **Part E: Episode, Patient, Hospital, and MSA Characteristics**

We will examine the characteristics listed below for balance on observable characteristics across the treatment and control groups in the 24-month pre period. This list may be not inclusive of all characteristics examined in the final analysis. Other characteristics not specified below will be marked as post-hoc additions.

Episode characteristics:

- Diagnosis related group 469 or 470
- Presence of hip fracture
- Type of procedure (defined by ICD-9 procedure code)

Patient characteristics:

- Age (in 5 year bins up until 90+)
- Gender
- Race (white/black/other)
- Urban/rural residence
- Original reason for Medicare enrollment: disability, age (ESRD excluded)
- Medicaid status (any Medicaid enrollment in prior 12 months)
- CCW conditions (27 total, used from prior calendar year)
  - Number of conditions
  - Proportion with  $\geq 6$  or  $\geq 9$  conditions

Hospital characteristics:

- Ownership type (profit, government, non-profit)
- Size (<100, 100-249, >250 beds)
- Teaching status (non-teaching, minor, major or unknown)
- Census region (North, South, Midwest, West)
- DSH percent (in quartiles)
- Medicare accountable care organization participation
- Pre-implementation average 90-day LEJR spending

MSA characteristics (averages of county-level measures within MSA):

- Median household income
- % Medicare advantage penetration
- % penetration of BPCI LEJR episodes

- SNF total beds

## **Part F: Statistical Analysis**

We will follow an intent-to-treat (ITT) approach for all main analyses in which we include all 75 MSAs randomized to the intervention (8 of these MSAs were subsequently dropped). This is to best preserve the initial randomization of MSAs to treatment groups performed by CMS. Secondary analyses using instrumental variables approaches will be considered to calculate a treatment-on-treated (ToT) effect estimate on the primary outcomes.

### Procedures for Missing Data

Because the data for this project will primarily use administrative Medicare claims data, we anticipate little to no missing data, particularly for the primary outcomes. If missingness for a key covariate exceeds 1-2% and is deemed important for the analysis, we will use multiple imputation to account for the missing data.

### Statistical Significance Levels

The overall significance level for the study will be set at  $P < 0.05$  for the three main outcomes, which we regard as independent study questions. For the multiple secondary analyses, these tests will be considered as exploratory and we will use a threshold of  $P < 0.05$ , taking into account clinical significance for results near the threshold of significance.

### Evaluation of Randomization/Balance Across Study Arms

We will evaluate balance across the episode, patient, hospital and MSA characteristics listed in Part E between the treatment and control groups using data from the pre-implementation period. We will use chi-squared and Student's t-tests as appropriate for two-sided tests of differences in means across the study arms. However, given the large sample size of episodes we anticipate in each arm, statistical significance will be only one factor in evaluating the adequacy of balance between the groups as clinically insignificant results may demonstrate statistically significant mean differences.

#### *Accounting for Variable Treatment Assignment Probability*

As shown in Part C.2. above, MSAs had different probabilities of being assigned to the treatment group depending on their quartile of average LEJR spending from July 1, 2013 to June 30, 2014 (as assessed by CMS). Because pre-period spending will be highly correlated with post-period spending, one of our primary outcomes, this could create baseline imbalance between treatment and control groups in unadjusted comparisons. We will adjust for treatment probability stratum in the statistical models for adjudicating outcomes below, which will provide within-stratum estimates of the effect of treatment assignment.

However, for assessing randomization, these differing treatment probabilities could be misleading about the effectiveness of randomization. To account for the varying

probabilities of treatment assignment, in addition to unweighted balance tables, we will also produce weighted balance tables that weight episodes, hospitals and MSAs such that the probability of the treatment or control MSAs being selected are equal within each stratum (i.e. analogous to “direct standardization”). The weights will be chosen to match each stratum to the treatment/control probability of the entire sample. The assignment probabilities and weights used are summarized in the Table below:

Stratum	Payment Quartile	Above/Below Median Population	Initial Treatment MSAs	Control MSAs	Treatment Probability	Control Probability	Treatment Weight (Overall Prob/Stratum Prob)	Control Weight (Overall Prob/Stratum Prob)
<b>Overall</b>	-	-	<b>75</b>	<b>121</b>	<b>0.383</b>	<b>0.617</b>	-	-
1	1 (low)	Below	10	23	0.303	0.697	1.263	0.886
2	1 (low)	Above	5	11	0.313	0.688	1.224	0.898
3	2	Below	7	12	0.368	0.632	1.039	0.977
4	2	Above	11	19	0.367	0.633	1.044	0.975
5	3	Below	9	13	0.409	0.591	0.935	1.045
6	3	Above	11	16	0.407	0.593	0.939	1.042
7	4 (high)	Below	11	13	0.458	0.542	0.835	1.140
8	4 (high)	Above	11	14	0.440	0.560	0.870	1.102

#### Unadjusted and Adjusted Statistical Analyses

The main model will use a “difference-in-differences” estimator to calculate the differential change in outcomes in the treatment vs. control groups in the post-period. The advantage of this model is adding an extra layer of robustness to reduce potential bias from imbalance in pre-period characteristics or outcomes between the treatment and control groups.

Our main pre-specified analysis to report the primary outcomes will be a regression model adjusted for patient characteristics and a hospital fixed effect. However, we will also compare these results to unadjusted statistical analyses (i.e., without the  $X_i$  term in equation (1) below), and regression models that use hospital and MSA characteristics in place of the hospital fixed effect.

For the unadjusted model, to test the hypotheses stated in Part B, we will estimate linear regression model with episodes ( $i$ ) nested within hospitals ( $j$ ) nested within MSAs ( $k$ ) as follows:

$$y_{ijk} = \beta_0 + \beta_1 \text{TREAT}_k * \text{POST} + X_i \beta_2 + \eta_j + \tau_i + \varepsilon_{ijk} \quad (1)$$

In equation (1),  $y_{ijk}$  is a continuous or dichotomous outcome corresponding to one of the primary or secondary outcomes specified in Part D.  $\beta_0$  is the intercept and  $\beta_1$  is the key coefficient on the intent-to-treat differential effect of treatment assignment vs. control in the post period ( $\text{TREAT}_k * \text{POST}$ ) of an episode as the MSA level.  $\text{POST}$  is a dummy variable specifying whether an episode occurred in the post-implementation period and  $\text{TREAT}$  is an indicator of whether an episode occurred in a treatment vs. control MSA. There are no main effects for  $\text{POST}$  and  $\text{TREAT}$  included because these are incorporated into hospital and quarter fixed effects as below.

$X_i$  is a vector of episode/patient characteristics, all specified above in Part E.  $\tau_i$  is a set of quarter fixed effects and  $\eta_j$  is a set of hospital-level fixed effects. The addition of the hospital fixed effect will control for any time-invariant differences between hospitals, and necessarily, the hospital's MSA as well. Adding these effects adds statistical power to the analysis by accounting for random differences between the treatment and control groups that arise by chance in the randomization process. The randomization stratum of an MSA is not included in this model because the hospital fixed effects account for these strata. We will account for clustered standard errors at the MSA level.

#### Testing for pre-intervention differences in trend

One of the basic assumptions behind difference-in-differences models is that in the absence of the treatment effect, both treatment and control groups would have stable and parallel trends in outcomes. To test this hypothesis, we will examine unadjusted models similar to equation (1) above, where a linear term for time (in quarters) will be used instead of the POST indicator. Significant differences in the  $\beta_1$  coefficient will be used to evaluate the parallel trends assumption.

In prior work we have found that testing for statistical differences in pre-trend may not be sufficient given that even small *insignificant* differences can bias findings. The differences in pre-trend should also be assessed in the context of the size of the difference in outcomes.

If we observe any differences in pre-trends that could change our interpretation of the results we will employ a permutation-based approach for our estimates. More specifically, if the estimated impact of CJR observed in our primary and secondary outcomes are small enough that they could be reasonably impacted by differential changes in pre-intervention trends we will use permutation tests.

This involves generating 100-1,000 samples of the MSAs in our study and randomly assigning treatment and control classification based on the distribution in the **Table** in Section C.2. Using these randomly permuted combinations of treatment and control, we will generate a distribution of estimates for the effect of interest. If the effect we observe is due to CJR treatment, then the effect we observe should be in the tail, ideally in the upper/lower 2.5<sup>th</sup> percentile or less, of the distribution of effect sizes from all the permuted samples. This permutation test may also be employed if the differences in the primary outcomes are marginal in terms of statistical significance.

The concept behind this approach is that instead of discarding the difference-in-differences estimator entirely, we will instead assess how much more extreme the effect we observe is vs. random allocation of treatment and control assignment.

#### Instrumental Variables Analysis

Because not all MSAs initially assigned to participate in CJR bundled payments actually ended up in the new payment model (8 MSAs dropped due to CMS regulation changes), our estimates

of  $\beta_1$  in Equation (1) above will be biased towards the null. To ask the question of the causal impact of CJR bundled payments for those MSAs that did participate, we will use instrumental variables (IV) analysis.

We will use assignment to the treatment group as our instrument for actual bundled payment implementation. Treatment group assignment is a valid instrument because:

- Study arm was randomly assigned and thus uncorrelated with outcome
- Study arm is assumed to have an effect on outcomes only through payment policy (i.e. study arm assignment meets the “exclusion criterion” for a valid instrument)
- Study arm is highly predictive of bundled payment implementation

The causal effect of bundled payment participation on the outcomes of interest in this analysis will be modeled along the lines of Equation (1) above, with the replacement of TREAT with an indicator for true exposure to bundled payments, BUNDLE:

$$y_{ijk} = \pi_0 + \pi_1 \text{BUNDLE}_k * \text{POST} + X_i \pi_2 + \eta_j + \tau_i + \varepsilon_{ijk} \quad (2)$$

We will estimate equation (2) by two stage least squares (2SLS), using the following first stage equation:

$$\text{BUNDLE}_k = \delta_0 + \delta_1 \text{TREAT}_k + \delta_3 \text{STRATUM}_k + \mu_{ijk}$$

In which the excluded instrument is the variable TREAT with the first stage coefficient of  $\delta_1$ .

Because the model is just identified through randomizations, the 2SLS estimate of  $\pi_1$  is given by the ratio of the reduced form in Equation (1) and first stage coefficients ( $\beta_1/\delta_1$ ).

### Planned Subgroup and Sensitivity Analyses

We plan to investigate other alternative specifications of our analyses above in subgroups of interest. Subgroups and sensitivity analyses may include but are not limited to the following. Any other further analyses conducted that are not specified below will be marked as post-hoc in the final analysis:

- Subgroup Analyses
  - Elective LEJR vs. hip fracture
  - Hip vs. knee replacement
  - Hospital subgroups:
    - ACO hospital participants
    - Teaching vs. non-teaching hospitals
    - For-profit vs. non-profit hospitals
    - Large (>250 beds) vs. smaller (<250 beds) hospitals
    - Top quartile of pre-period LEJR episode volume vs. lowest quartile
    - Top quartile of pre-period average episode costs vs. lowest quartile
- Sensitivity Analyses
  - Include the initial washout period in our main analyses

- Generalized linear models with log-link function or other functional forms for spending outcomes
- Logistic regression models for binary outcomes
- Post-period treatment estimator alone without difference-in-differences if pre-period data are adequately balanced