LEWINGROUP®

CMS Comprehensive Care for Joint Replacement Model: Performance Year 4 Evaluation Report – Appendices

Fourth Annual Report

HEALTH CARE AND HUMAN SERVICES POLICY, RESEARCH, AND ANALYTICS — WITH REAL-WORLD PERSPECTIVE.

Prepared for: **Centers for Medicare & Medicaid Services**

Submitted by: **The Lewin Group, Inc. with our partners: Abt Associates, GDIT, and Telligen**

September 2021

CMS Comprehensive Care for Joint Replacement Model: Performance Year 4 Evaluation Report – Appendices

Fourth Annual Report

Prepared for:

Jessica McNeely
Research and Rapid-Cycle Evaluation Group (RREG),
Center for Medicare & Medicaid Innovation (CMMI),
Centers for Medicare & Medicaid Services (CMS)
This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2014-00033I.

Submitted by:

The Lewin Group, Inc.

September 2021

List of Appendices

Appendix A: List of Acronyms & Glossary Terms

Exhibit A-1: List of acronyms

Acronym	Meaning
ACH	Acute Care Hospital
ACO	Accountable Care Organization
ADLs	Activities of Daily Living
APM	Alternative Payment Model
ASC	Ambulatory Surgical Center
BPCI	Bundled Payments for Care Improvement
CI	Confidence Interval
CJR	Comprehensive Care for Joint Replacement
CMMI	Center for Medicare & Medicaid Innovation
CMS	Centers for Medicare & Medicaid Services
CPT	Current Procedural Terminology
CY	Calendar Year
DiD	Difference-in-Differences
DME	Durable Medical Equipment
DSH	Disproportionate Share Hospital
ED	Emergency Department
EMR	Electronic Medical Record
ESRD	End-Stage Renal Disease
FFS	Fee-for-Service
FIM	Functional Independence Measure
FY	Fiscal Year
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems
HCC	Hierarchical Condition Category
HH	Home Health
HHA	Home Health Agency
ICS	Internal Cost Savings
IP	Inpatient
IPO	Inpatient Only
IPPS	Inpatient Prospective Payment System
IRF	Inpatient Rehabilitation Facility
IRF-PAI	Inpatient Rehabilitation Facility-Patient Assessment Instrument
IT	Information Technology
LEJR	Lower Extremity Joint Replacement
LOS	Length of Stay
LTCH	Long-Term Care Hospital
MA	Medicare Advantage
MCC	Major Complication or Comorbidity

Acronym	Meaning
MDS	Minimum Data Set
MSA	Metropolitan Statistical Area
MS-DRG	Medicare Severity-Diagnosis Related Group
NPRA	Net Payment Reconciliation Amount
OASIS	Outcome and Assessment Information Set
OP	Outpatient
OT	Occupational Therapy
PAC	Post-Acute Care
PDGM	Patient-Driven Groupings Model
PDPM	Patient-Driven Payment Model
PGP	Physician Group Practice
PRO	Patient-Reported Outcomes
PSW	Propensity Score Weighting
PT	Physical Therapy
PY	Performance Year
SNF	Skilled Nursing Facility
THA	Total Hip Arthroplasty
TKA	Total Knee Arthroplasty
VBP	Value-Based Payments

Exhibit A-2: Glossary of terms

Term	Definition
90-day post-discharge period (PDP)	The 90 days following discharge from the anchor hospitalization.
Acute care hospital (ACH)	A health care facility that provides inpatient medical care and other related services for acute medical conditions or injuries.
Ambulatory surgical center (ASC)	A health care facility that provides surgical care to patients not requiring hospitalization or services exceeding 24 hours.
Anchor hospitalization	The hospitalization that triggers the start of the episode of care.
Baseline time period	The period of time that precedes the intervention period as a basis for comparison in the difference-in-differences statistical technique. The baseline period includes episodes that were initiated from 2012 to 2014 and that ended between April 1, 2012 and March 31, 2015.
Beneficiary incentive	A programmatic flexibility available to hospitals participating in the CJR model. This allows participating hospitals to offer patients certain incentives not tied to the standard provision of health care, as long as it supports a clinical goal.
Bundle	The services provided during the episode that are linked for payment purposes.
CJR collaborator	Medicare-enrolled providers and suppliers engaged in caring for CJR beneficiaries that enter into sharing agreements with a participant hospital. Collaborators may be a SNF, HHA, LTCH, IRF, physician, non-physician practitioner, provider or supplier of outpatient therapy services, PGP, non-physician provider group practice, ACO, hospital, or critical access hospital.
CJR sharing arrangement	A financial arrangement between a participant hospital and a CJR collaborator for the sole purpose of making gainsharing payments or alignment payments under the CJR model.
Effective discount percentage	The effective discount percentage serves as Medicare's portion of the savings. A 3% effective discount percentage is used to set the prospective quality-adjusted target price. The effective discount percentage used at reconciliation varies based on the hospital's quality performance in the year and whether the hospital's average episode payment falls above or below its quality-adjusted target price. For hospitals receiving reconciliation payments, the effective discount percentages are: 1.5% for "excellent" quality, 2% for "good" quality, and 3% for "acceptable" quality. (Hospitals with "below acceptable" quality are ineligible to receive reconciliation payments.) For hospitals with repayment responsibility in PY2/3, the effective discount percentages were: 0.5% for "excellent" quality, 1% for "good" quality, and 2% for "acceptable" or "below acceptable" quality. For hospitals with repayment responsibility in PY4/5, the effective discount percentages are: 1.5% for "excellent" quality, 2% for "good" quality, and 3% for "acceptable" and "below acceptable" quality.
Episode benchmark price	The episode benchmark price represents the expected episode payments if treatment patterns and patient mix did not change from historical spending for LEJR episodes. In the first three years of the model, the episode benchmark price is based on a blend of hospital-specific and regional historical LEJR payments. In PY4/5, the episode benchmark price is based solely on regional amounts. The product of the episode benchmark price and the effective discount percentage equals the quality-adjusted target price.
Episode of care	For the CJR model, an episode of care is triggered by an inpatient hospitalization for an LEJR procedure in which a beneficiary is discharged under MS-DRG 469 (major joint replacement or reattachment of lower extremity with MCC) or 470 (major joint replacement or reattachment of lower extremity without MCC) and ends 90 days after discharge from the anchor hospitalization.
Gainsharing payment	A payment from a participant hospital to a CJR collaborator made pursuant to a CJR sharing arrangement. A gainsharing payment may be composed of reconciliation payments, internal cost savings, or both.

Term	Definition
Inpatient-only (IPO) list	A list of procedures that are covered by Medicare only when provided in the inpatient setting.
Internal cost savings (ICS)	The measurable, actual, and verifiable cost savings realized by the CJR-participating hospital resulting from care redesign undertaken by the hospital in connection with providing items and services to CJR model beneficiaries. Internal cost savings does not include savings realized by any individual or entity that is not a CJR participant hospital.
Metropolitan Statistical Area (MSA)	Counties associated with a core urban area that has a population of at least 50,000.
Net Payment Reconciliation Amount (NPRA)	The aggregate quality-adjusted target price minus the total dollar amount of Medicare fee-for-service payments for items and services included in the bundle, adjusted by stop gain or stop loss limits, if applicable.
Outpatient (OP) department	A hospital-based care setting for procedures covered by Medicare through the Outpatient Prospective Payment System. The 2-midnight rule provides guidance regarding the classification of inpatient or outpatient procedures.
Post-acute care (PAC)	Rehabilitation and palliative care services received by the beneficiary from IRFs, SNFs, HHAs, or LTCHs following a hospitalization.
Post-discharge home visit waiver	A waiver available to hospitals participating in the CJR model. Under this waiver, CMS waives the direct supervision requirement for home visits so that CJR beneficiaries may receive a limited number of home visits (up to nine per episode) by licensed clinical staff paid under the Medicare Physician Fee Schedule.
Post-discharge period (PDP)	Period of time starting on the day of the anchor hospitalization discharge. For the CJR model, the post-discharge period covers the 90 days after discharge.
Post-episode care	Under the CJR model, care that occurs after the 90-day post-discharge period.
Quality-adjusted target price	The quality-adjusted target price is based on three years of historical data and is a blend of the hospital historical episode payments and the regional average historical payments in the first three years of the CJR model. In PY4/5, the target price is based completely on the regional historical episode payment. The three years of historical data is rolling across performance years (2012-2014 for years 1 and 2, 2014-2016 for years 3 and 4, 2016-2018 for year 5). The quality adjustment at the beginning of the performance year assumes that the hospital's composite quality score falls in the "acceptable" range. The quality adjustment reflects the hospital's actual composite quality score at reconciliation. There are separate quality-adjusted target prices to account for MS-DRG and hip fracture status.
Reconciliation payment	A retrospective payment that Medicare makes to a CJR participant hospital if total fee-for-service payments for its episodes during a performance year are less than the aggregate quality-adjusted target price. If total fee-for-service payments for a CJR participant hospital's episodes are more than its aggregate quality-adjusted target price, the hospital repays the difference to Medicare in PY2-5.
Related items and services	Episode-related items and services paid under Medicare Part A or Part B, after exclusions are applied, that are included in the bundle. These include physicians' services; inpatient hospital services (including readmissions with certain exceptions discussed in the Final Rule); inpatient psychiatric facility services; LTCH services; IRF services; SNF services; HHA services; hospital outpatient services; outpatient therapy services; clinical laboratory services; DME; Part B drugs; and hospice.
Risk adjustment	A statistical process to adjust claims-based outcomes and ADL measures to take into account differences at the patient, episode, hospital, state, and MSA level that are related to the measures of interest. Without adequate risk adjustment, providers treating a sicker or more service-intensive patient mix would have worse outcomes than otherwise comparable providers serving healthier patients.

Term	Definition
Stop-loss/Stop-gain limits	Adjustments included in the NPRA calculation that vary by performance year. The stop-loss limit is the maximum amount a hospital will have to repay to CMS, and the stop-gain limit is the maximum amount that a hospital will receive from CMS as a reconciliation payment. They are based on a percentage of the quality-adjusted target price. The stop-loss limits are 5% in PY2, 10% in PY3, and 20% in PY4 and PY5. The stop-gain limits are 5% in PY1 and PY2, 10% in PY3, and 20% in PY4 and PY5.
Telehealth waiver	A waiver available to hospitals participating in the CJR model. Under this waiver, CMS allows Medicare coverage of telehealth services furnished to eligible beneficiaries regardless of their geographic region. Further, the originating site requirement is waived for eligible beneficiaries receiving telehealth services from their homes or places of residence.
Three-day hospital stay waiver	A waiver available to hospitals participating in the CJR model. Under this waiver, CMS waives the three-day hospital stay requirement for Part A skilled nursing facility coverage.

Appendix B: CJR Programmatic Flexibilities, Including Financial Arrangements, Beneficiary Incentives, and Program Rule Waivers

The CJR model allows hospitals to use fraud and abuse waivers issued by the Department of Health and Human Services to facilitate the implementation of care redesign interventions. Participating hospitals may or may not elect to use these waivers. Under the CJR model, hospitals may enter into financial arrangements with CJR collaborators, collaboration agents, downstream collaboration agents or provide incentives to CJR beneficiaries. Additionally, CMS waives certain Medicare program rules for beneficiaries in CJR episodes, such as: the direct supervision requirement for post-discharge home visits, specific requirements for furnishing telehealth services, and the three-day hospital stay requirement for coverage of skilled nursing facility (SNF) care. These waivers allow CJR beneficiaries to receive services under circumstances that would not otherwise be covered by Medicare.

The waivers allowed under the CJR model include:

- **Financial Arrangements –** Under the CJR model, hospitals may enter into sharing arrangements with certain collaborating providers and suppliers that are engaged in care redesign with the hospital and that furnish services to the beneficiary during an episode. Under such a sharing arrangement, hospitals may pass on a portion of their reconciliation payment, internal cost savings, or both (i.e., a gainsharing payment) to collaborating providers and suppliers. Sharing arrangements may also permit payments from a CJR collaborator to a participant hospital (i.e., an alignment payment) when the participating hospital has to repay CMS. Collaborators may be a SNF, home health agency (HHA), long-term care hospital (LTCH), inpatient rehabilitation facility (IRF), comprehensive outpatient rehabilitation facility (CORF), therapist in private practice, physician, non-physician practitioner, provider or supplier of outpatient therapy services, physician group practice (PGP), non-physician provider group practice (NPPGP), therapy group practice (TGP), accountable care organization (ACO), hospital, or critical access hospital. Under the CJR model, gainsharing payments must be made according to a pre-specified methodology.

  To be eligible to receive a gainsharing payment, collaborators must meet quality criteria for the performance year for which the participant hospital accrued the internal cost savings or earned the reconciliation payment that comprises the gainsharing payment. The quality of care criteria must be established by the participant hospital and directly related to the CJR episode. A CJR collaborator other than an ACO, PGP, NPPGP, or TGP must have directly furnished a billable item or service to a CJR beneficiary during a CJR episode that occurred in the same performance year for which the participant hospital accrued the internal cost savings or earned the reconciliation payment that comprises the gainsharing payment or was assessed a repayment amount. A CJR collaborator that is a PGP, NPPGP, or TGP must have billed for an item or service that was rendered by one or more PGP member, NPPGP member, or TGP member respectively to a CJR beneficiary

during a CJR episode that occurred during the same performance year for which the participant hospital accrued the internal cost savings or earned the reconciliation payment that comprises the gainsharing payment or was assessed a repayment amount and must have contributed to CJR activities and been clinically involved in the care of CJR beneficiaries during the same performance year for which the CJR participant hospital accrued the internal cost savings or earned the reconciliation payment that comprises the gainsharing payment or was assessed a repayment amount. A CJR collaborator that is an ACO must have had an ACO provider/supplier that directly furnished, or an ACO participant that billed for, an item or service that was rendered to a CJR beneficiary during a CJR episode that occurred during the same performance year for which the participant hospital accrued the internal cost savings or earned the reconciliation payment that comprises the gainsharing payment or was assessed a repayment amount and the ACO must have contributed to CJR activities and been clinically involved in the care of CJR beneficiaries during the same performance year for which the participant hospital accrued the internal cost savings or earned the reconciliation payment that comprises the gainsharing payment or was assessed the repayment amount. In the event that a hospital is due to make a repayment to CMS under the CJR model, the total amount of alignment payments received by the hospital from a CJR collaborator that is an ACO may not be greater than 50% of the amount the hospital owes CMS. With respect to a CJR collaborator other than an ACO, the total amount of alignment payments received by the hospital may not be greater than 25% percent of the amount the hospital owes CMS. CMS also requires that gainsharing agreements cannot incentivize CJR collaborators to reduce service or provide substandard care to Medicare beneficiaries.

- **Beneficiary Incentives –** Participating hospitals may provide certain in-kind items or services to CJR beneficiaries during an episode of care. The item or service must be reasonably connected to a beneficiary's medical care and either be preventive or advance a clinical goal. Incentives may include technology items, which can be used for telehealth visits.
- **Post-Discharge Home Visit Waiver –** The direct supervision requirement for home visits can be waived so that CJR beneficiaries may receive a limited number of home visits (up to nine post-discharge home visits per episode) by licensed clinical staff paid under the Medicare Physician Fee Schedule.
- **Telehealth Waiver –** Under the CJR model, geographic and originating site requirements that typically apply for Medicare coverage of telehealth services may be waived as long as services are furnished according to other coverage and payment criteria. Medicare coverage criteria typically require telehealth services be furnished to individuals in certain geographic areas, including rural, medically underserved areas. For the CJR model, CMS waived this provision, allowing Medicare coverage of telehealth services furnished to eligible beneficiaries regardless of their geographic region. Medicare coverage criteria also specify that Medicare may only cover telehealth services that are received in certain

clinical settings. For the CJR model, the originating site requirement is waived for eligible beneficiaries receiving telehealth services from their homes or places of residence.

- **Waiver of SNF 3-Day Rule –** Under traditional Medicare fee-for-service (FFS) rules, beneficiaries are not eligible for Medicare-covered SNF care unless they have a prior inpatient hospital stay of at least three consecutive days within 30 days of SNF admission. Under the SNF 3-day waiver, CJR participant hospitals can discharge a CJR beneficiary to an approved SNF without a qualifying 3-day inpatient stay when medically appropriate. This waiver became available in performance year 2 of the CJR model. A provision of this waiver is CJR beneficiaries may only be discharged to a SNF that is approved at the time of the beneficiary's admission. An approved SNF is one that received three or more stars on CMS' Five-Star Quality Rating System[1] for at least seven out of the past twelve months. CMS maintains a list of approved SNFs based on these requirements on the CJR model web site, which is updated quarterly.[2]

[1] www.medicare.gov/NursingHomeCompare/
[2] https://innovation.cms.gov/innovation-models/cjr

Appendix C: Methodology

I. Data Sources

A. Secondary data sources

Secondary data sources were used to:

1) Identify and characterize CJR participant hospitals and control group hospitals for risk adjustment and creation of weights for mandatory CJR hospitals and matched control groups for hospitals in voluntary metropolitan statistical areas (MSA) (Provider of Services file, Acute IPPS Final Rule data files, Medicare FFS claims, CJR programmatic data, and Bundled Payments for Care Improvement Salesforce Database);
2) Sample providers associated with CJR participant hospitals for participation in telephone interviews and surveys (CJR programmatic data and Medicare FFS claims);
3) Identify lower extremity joint replacement (LEJR) discharges, create LEJR episodes, characterize episodes and beneficiaries, and evaluate changes in LEJR discharge volume (Medicare FFS claims, Medicare FFS beneficiary enrollment data, Master Data Management (MDM), and Bundled Payments for Care Improvement Salesforce Database);
4) Generate payment, utilization, quality, and functional status and pain outcomes and savings to Medicare (Medicare FFS claims, Medicare standardized payments, and CJR programmatic data; and
5) Evaluate hospital performance in the CJR model as demonstrated by the amounts of reconciliation payments and repayments (CJR programmatic data, Medicare FFS claims, Medicare FFS beneficiary enrollment data, Provider of Services file, Acute IPPS final rule data files, and Bundled Payments for Care Improvement Salesforce Database).

Exhibit C-1 lists the secondary sources, their contents, purpose in this evaluation, and relevant date ranges used for this report.

Exhibit C-1: Secondary data sources

Data source	Date range	Dataset contents	Use
Area Health Resource Files (AHRF)	2015-2016 (Data is from 2012-2014)	County-level data aggregated to the MSA level. Variables include Medicare Advantage penetration, average Medicare beneficiary hierarchical condition category (HCC) score, dual eligible percentage, population per square mile, geography, and supply of health care facilities (SNF beds, LTCH beds) and health care professionals (primary care physicians, orthopedic surgeons, NPs/PAs, specialists).	Used to control for MSA Medicare Advantage penetration in the patient survey analysis.

Data source	Date range	Dataset contents	Use
Bundled Payments for Care Improvement Participant Database	Baseline and intervention	Identifies health care providers (hospitals, PAC providers, physicians, and physician practice groups) that are participating in the Bundled Payments for Care Improvement initiative, the time period of participation, and the models and episodes for which they are participating.	Used to identify LEJR discharges that are assigned to Bundled Payments for Care Improvement participants for exclusion. Used to identify hospitals as past Bundled Payments for Care Improvement LEJR participants for risk adjustment, creation of propensity score weights (PSW), and creation of matched control groups for hospitals in voluntary MSAs. Used to create a measure of Bundled Payments for Care Improvement dose for the volume analysis.
Bundled Payments for Care Improvement Advanced Participant Database	Intervention	Identifies health care providers (hospitals, physicians, and physician practice groups) that are participating in the Bundled Payments for Care Improvement Advanced initiative, the time period of participation, and the episodes for which they are participating.	Used to identify LEJR discharges in the control group that are assigned to Bundled Payments for Care Improvement Advanced participants for risk adjustment.
CJR programmatic data	Intervention	List of CJR participant hospitals, as well as their PY1, PY2, PY3 and PY4 quality-adjusted target prices, reconciliation (net payment reconciliation amount or NPRA), and hospital quality data.	Used to identify CJR participating hospitals, hospitals that continued mandatory participation in PY3, their start and end dates in the CJR model, their quality performance, and their reconciliation payments or repayment responsibility. Used total reconciliation payments and repayments to CMS to calculate savings to Medicare. Used average NPRA per episode to identify factors related to hospitals receiving reconciliation payments.
FY Acute IPPS Final Rule data files	FY 2016 (Data is from FY 2012-2014)	On an annual basis, CMS sets acute care hospital IPPS payment rates. Data files include fiscal year hospital-level information on provider identification number, bed count, medical residents per 1,000 beds, average daily census, DSH patient percentage, UCP per claim, Medicare days as a percent of total inpatient days, and section 401 status.	Used to risk adjust for acute care IPPS hospital characteristics. Used in the creation of PSW and matched control groups for hospitals in voluntary MSAs. Used to identify section 401 hospitals located in control group hospitals to exclude from the mandatory analysis.
MDM	Baseline and Intervention	Provider- and beneficiary- level information on participation in CMS Innovation Center payment demonstration programs. Includes beneficiary ID, program ID, and start and end dates of participation.	Used to identify beneficiaries involved in Pioneer, Next Generation, and Medicare Shared Savings ACO programs and control for their participation in our analyses. Used to apply the ACO exclusion for episodes starting on or after July 1, 2017 (MSSP track 3, CEC with downside risk, and Next Generation).

Data source	Date range	Dataset contents	Use
Medicare FFS beneficiary enrollment data	Baseline and Intervention	Enrollment data (from CME and MBSF) provide beneficiary Medicare Part A/B eligibility information.	Enrollment data were used to confirm beneficiary eligibility and provide beneficiary characteristics for analyses (e.g., risk adjustment models, LEJR volume analysis, creation of PSW and matched control groups for hospitals in voluntary MSAs). Enrollment data were used to measure the change in case-mix of CJR and control group patients between the baseline and the intervention periods.
Medicare FFS claims	Baseline and Intervention	Parts A and B claims data (from TAP files) provide claims for different services received during the anchor hospitalization and post-discharge period (e.g., dates and types of service). A minimum three month claims run out was used for episodes included in this report.	Claims were used to: 1) create the CJR episodes, describe service use, and create risk adjustment (e.g., Medicare beneficiary HCC score) and outcome variables (e.g., unplanned readmissions, emergency department visits, and number of days/visits in each PAC setting); 2) create PSW for hospitals in mandatory MSAs and matched control groups for hospitals in voluntary MSAs; 3) generate the number of LEJR discharges for the volume analysis; 4) identify outpatient TKA procedures in CJR and control markets for descriptive analyses and create outpatient TKA episodes; and 5) sample participants for primary data collection (patient survey, telephone interviews, and orthopedic surgeon survey).
Medicare IRF-PAI data	Baseline and Intervention	The IRF-PAI is a comprehensive assessment instrument administered by nursing staff to all Medicare beneficiaries when they are admitted to an IRF and at discharge (for stays longer than three days). The IRF-PAI collects information on patients' demographics, comorbidities, living arrangements, skin conditions, and functional, cognitive, respiratory, bladder, bowel, and swallowing status. A minimum six month run out of IRF-PAI data was used for episodes included in this report.	IRF-PAI data were used to measure the percent of patients who were admitted to an IRF within five days of discharge from the anchor hospitalization and improved in functional status (mobility) by the time they were discharged from the IRF. IRF-PAI data were also used to measure the change in case-mix of CJR patients and patients in the control group who were discharged from the hospital to an IRF, between the baseline and the intervention periods.

Data source	Date range	Dataset contents	Use
MDS 3.0 data	Baseline and Intervention	The MDS is a comprehensive assessment instrument administered by nursing staff to all Medicare beneficiaries when they are admitted to a Medicare-certified SNF, at discharge, as well as on days five, 14, 30, 60, 90, and quarterly, thereafter. The MDS collects information on patients' demographics, history and diagnoses, skin conditions, medications, care management, restraint use, preferences for routine and activities, and functional, sensory, cognitive, neuro/emotional, bladder, bowel, swallowing/nutritional, and pain status. A minimum six month run out of MDS data was used for episodes included in this report.	MDS data were used to measure the percent of patients who were admitted to a SNF within five days of discharge from the anchor hospitalization and improved in functional status (toilet use and transfer, locomotion, and walking in the corridor) by the time they were discharged from the SNF. Patients without self-reported moderate to severe pain was also measured. MDS data were also used to identify patients who were in a SNF or long-term nursing facility during the six months preceding the episode, and to measure the change in case-mix of CJR patients and patients in the control group who were discharged from the hospital to a SNF, between the baseline and the intervention periods.
Medicare OASIS data	Baseline and Intervention	The OASIS is a comprehensive assessment instrument administered by nursing staff to all Medicare beneficiaries at the initiation of home health care, at resumption of care following a hospitalization, and when the patient is discharged from home health care. The OASIS collects information on patients' demographics, history and diagnoses, living arrangements, skin conditions, medications, care management, therapy needs, use of emergent care, and functional, sensory, cognitive, neuro/emotional, respiratory, cardiac, bladder, bowel, and pain status. A minimum six month run out of OASIS data was used for episodes included in this report.	OASIS data were used to measure the percent of patients who started home health care within 14 days of discharge from the anchor hospitalization and improved in functional status (ambulation/ locomotion, bed transferring, and pain when moving around) by the time they were discharged from home health care. OASIS data were also used to measure the change in case-mix of CJR patients and patients in the control group who were discharged from the hospital to home health care, between the baseline and the intervention periods.
Medicare standardized payments	Baseline and Intervention	Medicare standardized payments for 100% of Part A and B claims received via the IDR. Produced by a CMS contractor.	Used to create Medicare standardized paid amounts (Part A and B) and allowed standardized payment amounts, including beneficiary out-of-pocket amounts. Used to estimate the impacts of the CJR model on total episode and service-level payments.
POS file	December 2016	Information on Medicare-approved facilities, including provider identification number, ownership status, size, medical school affiliation, and staffing.	Used to identify and characterize acute care hospitals actively engaged in Medicare for risk adjustment and creation of PSW and matched control groups for hospitals in voluntary MSAs.

Note: ACO = accountable care organization, AHRF = Area Health Resource Files, CEC = comprehensive ESRD care model, CME = common Medicare enrollment, CMS = Centers for Medicare & Medicaid Services, DSH = disproportionate share hospital, ESRD = end-stage renal disease, FFS = fee-for-service, FY = fiscal year, HCC = hierarchical condition category, IDR = integrated data repository, IPPS = Inpatient Prospective Payment System, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument, LEJR = lower extremity joint replacement,

MBSF = Medicare beneficiary summary file, MDM = Master Data Management, MDS = Medicare Minimum Data Set 3.0, MSA = metropolitan statistical area, MSSP = Medicare Shared Savings Programs, NPRA = net payment reconciliation amount, OASIS = Outcome and Assessment Information Set, PA = physician assistant, PAC = post-acute care, POS = provider of services, PPS = prospective payment system, PSW = propensity score weight, PY = performance year, SNF = skilled nursing facility, TAP = monthly Medicare claims file, TKA = total knee arthroplasty, TIN = tax identification number, UCP = uncompensated care payment.

B. Primary data sources

We collected and analyzed primary data from the orthopedic surgeon survey and telephone interviews to inform questions that are not readily answered by secondary data. We received survey responses from 249 orthopedic surgeons and conducted two rounds of telephone interviews with 72 providers. In this appendix we describe the methods employed during the fourth performance year. Prior primary data collection efforts are detailed in prior annual reports.[1, 2, 3]

1. Provider telephone interviews

In PY4, we conducted two rounds of telephone interviews, including one round with outpatient physical therapists (PTs) and one round with administrators from skilled nursing facilities (SNFs). The aim of these interviews was to better understand relationships with hospitals participating in the CJR model and how the CJR model has influenced the care LEJR patients receive. We conducted a 30-minute semi-structured telephone interview with 32 outpatient PTs who treated patients who had LEJR surgery at mandatory CJR hospitals. The aims of this data collection effort were to better understand the outpatient PT perspective on the CJR model and to identify variations in care practices for LEJR patients. The interview focused on five key domains including: early needs, care processes, coordination, patient recovery, and awareness of CJR and other CMS models. We conducted a 45-minute semi-structured telephone interview with Executive Directors and Directors of Nursing at 40 SNFs. The aim was to better understand how hospital participation in the CJR model influences SNFs that provide care to LEJR patients.

a. Interviewees

With the PT interviews, the team interviewed outpatient PTs working in CJR mandatory MSAs and billing Medicare exclusively for LEJR patients. With the SNF interviews, the team interviewed executive and nursing directors from SNFs providing care to Medicare fee-for-service (FFS) beneficiaries who had LEJR surgery.

b. Protocols

We developed semi-structured interview guides tailored to the topic of each round of telephone interviews. Interviewees were asked to complete a brief (6-8 question) web-based survey prior to the interview. The PT pre-interview survey asked eight closed-ended questions to gather

1 Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement model - first annual report appendices. https://innovation.cms.gov/files/reports/cjr-firstannrptapp.pdf. 2018: C3-C10.

2 Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement model - second annual report appendices. https://downloads.cms.gov/files/cmmi/cjr-secondannrpt-app.pdf. 2019: E5-E10.

3 Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement model - third annual report appendices. https://innovation.cms.gov/data-and-reports/2020/cjr-thirdannrpt-app.pdf. 2020: E5-E11.

descriptive information about the therapist and clinic where they work including: years in practice, the number of PTs and PT assistants at that clinic, the clinic ownership, the hospitals that refer to the clinic, and estimated volume of Medicare FFS LEJR patients the PT sees in a typical month. The SNF pre-interview survey asked six questions to gather descriptive information about the SNF including: estimated monthly volume of LEJR patients, participation in hospitals' preferred provider networks, health system affiliation, ownership by a multi-facility organization, and awareness of other Medicare initiatives (i.e. BPCI and ACOs). We used responses from the pre-interview surveys to tailor the interview protocols.

c. *Interviewee selection criteria*

For the PT interviews, we wanted to interview PTs who have substantial experience providing outpatient therapy to LEJR patients. We initially selected PTs who submitted claims for ten or more new patient physical therapy evaluations for Medicare FFS patients with LEJR between October 2017 and September 2018. When recruitment proved challenging, we expanded the sample to include PTs who submitted claims for eight or more new patient physical therapy evaluations in this timeframe.

Medicare claims showed that in CJR mandatory MSAs, 59% of new patient physical therapy evaluations were billed by PTs and 41% were billed by physicians or physicians' assistants. This is consistent with Medicare billing policy that allows providers who are co-located with PTs to bill for physical therapy services. We included both PTs who submitted claims for their own services and also PTs whose services were billed by other providers. To identify PTs whose claims were submitted through a physician national provider identifier (NPI), we identified the organization where they work using the Tax identification number (TIN) associated with these claims. We wanted to interview a total of 40 PTs. Based on experience with past rounds of telephone interviews, we anticipated a 30% participation rate.[4] Thus, we began by identifying a sample of 120 potential interviewees. We selected a random sample of 72 PTs (60% of 120) who bill Medicare directly for their services, and 48 (40% of 120) physician NPIs associated with physical therapy evaluations, to mirror the organizational structures suggested by claims data. When the participation rate did not reach 30%, we added 140 PTs to the sample.

For the SNF interviews, we used episode files and Medicare Part A institutional claims to identify SNF claims in the 90-day post-discharge period for LEJR episodes that ended between October 1, 2017 and September 30, 2018. SNF claims were identified using the SNF Centers for Medicare & Medicaid Services (CMS) certification number (CCN). We selected SNFs that had at least 10 patients in CJR episodes who were discharged directly from the hospital to the SNF between April 1, 2018 and March 31, 2019. We excluded episodes when discharge from the hospital was not directly to a SNF – for example, episodes during which a patient went to a SNF after a period at home or after a hospital readmission. There were 265 SNFs that met these criteria, including 220 with patients from CJR mandatory hospitals and 45 with patients from opt-in CJR hospitals in voluntary MSAs. Given the number of SNFs that met our inclusion criteria, further sampling was

[4] Note: Our ultimate participation rate was 12%. See section on "Interviewee recruitment" for details.

not required. We called the SNFs in this sample and asked for the names and contact information for their executive and nursing leaders. We obtained contact information for the Executive Director and Director of Nursing at 254 of the 265 SNFs.

d. Interviewee recruitment

For both rounds of telephone interviews, the NPI or CCN was linked to Master Data Management (MDM) data to obtain provider name, mailing/practice address, and telephone number. For PTs whose claims were submitted through another provider, we contacted the organization corresponding to the TIN, identified a rehabilitation manager or physical therapy director, and asked that person to recommend a PT who met the inclusion criteria and also provide that PT's email address.

We called the PT practices and SNFs in this sample and asked for the names and contact information for the desired interviewees (i.e., PTs or executive and nursing leaders at SNFs). We then contacted potential interviewees via email and included key information and related materials (i.e., frequently asked questions document and informed consent information). We encountered some challenges in recruiting PTs, who were less familiar with the CJR model and our evaluation efforts. We began attaching a letter from CMS providing information on the CJR model and attesting to our legitimacy as evaluators. We recruited both PT and SNF interviewees on a rolling basis. We successfully interviewed 32 (12%) of the 260 sampled PTs and 40 (30%) of the 133 contacted SNFs.

2. Orthopedic surgeon survey

Surgeons with direct experience in the CJR model are in the best position to provide information about whether or how CJR participant hospitals influence surgeons performing LEJRs. In formative work leading up to this survey, we determined that the most valuable information provided by surgeons would be why and how they engage with hospitals in response to the CJR model, and how this engagement has changed care practices and interactions with total knee arthroplasty (TKA) and total hip arthroplasty (THA) patients.

a. Survey sample

We used episode files and Medicare Part A institutional claims to identify the universe of surgeons listed as the operating surgeon for CJR episodes ending over a one-year period between 10/1/2017 and 9/30/2018 (n=5,462). This universe contained surgeon and hospital identifiers (surgeon NPIs and hospital CCNs); surgeon names, addresses, and other geographic and contact information; information about surgeons' practices (TINs); and information about surgeries, including the date of the first qualifying episode associated with each surgeon since the start of the CJR model, as well as counts of the number of surgeries (total LEJR, TKA, elective THA, and THA with a

fracture) performed by each surgeon. There were 4,261 unique surgeons when we aggregated the file to the NPI level; of which, 2,338 performed any LEJR surgeries at mandatory CJR hospitals.[5]

The survey instrument included questions about experiences "in the past three years" referring to the elapsed time since the CJR model began in April 2016 to the survey fielding in 2019 to capture changes during the CJR model. For this reason, we excluded from the sample surgeons with zero episodes during the first year of the CJR model (n=31). The sample was limited to surgeons with episodes in both the file identifying the universe of surgeons listed on claims in the defined one-year period, as well as those who have been treating CJR episodes since the first year of the model. Surgeons who performed fewer than 11 LEJR surgeries for episodes ending between October 1, 2017 and September 30, 2018 were excluded as they were considered low volume based on the episode volume distribution of the sample (n=1,441). The final sample included 866 surgeons.

The sample file included each surgeon's full name, business mailing address, telephone number, and practice location including zip code, and fax number. We supplemented this information by purchasing more reliable contact information for physicians, including email addresses, from the commercial vendor IQVIA.[6,7] IQVIA provided email addresses for 760 (88%) of the sampled surgeons, as well as information on gender, specialty, and employment

b. Survey domains

The thirty-six question survey instrument (Appendix M) was developed in collaboration with CMS to address seven domains: before surgery, intraoperative/in-hospital care, post-surgery care, longer-term outcomes, hospital performance monitoring and gainsharing, outpatient knee replacement, and information about respondents. Exhibit C-2 provides information about the survey domains and topics.

[5] When aggregating the file, we summed surgery counts across all affiliated hospitals for each surgeon. Over three-quarters (77%) of the 4,261 surgeons in the aggregated file performed surgeries at only one hospital (n=3,275). Among those who performed surgeries at multiple hospitals, 51 surgeons performed surgeries at both mandatory CJR mandatory hospitals and opt-in CJR hospitals in voluntary MSAs; these surgeons were excluded.

[6] Contact information from CMS data is limited. A 2013 OIG report concluded that provider data were inaccurate in 48 to 58 percent of records. IQVIA (formerly known as IMS Health and SK&A) uses several methods to update their records on an ongoing basis, including telephone calls to practice managers, Web research, pharmacy prescribing data, data from multiple sources, and automated data cleaning algorithms.

[7] Department of Health and Human Services (HHS), Office of Inspector General (OIG). Improvements Needed to Ensure Provider Enumeration And Medicare Enrollment Data Are Accurate, Complete, And Consistent. May 2013. Available at: https://oig.hhs.gov/oei/reports/oei-07-09-00440.pdf

Exhibit C-2: CJR orthopedic surgeon survey domains and topics of interest

Domain	Topics
Before Surgery	▪ Hospital guidance regarding patient selection, modifiable health risk factors, and whether this guidance is new or changed since the CJR model began ▪ Hospital guidance regarding patient selection, environmental risk factors, and whether this is new or changed since the CJR model began ▪ Referrals to services to help with modifiable risk factors ▪ Postponement of surgery to address environmental and modifiable health risk factors ▪ Importance of hospital influence on surgeon decision making regarding patient eligibility for surgery
Intraoperative Care	▪ Implementation of internal cost-savings strategies ▪ Hospital role in promoting these strategies
Post-Surgery Care	▪ Use of institutional post-acute care versus discharge to home ▪ Hospital role in influencing discharge destination ▪ Short-term recovery – range of motion, pain, and swelling
Longer-term Outcomes	▪ Surgeon perceptions of changes in patient mobility and joint pain under the CJR model
Hospital Performance Monitoring and Gainsharing	▪ Hospitals sharing performance metrics with surgeons ▪ Surgeon participation and interest in gainsharing
Outpatient Knee Replacement	▪ Whether surgeons perform knee replacement surgeries on an outpatient basis ▪ Hospital guidance and influence regarding the decision about inpatient versus outpatient surgery
About Respondents	▪ Tenure as surgeon ▪ Employer type ▪ Participation in other APMs ▪ Number of affiliated hospitals in which surgeon performs hip or knee replacement surgeries

Note: APM = alternative payment model.

II. Study Population

This section defines the CJR and control group populations, explains the weights used in the mandatory analyses to account for differences in sampling probabilities and creation of matched control groups for hospitals in voluntary MSAs, and outlines the additional eligibility criteria for hospitals and episodes.

A. Defining the CJR and control group populations

CMS selected MSAs eligible for CJR participation based on a stratified random sampling methodology in which MSAs were stratified into eight strata based on historical wage-adjusted episode payments and population size. Within each stratum, MSAs were randomly selected to participate in the CJR model (n=67 MSAs). This design allowed for a control group of hospitals in MSAs that were eligible but not selected by CMS to participate in the CJR model (n=104 MSAs). These MSAs represent what would have happened in CJR-type markets if the model was never implemented (i.e., the counterfactual).

In January 2018, CMS reduced the mandatory participation by about half by allowing all CJR hospitals in the 33 low-payment MSAs and CJR hospitals in the 34 high-payment MSAs that were designated as rural or low-volume a one-time opportunity to remain in the model. The 67 original CJR MSAs were ranked by average historical wage-adjusted episode payment and the top 34 MSAs with the highest payments were required to continue participation in the model (mandatory MSAs), while hospitals in the bottom 33 MSAs were given a one-time opportunity to opt-in (voluntary MSAs). This report covers the first four performance years of the model from April 1st 2016 to December 31, 2019. Our analysis primarily focused on episodes from hospitals that were mandated to participate in PY4 (mandatory analysis). This analysis excluded rural and low-volume hospitals in the mandatory MSAs that were allowed to opt-in to continue participation in CJR. Low-volume hospitals had less than 20 episodes over a three-year historical period (2012 to 2014) and rural hospitals were identified using the FY 2019 IPPS data (section 401 hospitals). In this report, we also present results for the hospitals in the 33 voluntary MSAs that opted to continue participation in PY3 (opt-in hospitals) and those that did not (non-opt-in hospitals).

Exhibit C-3 shows the names and core-based statistical area (CBSA) identification numbers of the CJR and control group MSAs included in the mandatory and voluntary analyses. The MSAs included in the mandatory analysis are starred, while the voluntary MSAs are unstarred. Section II.B provides additional detail about how the control group MSAs were identified and the weights generated for mandatory analyses. Section II.C provides additional detail about how the matched control groups were identified for the voluntary analyses.

Exhibit C-3: CJR and control group MSAs included in the mandatory and voluntary analyses

CJR		Control	
CBSA ID	MSA name, state	CBSA ID	MSA name, state
10420	Akron, OH*	10180	Abilene, TX*
10740	Albuquerque, NM	10580	Albany-Schenectady-Troy, NY
11700	Asheville, NC*	10900	Allentown-Bethlehem-Easton, PA-NJ*
12020	Athens-Clarke County, GA	11100	Amarillo, TX*
12420	Austin-Round Rock, TX*	11260	Anchorage, AK
13140	Beaumont-Port Arthur, TX*	12060	Atlanta-Sandy Springs-Roswell, GA
13900	Bismarck, ND	12700	Barnstable Town, MA*
14500	Boulder, CO	13460	Bend-Redmond, OR
15380	Buffalo-Cheektowaga-Niagara Falls, NY	13820	Birmingham-Hoover, AL*
16020	Cape Girardeau, MO-IL	14260	Boise City, ID
16180	Carson City, NV	14460	Boston-Cambridge-Newton, MA-NH
16740	Charlotte-Concord-Gastonia, NC-SC	14540	Bowling Green, KY*
17140	Cincinnati, OH-KY-IN*	15940	Canton-Massillon, OH
17860	Columbia, MO	15980	Cape Coral-Fort Myers, FL*
18580	Corpus Christi, TX*	16060	Carbondale-Marion, IL*
19500	Decatur, IL	16300	Cedar Rapids, IA
19740	Denver-Aurora-Lakewood, CO	16620	Charleston, WV
20020	Dothan, AL*	16700	Charleston-North Charleston, SC
20500	Durham-Chapel Hill, NC	16860	Chattanooga, TN-GA*
22420	Flint, MI	16980	Chicago-Naperville-Elgin, IL-IN-WI*
22500	Florence, SC*	17020	Chico, CA
23540	Gainesville, FL*	17780	College Station-Bryan, TX
23580	Gainesville, GA	17900	Columbia, SC*
24780	Greenville, NC*	17980	Columbus, GA-AL
25420	Harrisburg-Carlisle, PA*	18140	Columbus, OH
26300	Hot Springs, AR*	19100	Dallas-Fort Worth-Arlington, TX*
26900	Indianapolis-Carmel-Anderson, IN	19380	Dayton, OH*
28140	Kansas City, MO-KS	19660	Deltona-Daytona Beach-Ormond Beach, FL*
28660	Killeen-Temple, TX*	19820	Detroit-Warren-Dearborn, MI*
30700	Lincoln, NE	20260	Duluth, MN-WI
31080	Los Angeles-Long Beach-Anaheim, CA*	20740	Eau Claire, WI
31180	Lubbock, TX*	22020	Fargo, ND-MN
31540	Madison, WI	22520	Florence-Muscle Shoals, AL*
32820	Memphis, TN-MS-AR*	22900	Fort Smith, AR-OK
33100	Miami-Fort Lauderdale-West Palm Beach, FL*	23060	Fort Wayne, IN
33340	Milwaukee-Waukesha-West Allis, WI	23420	Fresno, CA
33700	Modesto, CA	24340	Grand Rapids-Wyoming, MI
33740	Monroe, LA*	24580	Green Bay, WI

CJR		Control	
CBSA ID	MSA name, state	CBSA ID	MSA name, state
33860	Montgomery, AL*	24860	Greenville-Anderson-Mauldin, SC*
34940	Naples-Immokalee-Marco Island, FL	25060	Gulfport-Biloxi-Pascagoula, MS*
34980	Nashville-Davidson--Murfreesboro--Franklin, TN	25540	Hartford-West Hartford-East Hartford, CT
35300	New Haven-Milford, CT*	25620	Hattiesburg, MS*
35380	New Orleans-Metairie, LA*	25940	Hilton Head Island-Bluffton-Beaufort, SC*
35620	New York-Newark-Jersey City, NY-NJ-PA*	26140	Homosassa Springs, FL*
35980	Norwich-New London, CT	26420	Houston-The Woodlands-Sugar Land, TX*
36260	Ogden-Clearfield, UT	26580	Huntington-Ashland, WV-KY-OH
36420	Oklahoma City, OK*	26620	Huntsville, AL*
36740	Orlando-Kissimmee-Sanford, FL*	26980	Iowa City, IA
37860	Pensacola-Ferry Pass-Brent, FL*	27140	Jackson, MS*
38300	Pittsburgh, PA*	27860	Jonesboro, AR*
38940	Port St. Lucie, FL*	27900	Joplin, MO
38900	Portland-Vancouver-Hillsboro, OR-WA	29180	Lafayette, LA*
39340	Provo-Orem, UT*	29200	Lafayette-West Lafayette, IN
39740	Reading, PA*	29340	Lake Charles, LA*
40980	Saginaw, MI	29420	Lake Havasu City-Kingman, AZ
41860	San Francisco-Oakland-Hayward, CA	29460	Lakeland-Winter Haven, FL*
42660	Seattle-Tacoma-Bellevue, WA	29620	Lansing-East Lansing, MI
42680	Sebastian-Vero Beach, FL*	30460	Lexington-Fayette, KY*
43780	South Bend-Mishawaka, IN-MI	30620	Lima, OH*
41180	St. Louis, MO-IL	30780	Little Rock-North Little Rock-Conway, AR
44420	Staunton-Waynesboro, VA	31140	Louisville/Jefferson County, KY-IN*
45300	Tampa-St. Petersburg-Clearwater, FL*	31420	Macon, GA*
45780	Toledo, OH*	31700	Manchester-Nashua, NH
45820	Topeka, KS	33460	Minneapolis-St. Paul-Bloomington, MN-WI
46220	Tuscaloosa, AL*	34820	Myrtle Beach-Conway-North Myrtle Beach, SC-NC
46340	Tyler, TX*	34900	Napa, CA
48620	Wichita, KS	35840	North Port-Sarasota-Bradenton, FL*
		36100	Ocala, FL
		36540	Omaha-Council Bluffs, NE-IA
		37900	Peoria, IL
		37980	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD*
		38060	Phoenix-Mesa-Scottsdale, AZ
		38860	Portland-South Portland, ME
		39300	Providence-Warwick, RI-MA
		39460	Punta Gorda, FL*
		39580	Raleigh, NC
		40140	Riverside-San Bernardino-Ontario, CA*
		40220	Roanoke, VA

CJR		Control	
CBSA ID	MSA name, state	CBSA ID	MSA name, state
		40340	Rochester, MN
		40380	Rochester, NY
		40900	Sacramento--Roseville--Arden-Arcade, CA
		41500	Salinas, CA
		41620	Salt Lake City, UT*
		41740	San Diego-Carlsbad, CA
		41940	San Jose-Sunnyvale-Santa Clara, CA
		41980	San Juan-Carolina-Caguas, PR
		42200	Santa Maria-Santa Barbara, CA
		42220	Santa Rosa, CA
		42340	Savannah, GA
		43340	Shreveport-Bossier City, LA*
		43620	Sioux Falls, SD
		44060	Spokane-Spokane Valley, WA
		44100	Springfield, IL
		44180	Springfield, MO
		41100	St. George, UT
		46060	Tucson, AZ
		46140	Tulsa, OK
		46520	Urban Honolulu, HI
		47940	Waterloo-Cedar Falls, IA*
		48300	Wenatchee, WA
		48900	Wilmington, NC
		49340	Worcester, MA-CT*
		49620	York-Hanover, PA*
		49660	Youngstown-Warren-Boardman, OH-PA*

Source: https://innovation.cms.gov/initiatives/CJR. Information for control group MSAs provided by CMS.

Notes: An asterisk indicates that the MSA was included in the mandatory analysis. MSAs without an asterisk were included in the voluntary opt-in and non-opt-in analyses.
CBSA = core-based statistical area, MSA = metropolitan statistical area.

B. Creation of the analytic weights for the mandatory analysis

1. Average treatment effect of the treated (ATT)

For the original design of the model, the probability of an MSA being selected to participate in the CJR model varied across the strata, with CMS proportionally under-sampling MSAs in the lower average episode payment strata (stratum 1, 2, 5, and 6) and over-sampling MSAs in higher average episode payment strata (stratum 3, 4, 7, and 8). Exhibit C-4 shows the count of CJR and control group MSAs by stratum and the proportion of MSAs in each stratum that make up the CJR and control groups.

Exhibit C-4: CMS' original stratified random sample of CJR MSAs

MSA population	MSA sampling stratum	MSA average episode payment	# MSAs eligible for sampling	CJR sample		Control group sample	
				# CJR MSAs	Proportion of MSAs selected for CJR	# Control group MSAs	Proportion of MSAs in the control group
Less than median population	1	Lowest quartile	25	8	32.0%	17	68.0%
	2	2nd lowest quartile	18	6	33.3%	12	66.7%
	3	3rd lowest quartile	19	8	42.1%	11	57.9%
	4	Highest quartile	22	11	50.0%	11	50.0%
More than median population	5	Lowest quartile	15	5	33.3%	10	66.7%
	6	2nd lowest quartile	28	10	35.7%	18	64.3%
	7	3rd lowest quartile	22	9	40.9%	13	59.1%
	8	Highest quartile	22	10	45.5%	12	54.5%
		Total	171	67		104	

Source: CJR evaluation team analysis of the Medicare Program Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services; A Final Rule by the Centers for Medicare & Medicaid Services, 80 FR 73273 (November 24, 2015) (codified at 42 CFR 510).

Note: MSA = metropolitan statistical area.

We used an ATT analysis to evaluate the impact of CJR on mandatory hospitals. For this analysis, we constructed the control group using the following steps:

Step 1. We began with the 104 non-CJR MSAs.

Step 2. We identified and excluded low-volume and rural hospitals from the 104 non-CJR MSAs because these hospitals were excluded from mandatory participation in the CJR group.

Step 3. We applied MSA-level weights to the 104 non-CJR MSAs based on the exact probability that the MSA was selected into the 34 mandatory CJR MSAs through the two-step selection process.

To construct the weights in Step 3, we first calculated **the probabilities of the first-stage selection** for each MSA, i.e., the probability that the MSA was randomly selected to be in the original set of 67 CJR MSAs. These probabilities equaled the proportion of MSAs randomly selected for CJR from each MSA sampling stratum.

Next, we calculated **the probabilities of the second stage selection**, i.e., the probability that the MSA was selected into the 34 mandatory CJR MSAs given that it was selected in the first stage. Those second stage selection probabilities were more complex to calculate because the MSAs for the 34 mandatory CJR MSAs were not selected randomly and so we could not rely on simple proportions.[8] We therefore calculated exact probabilities using combinatorics. We used the exact

[8] They were selected by ranking the original 67 CJR MSAs by historical average episode payment and retaining the top half of the sample (i.e., retaining the 34 MSAs with the highest historical average episode payment).

probabilities to construct MSA-level weights such that the weighted control group was representative of the CJR group. Specifically,

- *Weight for 'mandatory CJR hospitals' = 1*
- *Weight for control group hospitals* =
$$\frac{(probability\ in\ treatment)}{(probability\ in\ control)} = \frac{(probability\ in\ 34\ mandatory\ CJR\ MSAs)}{(probability\ in\ 104\ non\ CJR\ MSAs)}$$

  Note: These are MSA stratum-level weights so all control group hospitals in the same MSA will have the same weight.

We compared the exact probabilities with simulated probabilities that we produced by simulating the two-stage selection process 1,000 times, summing the number of times each MSA was selected into the 34 mandatory CJR MSAs, and dividing the sum by 1,000. The exact probabilities from the combinatorics-based solution and the simulated probabilities are the same (rounded to the 10^{th} of a percent).

Exhibit C-5 shows the analytic weights calculated for control group MSAs included in the mandatory analysis.

Exhibit C-5: Analytic weights for control group MSAs included in the mandatory analysis

MSA sampling stratum	MSA	Weight
4	All MSAs	1.00
8	All MSAs	0.83
7	Birmingham-Hoover, AL	0.69
7	Cape Coral-Fort Myers, FL	0.69
7	Chattanooga, TN-GA	0.68
7	Columbia, SC	0.11
7	Greenville-Anderson-Mauldin, SC	0.69
3	Gulfport-Biloxi-Pascagoula, MS	0.73
3	Hattiesburg, MS	0.73
3	Huntsville, AL	0.71
3	Jonesboro, AR	0.73
7	Lexington-Fayette, KY	0.69
3	Lima, OH	0.73
7	Louisville/Jefferson County, KY-IN	0.51
3	Macon, GA	0.73
3	Manchester-Nashua, NH	0.00
7	North Port-Sarasota-Bradenton, FL	0.69
3	Ocala, FL	0.67
7	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.56
3	Punta Gorda, FL	0.73
7	Riverside-San Bernardino-Ontario, CA	0.02
7	Salt Lake City, UT	0.05
3	Waterloo-Cedar Falls, IA	0.73
3	Wilmington, NC	0.00
7	Worcester, MA-CT	0.69
7	York-Hanover, PA	0.69

Source: CJR evaluation team analysis of the Medicare Program; Cancellation of Advancing Care Coordination Through Episode Payment and Cardiac Rehabilitation Incentive Payment Models; Changes to Comprehensive Care for Joint Replacement Payment Model: Extreme and Uncontrollable Circumstances Policy for the Comprehensive Care for Joint Replacement Payment Model; A Final Rule by CMS, 82 FR 57066 (December 1, 2017) (codified at 42 CFR 510 and 42 CFR 512).

Note: MSA = metropolitan statistical area.

2. Propensity score weight

Next, we adjusted the ATT weights to account for CJR participant hospitals shifting a lower share of TKAs to the hospital outpatient setting. We included outpatient TKAs in the control group and further adjusted the weights on these outpatient TKAs to create balance with the CJR group.

TKA was removed from the inpatient only list in January 2018. As a result, Medicare pays for TKAs performed in the hospital outpatient department; however, the CJR model only includes

inpatient LEJRs as episodes. Following the policy change, both mandatory CJR and control group hospitals began performing TKAs in the outpatient setting, however mandatory CJR hospitals shifted fewer TKAs to the outpatient setting. Our analyses indicated that a portion of the CJR inpatient TKAs would have been outpatient in the absence of the CJR model. As a result of this differential response to the outpatient TKA policy, an appropriate counterfactual for the CJR episodes would need to include patients who would have received their TKA in the inpatient setting if they had been treated in a CJR hospital, but instead received their TKA in the outpatient setting because they were treated at a control group hospital.

To construct an appropriate counterfactual, we employed the propensity score weighting (PSW) method and included *all* control outpatient TKAs in the DiD model, weighted by the hypothetical probability of an outpatient TKA being inpatient if the hospital had been participating in the CJR model. The probability weights were constructed to ensure that the weighted sum of all control group outpatient TKAs balances the CJR inpatient TKAs predicted to have been inpatient TKAs in the absence of the CJR model. Outpatient TKAs were not included in the CJR group.

A logit regression was used to model the probability that a TKA in the CJR or control groups would be performed in the inpatient or outpatient setting. This logit included CJR status, hospital TKA volume in 2017, hospital average length of stay for TKAs in 2017, and all other risk adjustment variables included in our difference-in-differences (DiD) models. The coefficient on CJR status predicts the proportion of CJR inpatient TKAs that were inpatient due to the CJR model. A second logit model was run on the CJR-treated inpatient and outpatient TKAs to predict TKA setting (inpatient or outpatient). Then, the estimated coefficients from that model were used to predict the probability of a control outpatient TKA being inpatient had the episode been performed at a CJR hospital. The final weight for each control group TKA episode was:

$$w * \left(\hat{p} * \frac{N}{\sum \hat{p}}\right)$$

where:

- w is the original sampling weight for the hospital at which the outpatient TKA was performed.
- $\hat{p}$ is the estimated probability that a control TKA would have been inpatient had it been performed at a CJR hospital.
- N is the number of control group outpatient TKAs needed to correct the imbalance in outpatient TKA shares between CJR and control groups.
- $\sum \hat{p}$ is the sum of all the predicted probabilities for control group TKAs.

C. Creation of the matched control groups for the voluntary analyses

CJR hospitals located in the 33 voluntary MSAs were given a one-time opportunity in January 2018 to opt to continue participation in the CJR model for PY3 through PY5. We classify these hospitals into two groups: "opt-in CJR hospitals" are hospitals that opted to continue their participation, and "non-opt-in CJR hospitals" are hospitals that did not opt-in and thus their participation ended as of January 1, 2018.

To account for this selection, we constructed a subset group of matched control hospitals to use as a counterfactual when evaluating the impact of the CJR model on each CJR hospital group. We first took all hospitals located in control MSAs (MSAs eligible but not selected to participate in the CJR model) from sampling strata that also had CJR voluntary MSAs. More specifically, all control MSAs in sampling strata 1, 2, 3, 5, 6, and 7 were included; sampling strata 4 and 8 were used in the mandatory analysis only (Exhibit C-4). Second, we selected specific hospitals located in these control MSAs that resembled the voluntary CJR hospitals on a variety of baseline characteristics. This was performed by separate one-to-one nearest neighbor hospital-level propensity score matching without replacement for opt-in CJR hospitals and non-opt-in CJR hospitals. Each propensity score matching procedure used a logistic regression to estimate propensity scores and included 36 hospital-level covariates calculated using data from our baseline period.[9] This created a group of matched control hospitals for each of the two groups, specifically 74 opt-in control hospitals and 200 non-opt-in control hospitals, to be used in separate corresponding DiD and descriptive analyses. Because we created matched control groups for the opt-in and non-opt-in hospitals, we did not need to use analytic weights in these analyses. Additional details pertaining to each matched control group is presented in the following subsections.

1. Opt-In matched control group balance assessment

For the 74 opt-in CJR hospitals we selected 74 matched control group hospitals to serve as a counterfactual in our analyses. Comparing this matched control group to the opt-in CJR group, all but one of the variables used in the propensity score matching procedure had standardized mean

[9] The 36 hospital-level covariates included: indicators for hospital ownership, number of hospital beds, total TKA episode volume, total THA episode volume, total LEJR MS-DRG 469 episode volume, total LEJR MS-DRG 470 episode volume, indicator for participation in BPCI LEJR, percent of total LEJR volume that was in BPCI, indicators for Census Division, average HCC score, average age, percent of LEJR patients in age categories (20-64, 65-79, 80+), percent of LEJR patients that were female, percent of LEJR patients in race/ethnicity categories, percent of LEJR patients eligible for Medicaid, percent of LEJR patients with disability excluding ESRD, percent of LEJR patients flagged with obesity, percent of LEJR patients flagged with hypertension, percent of LEJR patients flagged with dementia, and percent of LEJR patients with prior care use six months prior to anchor hospitalization (ACH stay, IRF stay, SNF stay, HH use, any prior care).

differences of less than 0.2.[10,11] The distributions of propensity scores between the matched control group and the opt-in CJR group exhibited common support and appeared similar.[12]

2. Non-Opt-In matched control group balance assessment

For the non-opt-in CJR hospitals we used a caliper in our matching procedure to ensure that the distribution of propensity scores of the non-opt-in CJR hospitals and the matched control hospitals exhibited common support and appeared similar.[13] When matching each non-opt-in CJR hospital with one control hospital, the resulting match had to be within a selected absolute difference (i.e., not exceed a specified threshold) of log-odds propensity score between the two hospitals. The caliper was based on the standard deviation of the estimated log-odds propensity score and assessed among various thresholds to determine the optimal value. We employed a 0.05 caliper, which excluded some non-opt-in CJR hospitals from all our analyses using the matched control group.

For the remaining 200 non-opt-in CJR hospitals, we selected 200 matched control group hospitals to serve as a counterfactual in our analyses.[14] With the caliper, all matching variables had standardized mean differences within +/- 0.2.

3. Overlap of voluntary and mandatory control groups

The propensity score matching procedures were performed separately and independently for each of the two groups of CJR hospitals in voluntary MSAs (opt-in and non-opt-in). As a result, the matching procedures considered the same set of potential control group hospitals and were permitted to choose the same individual hospitals. This methodological choice was made based on conceptual factors and assessment of empirical evidence of the quality of the matched control groups. Of the hospitals chosen in the two matched control groups, 31 control hospitals were included in both groups.

Moreover, since the analytic weights used for the analysis of mandatory CJR hospitals included control MSAs from strata 3 and 7, the matching procedures also considered some control hospitals that were included in the mandatory control group. This methodological choice was made to account for these MSA strata not having a certain chance of being hypothetically selected as a

[10] The indicator for Census South Atlantic Division had a standardized mean difference of 0.24 between the opt-in CJR and the matched control group. This was driven by there being 9 opt-in CJR hospitals and only 4 matched control hospitals. Given the similarities in all other matching variables, we do not think this slight geographical imbalance is of concern.

[11] Stuart, E.A. (2010). Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1), 1.

[12] The distributions of the log odds of the propensity score between the opt-in CJR hospitals and the matched control hospitals resulted in failing to reject the null hypothesis of the Kolmogorov–Smirnov test that the distributions were equal (p=0.65).

[13] When using the caliper, the distributions of the log odds of the propensity score between the non-opt-in CJR hospitals and the matched control hospitals resulted in failing to reject the null of the Kolmogorov–Smirnov test that the distributions were equal (p=0.46).

[14] Two non-opt-in CJR hospitals did not have LEJR episode volume in the baseline and thus were also excluded from our analyses.

"mandatory" MSA. Thus, 18 control hospitals chosen in the matched control group for opt-in CJR hospitals and 32 control hospitals chosen in the matched control group for the non-opt-in CJR hospitals are included in the mandatory control group with nonzero analytic weights.

D. Additional eligibility criteria for hospitals and episodes

1. Hospital criteria

For inclusion in the analysis, hospitals had to be acute care hospitals (ACH) paid under the IPPS that performed LEJR for Medicare beneficiaries in the baseline or intervention periods.

2. Episode definition

For both the CJR and control group populations, the beginning of an episode is triggered by an admission to a CJR participating or control group hospital (called an anchor hospitalization) with a resulting discharge in Medicare Severity-Diagnosis Related Group (MS-DRG) 469 or 470 (LEJR with major complications or comorbidities and LEJR without major complications or comorbidities, respectively). The end of the episode is 90 days after the anchor hospital discharge.

Medicare beneficiaries who met and maintained the following eligibility throughout the period were included in the analysis:

- enrolled in Medicare Parts A and B;
- Medicare was the primary payer (i.e., not enrolled in any managed care plan or covered under other health plans); and
- not eligible for Medicare based on end-stage renal disease (ESRD).

As specified in the Final Rule, episodes were cancelled in the CJR model and excluded from the analysis if:

- the patient no longer met the eligibility criteria described in the preceding paragraph;
- the patient was readmitted to a participating hospital during the episode and discharged under MS-DRG 469 or 470 (in which case the first episode is canceled and a new CJR episode begins);
- the patient died at any time during the episode period;
- the episodes started on or after July 1, 2017 and were prospectively assigned to a Next Generation ACO, a Medicare Shared Savings Program ACO track 3, or a Comprehensive ESRD Care Model ACO with downside risk;[15] or

[15] This additional exclusion criterion was added with the January 2017 Final Rule, Advancing Care Coordination Through Episode Payment Models (EPMs); Cardiac Rehabilitation Incentive Payment Model; and Changes to the Comprehensive Care for Joint Replacement Model (CJR). Available at: https://www.federalregister.gov/documents/2017/01/03/2016-30746/medicare-program-advancing-care-coordination-through-episode-payment-models-epms-cardiac

- the episodes were attributed to the Bundled Payments for Care Improvement initiative.[16]

To estimate the all-cause mortality rate measure, we retained episodes that were canceled due to death of patient, but otherwise met all other eligibility criteria.

We also excluded episodes that lacked certain beneficiary information used to risk-adjust outcomes (age, sex, and six months of Medicare FFS enrollment history prior to the LEJR hospital admission).

We also created outpatient TKA episodes for inclusion in the control group, as described above in Section II.B.2. Beginning in the January 2018, CMS removed TKA from the inpatient only list, allowing Medicare coverage for TKAs provided in the hospital outpatient setting. Evidence suggests that the CJR model influences the choice of inpatient or outpatient setting, which would bias impact estimates that are based only on inpatient LEJR episodes that are included under the CJR model. (Annual Report 3 includes additional information about outpatient TKA and the CJR model.[17]) Therefore, we also include outpatient TKA episodes in the control group and apply a weight based on their probability of being an inpatient TKA in the absence of CJR to obtain impact estimates of the CJR model. For the outpatient TKAs, the beginning of the episode was triggered by a TKA performed in the outpatient department of a CJR participating or control group hospital (CPT code 27447 assigned to C-APC 5115 with status indicator "J1" in Part B institutional claims). The end of the episode is 90 days after the outpatient procedure and beneficiaries had to meet and maintain the CJR eligibility criteria throughout the episode to be included in the analysis.

[16] Episodes initiated at CJR participant hospitals could be attributed to a physician group practice (PGP) participating in the Bundled Payments for Care Improvement initiative or to skilled nursing facilities, inpatient rehabilitation facilities, long-term care hospitals or home health agencies participating in the Bundled Payments for Care Improvement Initiative Model 3.

[17] Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement model - third annual report. https://innovation.cms.gov/data-and-reports/2020/cjr-thirdannrpt. 2020: 31-37.

III. Impact of the CJR Model on Claims and Assessment-based Outcomes

A. Measures of impact on payments, utilization, and quality

In this section we present the episode-level outcome measures that were constructed to assess the impact of the CJR model on Medicare payments, utilization, and quality during the first CJR performance year. Exhibit C-6 and C-7 list each claims-based and assessment-based measure respectively.

Exhibit C-6: Claims-based payment, utilization, and quality measures

Measure category	Measure name/description
Medicare payments[a]	Total Medicare standardized allowed amounts included in the episode, inpatient anchor hospitalization through the 90-day PDP
	Medicare standardized allowed amounts included in the inpatient anchor hospitalization
	Medicare standardized allowed amounts per episode, by service, 90-day PDP[b]
	Medicare standardized allowed amounts, 30-Day PEP[c]
Utilization	First post-acute discharge was to IRF
	First post-acute discharge was to SNF
	First post-acute discharge was to HHA
	First post-acute discharge was home without HHA
	Any HHA visits, 90-day PDP
	Number of IRF days, 90-day PDP[d]
	Number of SNF days, 90-day PDP[d]
	Number of HHA visits, 90-day PDP[d]
	Number of HHA PT/OT visits, 90-day PDP[d]
	Number of PT/OT visits[d]
Quality	Unplanned readmission, 90-day PDP
	Emergency department visit, 90-day PDP
	All-cause mortality, inpatient stay and 90-day PDP[e]
	Incidence of any complications, 90-day PDP[f]

Source: All measures are constructed from Medicare fee-for-service claims data.

Notes: HHA = home health agency, IRF = inpatient rehabilitation facility, OT = occupational therapy, PAC = post-acute care, PDP = post-discharge period, PEP = post-episode period, PT = physical therapy, SNF = skilled nursing facility.

[a] Payments are the standardized Medicare allowed amounts. Standardization removes wage adjustments and other Medicare payment adjustments. Allowed amounts include beneficiary cost sharing.

[b] Services include inpatient readmissions, IRF, SNF, HHA, and services covered under Medicare Part B.

[c] Services include all health care services covered under Medicare Part A and Part B.

[d] The eligible sample for PAC days and visits is among those with any use.

[e] Under the CJR model, death during the anchor hospitalization or 90-day PDP cancels the episode. Therefore, to estimate the all-cause mortality rate, this analysis includes CJR and control group episodes as well as beneficiary admissions at CJR and control group hospitals that would have been identified as episodes if the beneficiaries had not died during the anchor hospitalization or 90-day PDP.

[f] THA/TKA complications is measured among elective episodes only.

Exhibit C-7: Assessment-based functional status and pain measures

First PAC setting	Outcome name
IRF	Average change in mobility score
SNF	Improved transfer, locomotion on unit, and walking in corridor
	Improved toilet use
	Without self-reported pain
HHA	Improved ambulation/ locomotion
	Improved bed transferring
	Reduced pain

Source: IRF measures are constructed from PAI data, SNF measures are constructed from MDS data, and HHA measures are constructed from OASIS data.

Note: HHA = home health agency, IRF = inpatient rehabilitation facility, MDS = minimum data set, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, PAI = patient assessment instrument, SNF = skilled nursing facility.

B. Measures of unintended consequences

Our evaluation of unintended consequences of the CJR model focused on changes in patient mix. Exhibit C-8 lists the patient characteristics from claims and enrollment data that we monitored. While the impact analysis on payment, utilization, and quality controlled for changes in these patient characteristics, we also monitored changes in these characteristics separately to directly examine changes in patient mix.

Exhibit C-8: Measures of patient mix

Type of unintended consequence	Measure name/description
Changes in patient mix	Age
	Sex
	Race/ethnicity
	Medicaid eligibility
	Disability, no ESRD
	Congestive heart failure
	HCC score
	Dementia
	Obesity
	Hypertension
	Diabetes
	Prior utilization (in the six months prior to the anchor hospitalization) ▪ Inpatient ACH stay ▪ IRF stay ▪ SNF stay ▪ Home health use ▪ Any prior care[a]

Source: Patient mix measures are constructed from Medicare fee-for-service claims and beneficiary enrollment data.

Notes: ACH = acute care hosptial, ESRD = end-stage renal disease, HCC = hierarchical condition category, IRF = inpatient rehabilitation facility, SNF = skilled nursing facility.

[a] Any prior care includes inpatient hospital, psychiatric hospital, emergency department visits, skilled nursing facility, inpatient rehabilitation facility, home health, long-term care hospital, and hospice during the six months prior to anchor hospitalization.

C. Analytic methodology

While the CJR and control group populations are overall quite similar in terms of market, hospital, and patient characteristics,[18] there may be unobserved differences that impact outcomes. To control for both observed and unobserved differences and to isolate the impact of the CJR model on outcomes, we used a DiD regression approach supplemented by risk adjustment.

1. DiD estimator

The DiD approach quantifies the impact of the CJR model by comparing changes in outcomes between the baseline and intervention periods for the CJR population and the control group population. One of the main advantages of this approach is that it can successfully isolate the effect of unobserved characteristics of treatment and control groups that are time invariant.[19]

a. Baseline period

The baseline period for our evaluation encompasses episodes that started between January 1, 2012 and December 31, 2014 and ended between April 1, 2012 and March 31, 2015.

b. Intervention period

The intervention period for this Annual Report follows the definition of the first through fourth performance years in the Final Rule: episodes starting on or after April 1, 2016 and ending by December 31, 2019.[20]

The DiD model uses an outcome measure, Y, and estimates the differential change in Y for beneficiaries receiving care from CJR participant hospitals between the baseline and the intervention periods relative to that same change for beneficiaries receiving care from hospitals in the control group.

To illustrate the DiD approach, we define:

- $Y_{i,k,t}$ is the outcome for the i^{th} episode with an LEJR at hospital k in period t ($t = 1$ during the CJR intervention quarters and zero otherwise)
- $CJR_{i,k}$ is an indicator that takes the value of 1 if the i^{th} episode was initiated by a CJR participant hospital k and takes the value of 0 otherwise
- $X_{i,k,t}$ are hospital, geographic, and patient characteristics in period t
- $E[Y|t, CJR, X]$ is the expected value of outcome measure Y conditional on values of t, CJR, and X

[18] CMS. Comprehensive Care for Joint Replacement Model - Third Annual Report. Last updated 14 January 2021. Available at: https://innovation.cms.gov/data-and-reports/2020/cjr-thirdannrpt

[19] While the DiD model controls for unobserved heterogeneity that is fixed over time, it does not control for unobserved heterogeneity that varies over time.

[20] CMS. Medicare Program; Comprehensive Care for Joint Replacement Payment Model for Acute Care Hospitals Furnishing Lower Extremity Joint Replacement Services: final rule (42 CFR Part 510). *Fed Regist*. 2015; 80(226): 73273-73554.

The DiD estimator is:

$$DiD = [E(Y \mid t=1, CJR = 1, X) - (E(Y \mid t=0, CJR = 1, X)] - [E(Y \mid t=1, CJR = 0, X) - (E(Y \mid t=0, CJR = 0, X)] \quad \textbf{(1)}$$

To illustrate the calculation of the DiD, consider the linear model listed below:

$$Y_{i,k,t} = b_0 + b_1 \cdot t + b_2 \cdot CJR_{i,k} + b_3 \cdot CJR_{i,k} \cdot t + X_{i,k,t}' \cdot B + u_{i,k,t} \quad \textbf{(2)}$$

- The value of coefficient b_1 captures aggregate factors that could cause changes in outcome *Y* in the intervention period relative to the baseline period that are common across CJR and control group episodes.
- Coefficient b_2 captures the relative differences in outcomes between CJR and control group episodes.
- Coefficient b_3 determines the differential in outcome *Y* experienced by beneficiaries receiving services from CJR hospitals during the CJR intervention period relative to control group episodes in the intervention period, and represents the DiD estimator.
- The vector of coefficients *B* measures the differential effects of risk factors (X) on the outcome variable.

To calculate separate DiDs for each of the four performance years during the intervention period, Equation 2 was modified to include four-time period indicators t_1 (equals 1 during PY1 intervention period and zero otherwise), t_2 (equals 1 during PY2 intervention period and zero otherwise), t_3 (equals 1 during PY3 intervention period and zero otherwise), and t_4 (equals 1 during PY4 intervention period and zero otherwise).

$$Y_{i,k,t} = b_0 + b_1 \cdot t_1 + b_2 \cdot t_2 + b_3 \cdot t_3 + b_4 \cdot t_4 + b_5 \cdot CJR_{i,k} + b_6 \cdot CJR_{i,k} \cdot t_1 + b_7 \cdot CJR_{i,k} \cdot t_2 + b_8 \cdot CJR_{i,k} \cdot t_3 + b_9 \cdot CJR_{i,k} \cdot t_4 + X_{i,k,t}' \cdot B + u_{i,k,t} \quad \textbf{(3)}$$

- Coefficient b_6 determines the differential in outcome Y experienced by beneficiaries receiving services from CJR providers during the CJR PY1 intervention period relative to control group episodes in the PY1_intervention period, and represents the DiD estimator for PY1.
- Coefficient b_7, b_8, and b_9 represent the DiD estimators for PY2, PY3, and PY4 respectively.

Finally, to calculate the DiD estimate for outcome measures that were risk-adjusted with non-linear models, we used the regression model's coefficient estimates to calculate each of the four conditional expectations that make up the DiD estimator in Equation 1. In these cases, the standard errors were computed using the Delta method.[21] For all DiD models, statistical significance was assessed at the 10% level.

[21] The delta method expands a function of a random variable about its mean, usually with a Taylor approximation, and then takes the variance. Specifically, if Y= f(x) is any function of a random variable X, we need only calculate the variance of X and the first derivative of the function to approximate the variance of Y. Let μ_x be the mean of X and

This approach was used for mandatory and voluntary analyses. For the mandatory analysis, we applied the analytic weight described above in Section II.B. We used the matched control group, described in Section II.C, for the voluntary analysis, which did not require an analytic weight.

c. Assumptions of DiD estimators

One critical assumption of an unbiased DiD estimate is that the treatment and control group outcomes follow parallel trends for the outcome of interest during the baseline period. Another assumption is that these parallel trends would have remained the same in the period when the policy is actually implemented in the absence of the policy intervention. While the first assumption can be tested if sufficient baseline data on the CJR and control groups are available, the second assumption is untestable.

We evaluated the parallel trends assumption two ways: visually inspected trends for all outcomes; and statistically tested that the CJR and control group outcomes follow parallel trends during the baseline period. We estimated episode-level models for each outcome using baseline data and used both linear and joint F-tests of equality to conclude whether there is evidence to reject the parallel trend assumption. We considered outcomes to fail parallel trends if for both tests we reject the null hypothesis of seemingly parallel trends at the 10% significance level.

For the joint F-test, we report the p-value of an F-test that tests if the differential between the CJR and control group are jointly equal across discrete four-quarter time periods. We included dummy variables for each of the three baseline years; interaction terms between the CJR group indicator and each of the year dummies, along with all the risk-adjustment variables that we include in the DiD models (described in the Section III.C.2).

The joint F-test model is:

$$Y_{i,k,t} = b_0 + b_1 \cdot Year_{1,i} + b_2 \cdot Year_{2,i} + b_3 \cdot Year_{3,i} + b_4 \cdot Year_{1,i} \cdot CJR_k + b_5 \cdot Year_{2,i} \cdot CJR_k + b_6 \cdot Year_{3,i} \cdot CJR_k + X_{i,k}{}' \cdot B + u_{i,k,t}$$

where:

- $Y_{i,k,t}$ is the outcome for the i^{th} episode with an LEJR at hospital k in the baseline period in year t.
- $Year_{i,t}$ is an indicator that takes the value of 1 if the i^{th} episode was initiated during year t of the baseline period and takes the value of 0 otherwise
- $CJR_{i,k}$ is an indicator that takes the value of 1 if the i^{th} episode was initiated by a CJR participant hospital k and takes the value of 0 otherwise
- $X_{i,k}$ are hospital, geographic, and patient characteristics in the baseline period

f'(x) be the first derivative, a Taylor expansion of Y = f(x) about μ_x gives the approximation: Y = f(x) ≈ f(μ_x) + f'(μ_x)(x − μx). Taking the variance of both sides yields: Var(Y) = Var(f(X)) ≈ [f'(μ_x)]^{2}Var(X). For example, suppose Y = X^2. Then f(x) = X^2 and f'(x) = 2x, so that Var(Y) ≈ $(2\mu_x)^2$ Var(X).

and the test is:

$$H_0 : b_4 = b_5 = b_6$$
$$H_1 : b_4 \neq b_5, \text{ or } b_4 \neq b_6, \text{ or } b_5 \neq b_6$$

For the linear test, we report the p-value of a linear slope coefficient of the quarterly difference between the CJR and control group. We included a quarterly indicator; interaction term between the CJR group indicator and the quarterly indicator, along with all the risk-adjustment variables that we include in the DiD models.

The linear test model is:

$$Y_{i,k,t} = b_0 + b_1 \cdot Quarter_{i,t} + b_2 \cdot CJR_k + b_3 \cdot Quarter_{i,t} \cdot CJR_k + X_{i,k}' \cdot B + u_{i,k,t}$$

where:

- $Y_{i,k,t}$ is the outcome for the i^{th} episode with an LEJR at hospital k in the baseline period in quarter t.
- $Quarter_{i,t}$ is an indicator that takes the value of 1 if the i^{th} episode was initiated during quarter t of the baseline period and takes the value of 0 otherwise
- $CJR_{i,k}$ is an indicator that takes the value of 1 if the i^{th} episode was initiated by a CJR participant hospital k and takes the value of 0 otherwise
- $X_{i,k}$ are hospital, geographic, and patient characteristics in the baseline period

and the test is:

$$H_0 : b_3 = 0$$
$$H_1 : b_3 \neq 0$$

For mandatory CJR hospitals, the following outcomes failed both the linear and joint F-tests of parallel trends:

- HHA payments ($p<0.01$ for both linear and joint tests),
- 30-day post-episode payments ($p<0.01$ for the joint test and $p<0.05$ for the linear test),
- any HHA use ($p<0.05$ for both linear and joint tests),
- Number of outpatient physical therapy/occupation therapy (PT/OT) visits ($p<0.10$ for the joint test and $p<0.05$ for the linear test), and
- Without self-reported pain for those first discharged to SNFs ($p<.10$ for the joint test and $p<0.05$ for the linear test).

For opt-in hospitals in voluntary MSAs, no outcomes failed both the linear and joint F-tests of parallel trends.

For the non-opt-in hospitals in voluntary MSAs, the following outcomes failed both the linear and joint F-tests of parallel trends:

- 30-day post-episode payments ($p<0.10$ for both linear and joint tests),

- Readmission payments ($p<0.05$ for the joint test and $p<0.10$ for the linear test), and
- Reduced pain for those first discharged to HHAs ($p<0.05$ for the joint test and $p<0.01$ for the linear test).

Results for the parallel trends tests are included in Appendix K.

d. Analysis of the impact of CJR on subpopulations

Our analysis of the differential impact of the CJR model on subpopulations with historically poorer access to care and health outcomes is based on the DiD methodology. We study the differential impact of the CJR model by estimating the impact of the CJR model on a target subpopulation and a comparison subpopulation, and then estimate the difference between the two CJR model impacts to determine if the CJR model impacts the target subpopulation differently than the comparison subpopulation. The estimation of both differential impacts takes place in a single regression, subject to the constraint that the coefficients on risk-adjustment variables are the same for both the target and comparison subpopulations.

For this report, we studied three subpopulations: patients who are Black or African American, patients who are eligible for both Medicare and Medicaid (dually eligible), and patients who are Black or African American *and* dually eligible. The comparison subpopulations are, respectively, white patients, patients who are not dually eligible, and patients who are both white and not dually eligible. The estimated differential impact represents how the difference in the risk-adjusted average outcome between the target (e.g., Black/African American patients) and comparison (e.g., white patients) subpopulations changed between the baseline and intervention periods as a result of the CJR model. In other words, it represents the difference between the effect of the CJR model on the target subpopulation and the effect of the CJR model on the comparison subpopulation.

In the analysis of the impact of CJR on subpopulations with historically poorer access to care and health outcomes, we control for changes in patient characteristics through risk adjustment just as we do in the main analyses. However, despite rigorous risk adjustment, a change in the complexity of the subpopulation could still effect the DiD estimates. For example, DiD estimates that show a decrease in payment or an increase in quality for a subpopulation could be due in part to CJR hospital participants selecting less complex patients from this subpopulation in the intervention period.

2. Risk adjustment to control for differences in beneficiary demographics and clinical risk factors

a. Claims-based risk adjustments

In the DiD models, we controlled for potential differences in beneficiary demographics, clinical characteristics observed before hospitalization, and provider characteristics (represented by $X_{i,i,t}$ in Equation 2 above). Demographic factors included age categories, sex, age and sex interactions, race/ethnicity indicators, Medicaid eligibility status, and disability status. All outcomes were risk adjusted for the episode's hip fracture status, procedure type (hip or knee),

and MS-DRG (469 or 470). To control for participation in other Medicare initiatives, we used a dummy variable that indicated whether the beneficiary was in the Medicare Shared Savings Program (MSSP), Pioneer ACO Model, or Next Generation ACO Model during the episode.[22] To control for prior health conditions, we used HCC indicators for the 12 months preceding the anchor hospitalization,[23] as well as indicators for obesity, diabetes, hypertension, and tobacco use, generated from the claims data. To further control for case-mix differences, we included measures of prior care use in the following settings: acute care IPPS hospital, emergency department visits, long-term care hospital (LTCH), SNF, IRF, hospice, other Part A inpatient, custodial nursing facility, and home health agency (HHA).

We also controlled for provider characteristics that might be related to the outcomes of interest, such as hospital bed count, for-profit status, and previous Bundled Payments for Care Improvement initiative LEJR experience and previous Bundled Payments for Care Improvement initiative experience in a clinical episode other than LEJR. In October 2018, the Bundled Payments for Care Improvement Advanced initiative began. This CMMI model also includes LEJR as a clinical episode and aims to reduce payments, while maintaining or improving quality. CJR participant hospitals could not participate the in Bundled Payments for Care Improvement Advanced initiative for LEJR clinical episodes; however, hospitals and surgeons in the control group could participate. We found that 46% of mandatory control group episodes that started on or after October 1, 2018 were attributed to the Bundled Payments for Care Improvement Advanced initiative. To account for contamination in our control group by this other CMMI model, we included an indicator variable that identifies control group LEJR episodes performed by surgeons or at hospitals participating in the Bundled Payments for Care Improvement Advanced model.

While the same demographic and enrollment status indicators were included for all outcomes, we considered alternative aggregation levels to control for prior care use, prior health conditions, and regional characteristics (Exhibit C-9). To assess different specifications, we split the sample into a model development and a validation sample and estimated each model using data from the model development sample. We then evaluated the models' goodness of fit (Akaike Information Criterion (AIC), Bayesian Information criterion (BIC) criteria, and R-square) in the model development sample and their predictive performance in the validation sample.

---

[22] Beneficiaries with episodes during or after July 2017 that were aligned with MSSP track 3, Next Generation ACO, or Comprehensive End Stage Renal Disease Care Model and were excluded from the CJR model.

[23] The Hierarchical Condition Category (CMS-HCC) model is a prospective risk-adjustment model used by CMS to adjust Medicare Part C capitation payments for beneficiary health spending risk. The model adjusts for demographic and clinical characteristics. The clinical component of the model uses diagnoses from qualifying services grouped into numerous HCC indicators. The HCC indicators in the risk adjustment model included: sepsis, different types of cancer, diabetes, obesity, malnutrition, rheumatoid arthritis, coagulation defects, dementia, drug/alcohol dependence, mood disorder, Parkinson's disease, seizure disorders, cardio-respiratory failure, congestive heart failure, angina, heart arrhythmias, stroke, vascular disease, chronic obstructive pulmonary disease, macular degeneration, kidney disease, and renal failure. Pope, Gregory C.; Kautter, John; Ellis, Randall P.; Ash, Arlene S.; Ayanian, John Z.; Iezzoni, Lisa I.; Ingber, Melvin J.; Levy, Jesse M.; and Robst, John, "Risk adjustment of Medicare capitation payments using the CMS-HCC model" (2004). *Quantitative Health Sciences Publications and Presentations.* Paper 723.

Exhibit C-9: Predictive risk factors used to risk-adjust claims-based outcomes

Domain	Variables
Characteristics of the procedure	▪ Anchor MS-DRG ▪ Hip fracture status ▪ Procedure type (hip or knee)
Patient demographics and enrollment	▪ Age (under 65, 65-79, 80+) ▪ Sex ▪ Race/ethnicity ▪ Medicaid status ▪ Disability status at enrollment in Medicare (not ESRD) ▪ Attribution to Medicare Shared Savings Program, Pioneer ACO Model, or Next Generation ACO Models during the CJR episode
Prior health conditions	▪ CMS-HCC version 21 indicators from qualifying services and diagnoses (those meeting a threshold of at least 1%) from claims and data for 12 months preceding the anchor hospitalization ▪ Obesity indicator ▪ Diabetes indicator ▪ Hypertension indicator ▪ Tobacco use indicator
Prior use	▪ Prior care use (any acute care inpatient, emergency department visits, IRF, SNF, HHA, hospice, other Part A inpatient, LTCH, and custodial nursing facility service) variables used in risk adjustment varied by model[a] • Binary indicators for any care use in the six months preceding the start of the episode • Binary indicators for any care use in the one month preceding the start of the episode • Number of days of care use in the six months preceding the start of the episode
Geography	▪ State indicators
Hospital provider characteristics	▪ Bed count ▪ For-profit status ▪ Bundled Payments for Care Improvement LEJR experience ▪ Bundled Payments for Care Improvement experience in a clinical episode other than LEJR ▪ LEJR performed by surgeons or at hospitals participating in the Bundled Payments for Care Improvement Advanced model for LEJR clinical episodes (control group only)

Source: Risk adjustment variables were contructed from Medicare fee-for-service claims and beneficiary enrollment data, December 2016 POS, FY 2016 CMS Annual IPPS, CMS Master Data Management, Bundled Payments for Care Improvement initiative participant list, and Bundled Payments for Care Improvement Advanced initiative participant list.

Notes: ACO = accountable care organization, ED = emergency department, ESRD = end-stage renal disease, FY = fiscal year, HCC = hierarchical condition category, HHA = home health agency, IPPS = inpatient prosepctive payment system, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, LTCH = long-term care hospital, MS-DRG = Medicare Severity-Diagnosis Related Group, POS = provider of services, SNF = skilled nursing facility.

[a] The optimal specification for each prior use variable was chosen using the goodness of fit criteria for each outcome. The binary 6-month indicators were used for: SNF payment, IRF payment, HHA payment, Part B payment, unplanned readmissions, ED use, number of SNF days, and first discharge setting. The binary 1-month indicators were used for: complications and mortality. The indicators for number of days in the past 6 months were used for: total episode payment, readmissions payment, 30-day post-episode payment, number of IRF days, number of HHA visits, number of outpatient PT/OT visits, and number of HHA PT/OT visits.

b. Assessment-based risk adjustment

We applied risk-adjustment models endorsed by the National Quality Forum (NQF) and adopted by CMS for the IRF (average change in mobility score),[24] SNF (improved transfer, locomotion on unit, and walking in corridor),[25,26] and HHA settings (improved ambulation/locomotion, improved bed transferring, and improvement in the frequency of pain when moving around).[27,28,29] We made some modifications to the risk-adjustment models for these measures to better align with the needs of the evaluation. The risk-adjustment model for the SNF measure "Improved Status in Toilet Use" was designed specifically for the CJR model evaluation. We relied on clinical and PAC experts to draft an exhaustive list of assessment-based risk factors to potentially control for, and used a stepwise regression approach to develop a parsimonious risk adjustment model for this outcome measure.

For all measures, we dropped certain assessment-based covariates from the existing risk adjustment models in the following three scenarios: first, if they had a low prevalence (less than 1%) in the CJR population and were not statistically significant risk factors; second, if they were perfect predictors of the outcome (i.e., the outcome was always the same for a given value of the covariate); or third, if they had p-values greater than 0.05 and did not significantly improve the model's goodness of fit (c-statistic and pseudo-R-squared for logistic regressions and R-squared, AIC, and BIC criteria for ordinary least squares (OLS) regressions).

All risk adjustment models controlled for the length of the anchor hospitalization and the patients' functional status at the start of care. All SNF and HHA outcomes controlled for whether the patients were readmitted to the SNF or HHA provider after the anchor hospitalization. We also controlled for potential differences in characteristics of the procedure, patient demographics and enrollment, prior health conditions, utilization measures preceding the start of the anchor hospitalization, geography, and hospital provider characteristics (Exhibit C-10). We considered alternative aggregation levels to control for prior service use (Exhibit C-10) and selected a specific subset of prior service use variables for each outcome that improved the model's goodness of fit. For the SNF measures, we included additional Minimum Data Set (MDS)-based risk-factors based

[24] RTI International (2015). Inpatient Rehabilitation Facility Quality Reporting Program: Specifications for the Quality Measures Adopted through Fiscal Year 2016 Final Rule. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/IRF-Quality-Reporting/Downloads/IRF_Final_Rule_Quality_Measure_Specifications_7-29-2015.pdf

[25] RTI International (2016). MDS 3.0 Quality Measures User's manual, version 10.0. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/NursingHomeQualityInits/Downloads/MDS-30-QM-Users-Manual-V10.pdf

[26] The without self-reported pain measure for the SNF setting is NQF-endorsed and not risk-adjusted.

[27] CMS (2016). Home health agency quality measures: technical documentation of oasis-based patient outcome measures, Revision 5. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/HHQIQualityMeasures.html

[28] Nuccio EJ, Richard AA, Hittle DF (2011). Home health agency quality measures: logistic regression models for risk adjustment. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Downloads/HHQILogisticRegressionModelsforRiskAdjustment.pdf

[29] Hittle DF, Nuccio EJ (2017). Home health agency patient-related characteristics reports: technical documentation of measures. Available at: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HomeHealthQualityInits/Downloads/HHQILogisticRegressionModelsforRiskAdjustment.pdf

on t-tests and their ability to improve the model's goodness of fit. These additional factors spanned several MDS domains, including cognitive, mood and behavior status, bowel and bladder status, health condition, functional status, skin condition, and psychiatric/mood disorder. Finally, we controlled for the number of days (up to 14 days) between discharge from the anchor hospitalization and the start of home health care for patients who were discharged from the hospital directly to home health care. It is likely that patients' functional status will substantively improve over the days following their anchor hospitalization discharge.

Exhibit C-10: Predictive risk factors used to risk-adjust assessment-based outcomes

Domain	Variables
Characteristics of the procedure	▪ Anchor MS-DRG ▪ Hip fracture status ▪ Procedure type (hip or knee)
Patient demographics and enrollment	▪ Age ▪ Sex ▪ Medicaid status ▪ Disability status at enrollment in Medicare (not ESRD) ▪ Alignment to Medicare Shared Savings Program, Pioneer, or NextGen ACO during CJR episode
Prior health conditions	▪ HCC score from qualifying services and diagnoses from Medicare claims data for 12 months preceding admission to the anchor hospitalization
Prior use	▪ Prior care use (any acute care inpatient, emergency department visits, IRF, SNF, HHA, hospice, other Part A inpatient, LTCH, and custodial nursing facility service) variables used in risk adjustment varied by model[a] • Binary indicators for any care use in the six months preceding the start of the episode • Binary indicators for any care use in the one month preceding the start of the episode • Number of days of care use in the six months preceding the start of the episode
Geography	▪ State indicators
Hospital provider characteristics	▪ Bundled Payments for Care Improvement LEJR experience ▪ Bundled Payments for Care Improvement experience in a clinical episode other than LEJR ▪ LEJR performed by surgeons or at hospitals participating in the Bundled Payments for Care Improvement Advanced model for LEJR clinical episodes (control group only)
Anchor inpatient stay	▪ Length of inpatient stay, and length of stay squared
PAC assessment-based measures (MDS, OASIS, IRF-PAI) at the start of the PAC stay	▪ SNF readmission or HHA resumption of care after being discharged from the anchor hospitalization ▪ Functional status at PAC admission with respect to the outcome being measured ▪ Days between discharge from the anchor hospitalization and the start of home health care ▪ Assessment-specific measures of factors related to cognitive status, mood and behavior status, bowel and bladder status, health conditions, functional status, skin condition, and psychiatric/mood disorders

Source: Risk adjustement variables were contructed from Medicare fee-for-service claims and beneficiary enrollment data, IRF PAI, SNF MDS, HH OASIS, CMS Master Data Management, Bundled Payments for Care Improvement initiative participant list, and Bundled Payments for Care Improvement Advanced initiative participant list.

Notes: ACO = accountable care organization, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, HHA = home health agency, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility-Patient Assessment Instrument, LEJR = lower extremity joint replacement, LTCH = long-term care hospital,

MDS = Minimum Data Set, MS-DRG = Medicare Severity-Diagnosis Related Group, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, SNF = skilled nursing facility.

[a] The optimal specification for each prior use and assessment-based variable was chosen using the goodness of fit criteria for each outcome. The binary 6-month indicators were used for: HHA ambulation, HHA bed transfer, HHA pain, SNF motion, SNF toileting, and IRF mobility. The binary 1-month prior SNF use indicator was included in the SNF motion model.

3. Model types

We used a variety of models including logistic, Poisson, multinomial logit, OLS regressions, and two-part models (Exhibit C-11). Models were estimated depending on the type and characteristics of the outcome measure. For example, logistic models were estimated for the discrete quality outcomes (i.e., all claims-based quality of care measures and the assessment-based measures for improved functional status). A multinomial logit model was applied to first-discharge setting. OLS models were estimated for the continuous total number of days or visits measures (e.g., number of SNF days, number of IRF days, number of HHA visits, number of HHA PT/OT visits, and number of PT/OT visits) as well as total episode payments, Part B payments, and the assessment-based quality measure for the average change in mobility score for IRF patients. Two-part models were favored for payment outcomes where more than 5% of individuals had zero payments for the particular outcome. These payment outcomes included the individual Part A payments that exhibited zero-mass and skewness.

Exhibit C-11: Outcomes by model type

Model type	Outcomes
Ordinary least squares (OLS)	▪ Total episode payments ▪ Part B payments ▪ Number of IRF days ▪ Number of SNF days ▪ Number of HHA visits ▪ Number of PT/OT Visits, outpatient ▪ Number of PT/OT Visits, HHA ▪ Average change in mobility score, IRF
Two part models (Probit/OLS)	▪ Readmission payments ▪ IRF payments ▪ SNF payments ▪ HHA payments ▪ 30-day PEP payments
Multinomial logistic	▪ First post-acute discharge was to IRF ▪ First post-acute discharge was to SNF ▪ First post-acute discharge was to HHA ▪ Discharge to home without home health
Logistic	▪ Any HHA visits ▪ Unplanned readmission ▪ Emergency department visit ▪ Complications, among elective episodes ▪ All-cause mortality ▪ Improved status in transfer, locomotion, and walking in the corridor, SNF ▪ Improved status in toilet use, SNF ▪ Without self-reported moderate to severe pain, SNF ▪ Improved status in ambulation/locomotion, HHA ▪ Improved status in bed transferring, HHA ▪ Improvement in the frequency of pain when moving around, HHA

Note: HHA = home health agency, IRF = inpatient rehabilitation facility, OLS = ordinary least squares, OT = occupational therapy, PEP = post-episode period, PT = physical therapy, SNF = skilled nursing facility.

Estimates from the multivariate regression models were used to construct model-predicted outcomes under two scenarios (baseline and intervention) for both CJR and control group hospitals. To control for changes in service and case mix over time, as well as differences between CJR and non-CJR beneficiaries, we used the same reference population of beneficiaries to calculate predicted outcomes for CJR and control group episodes. The reference population used in this report is all CJR beneficiaries during the baseline and intervention period. Given the design of the CJR model (randomly sampling MSAs to participate), we accounted for clustering at the MSA level in the estimation of our standard errors in all of our regression models for mandatory hospitals. In our regression models for opt-in and non-opt-in hospitals, we accounted for clustering at the hospital level in the estimation of our standard errors because the decision whether or not to continue participation in PY3 was at the hospital level.

4. Sensitivity analyses

ACO Participation: Similar to what was done in our third annual report, a number of sensitivity analyses were performed on the findings for the claims-based outcomes in the main analysis.[30] We excluded episodes generated under MSSP, Pioneer ACO, or Next Generation ACO to identify whether these exclusions would change the DiD estimate. We found that this exclusion did not materially affect any of the findings in the main analysis.

Beneficiaries without FFS Coverage in the Six Months Prior to the LEJR: In addition, 4.9% of the LEJR episodes were not included in the risk-adjusted DiD estimation because they did not have information related to prior health care conditions due to the lack of fee-for-service coverage in the six months prior to the anchor hospitalization. To explore the change in total episode payment with these episodes included, we utilized two methods to impute the values of prior health care condition variables for use in our risk-adjusted DiD estimation. In the first method, we imputed values for each prior use variable by taking the average value of that variable, stratified by beneficiary age and MSA. In the second method, for each prior use variable, we used the sample of observations with no missing values to train a model using all non-prior use variables to predict the value of the prior use variable. We then used the model to obtain out-of-sample predictions, which we used as imputed values for the sample of observations with the missing prior use variables. Regardless of imputation method, our findings were robust. The total payment DiD including these episodes was comparable to the DiD excluding these episodes.

Bundled Payments for Care Improvement (BPCI) Initiative: We also ran additional sensitivity tests to examine the impact of prior hospital participation in the Bundled Payments for Care Improvement (BPCI) Initiative for LEJR clinical episodes on the DiD estimates. There is an imbalance between the mandatory CJR and control groups in the number of intervention episodes contributed by former BPCI LEJR hospitals. We found that a larger number of BPCI Model 1 participants were included in the mandatory CJR group than the control group, and BPCI Model 1 ended on March 31, 2016, just before the start of the CJR model. Also, we found that more BPCI LEJR hospitals exited BPCI prior to the end of the BPCI initiative in mandatory CJR MSAs than in the control group likely because they wanted to join the CJR model. As a result, during the CJR intervention and prior to the end of the BPCI Initiative, 22% of intervention episodes from mandatory CJR hospitals were contributed by former BPCI LEJR hospitals, compared to 9% of control group episodes.

In our main analysis, we handled BPCI episodes and hospitals following the below specifications: 1) we excluded BPCI LEJR episodes, i.e., episodes from BPCI LEJR hospitals during the time period in which they were participating in BPCI LEJR; 2) we included episodes from former BPCI LEJR hospitals from time periods in which they were not participating in BPCI LEJR (the time

[30] Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement model - third annual report appendices. https://innovation.cms.gov/data-and-reports/2020/cjr-thirdannrpt-app. 2020: E36-37.

period prior to joining BPCI and the time period after exiting BPCI);[31] and 3) we controlled for prior BPCI LEJR participation in the DiD. However, the imbalance in the number of episodes from former BPCI LEJR hospitals between the CJR and control group MSAs raises concerns about the comparability of the treatment and control groups.

We performed two sensitivity tests with the intention of balancing the contribution of intervention episodes from former BPCI LEJR hospitals across the mandatory CJR and control group samples. In the first test, we excluded episodes contributed by former BPCI LEJR hospitals. The DiD for total payments increased from -$1,511 to -$1,256, a $255 difference.

In the second test, we included episodes from BPCI LEJR hospitals during the time period in which they were participating in BPCI. In other words, we ignored BPCI attribution and included BPCI episodes in the mandatory CJR and control groups in the analysis. The DiD for total payments increased from -$1,511 to -$1,344, a $167 difference.

Based on our two sensitivity tests, the larger contribution of intervention episodes by former BPCI LEJR hospitals does appear to overestimate the reductions in average episode payments due to the CJR model by roughly $200 per episode.

Geographic and Time Risk-Adjusters: Finally, we performed a sensitivity analysis in which we varied the geographic and time risk-adjusters in our regression model. In addition to beneficiary, hospital, and geographic controls, our standard specification includes state indicators and omits time indicators. In this sensitivity analysis, we estimated the DiD model sixteen times, each time accounting for pairings of a different set of geographic indicators—region, state, MSA, or hospital—and time indicators—none, year, seasonal quarter, or year and seasonal quarter. In each pairing, the DiD estimate was statistically significant at the 1% level and ranged from -$1,513 to -$1,313 (Exhibit C-12). This suggests that the CJR model led to a large and statistically significant relative reduction in total payments, regardless of the exact specification of geography and time used.

[31] There is an exception for the ITT analysis because it includes voluntary CJR and control group MSAs. BPCI LEJR hospitals in voluntary CJR or control MSAs that exited BPCI after January 2018 are not included in the ITT analysis because hospitals in voluntary MSAs could not opt in to the CJR model after January 2018.

Exhibit C-12: The CJR model's relative reduction in total payments is robust to the set of risk-adjusters used

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention). To account for CJR participant hospitals shifting a lower share of TKAs to the hospital outpatient setting, the control group includes outpatient TKAs, which are weighted to balance the CJR group.

Notes: The estimated relative changes in total payments are the result of DiD models, each with the corresponding set of geographic- and time-level indicators. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded shapes, respectively. The error bars represent 90% confidence intervals.

Because CJR participant hospitals shifted a lower share of TKAs to the hospital outpatient setting, the control group includes outpatient TKA episodes that have been weighted to balance the episode volume in the CJR hospitals.

DiD = difference-in-differences, MSA = metropolitan statistical area, TKA = total knee arthroplasty.

IV. Outpatient TKA descriptive methods

CMS removed TKA from the inpatient only list, effective January 2018, and Medicare now covers TKAs performed in the hospital outpatient setting.[32] We monitored the occurrence of outpatient TKA in CJR and control group hospitals and MSAs from January 1, 2018 through October 3, 2019.[33]

We identified TKAs performed on outpatients using Part B claims data (CPT code 27447 assigned to C-APC 5115 with status indicator "J1") and TKAs performed on inpatients using Part A claims data (MS-DRG 469 or 470 with ICD codes on the claim indicating a knee procedure). We excluded TKA discharges that did not meet CJR episode eligibility.

We calculated percent of TKAs performed as outpatients overall (in 2018 and 2019) and by quarter for the mandatory CJR and control groups by dividing the number of outpatient TKAs meeting episode eligibility by the sum of all TKAs meeting episode eligibility (TKAs performed on inpatients and outpatients).

For mandatory hospitals, we observed a nine percentage point difference in percent of TKAs performed in the outpatient setting since 2018 (CJR, 22.4% vs. control, 31.5%). For the two hospital groups in voluntary MSAs, the difference in percent of TKAs sent to the hospital outpatient departments was smaller (opt-in hospitals, 35.1% vs. 37.5% in the matched control group; non-opt-in hospitals, 30.4% vs. 32.3% in the matched control group).

[32] CMS. Medicare Program: Hospital Outpatient Prospective Payment and Ambulatory Surgical Center Payment Systems and Quality Reporting Program. https://s3.amazonaws.com/public-inspection.federalregister.gov/2017-23932.pdf. 2017.

[33] To match the episode inclusion criteria for the DiD analyses, inpatient and outpatient TKA discharges were included if the 90-day post-discharge period occurred on or before December 31, 2019. A patient discharged on October 3 would have a 90-day post-discharge period ending on December 31.

V. Savings to Medicare due to the CJR model

We calculated Medicare savings by subtracting reconciliation payments to CJR participant hospitals from the change in non-standardized paid amounts due to the CJR model. Medicare savings was calculated on both a total and a per-episode basis.

Medicare savings = Change in non-standardized paid amounts – Reconciliation payments

A. Change in non-standardized paid amounts

The change in non-standardized paid amounts was based on estimates from a DiD model of per-episode standardized paid amounts. The DiD estimates were multiplied by negative one and converted to non-standardized paid amounts using a ratio of non-standardized to standardized Medicare paid amounts from CJR intervention episodes (Exhibit C-13). This method produced a per-episode estimate of the change in non-standardized paid amounts. The total change in non-standardized paid amounts was produced by multiplying the per-episode estimate by the total number of episodes.

Exhibit C-13: Ratios of non-standardized to standardized Medicare paid amounts by hospital group

Time period	Mandatory hospitals	Opt-in hospitals	Non-opt-in hospitals
Baseline	1.032	1.035	1.000
PY1	1.032	1.037	1.002
PY2	1.032	1.040	1.003
PY3	1.036	1.037	1.002
PY4	1.042	1.045	1.009
Cumulative	1.033	1.038	1.002

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

Notes: The ratio is calculated as the average non-standardized (actual) paid amounts divided by the average standardized paid amounts for episodes. outpatient TKAs are included in the calculation of the ratios for mandatory and opt-in hospitals. The anchor payment (MS-DRG payment for inpatient episodes or APC payment for outpatient TKAs) was subtracted from the total episode payment before calculating the ratio.

APC = ambulatory payment classification, MS-DRG = Medicare Severity-Diagnosis Related Group, PY = performance year, TKA = total knee arthroplasty.

B. Reconciliation payments

Reconciliation payments are defined as total payments made to CJR participants by Medicare net of repayments from CJR participants to Medicare. Reconciliation payments can be positive or negative. In the program literature, they are often referred to by the term "net payment reconciliation amounts" or "NPRA." These data were provided by the CMS CJR payment contractor. Reconciliation payments per episode were calculated by dividing total reconciliation payments by the total number of CJR episodes.

C. Hospital group estimates versus cumulative estimates

We reported estimates for three different hospital groups: mandatory CJR hospitals (excluding those with a low-volume or rural exemption), opt-in hospitals in voluntary MSAs, and non-opt-in hospitals in voluntary MSAs.

For each hospital group, a comparison group of episodes from control hospitals was constructed, and a DiD model was used to produce an estimate of per-episode reductions in standardized paid amounts.[34]

We also reported estimates for all hospital groups combined. We could have added together the hospital group estimates, or constructed a weighted average of the per-episode estimates from each hospital group. However, these approaches would have led to overly conservative (too wide) confidence intervals and ranges. Instead, we pooled the three regressions together into a single overarching model. This allowed us to construct an accurate confidence interval for the weighted average of per-episode reductions in standardized payments.

D. Performance year estimates versus total estimates

We reported Medicare savings estimates for all four performance years combined, and on a per-performance year basis. The performance year estimate was derived from a DiD model that compared episode payments in a given performance year to episode payments during the baseline period. Thus, for instance, the PY2 estimate was determined by comparing the change in standardized payments per episode between PY2 and the baseline period in CJR hospitals to that same change in control hospitals.

Estimates of the total savings to Medicare over multiple performance years for a specific hospital group were constructed from the performance year estimates. We constructed a weighted average of the performance year estimates, with the weights reflecting the proportion of all episodes that occurred in a given performance year. We also estimated a confidence interval for the weighted average, allowing for the construction of our estimated ranges on total savings.

Estimates of the total savings to Medicare including all performance years and all hospitals were constructed by pooling hospital groups as described above, and then constructing a weighted average of the performance year estimates from the pooled regression.

E. Considerations

We excluded low-volume and rural hospitals in mandatory MSAs from our main estimates of Medicare savings. We do not include the low-volume and rural hospitals located in the 34 mandatory MSAs in the analysis of opt-in hospitals in voluntary MSAs because low-volume and rural hospitals differ in important ways that are likely to affect performance in the model. Further,

[34] To construct an appropriate counterfactual, we employed the propensity score weighting (PSW) method and included *all* control outpatient TKAs in the DiD model, weighted by the hypothetical probability of an outpatient TKA being inpatient if the hospital had been participating in the CJR model.

unlike the voluntary hospitals, the low-volume and rural hospitals are located in MSAs with higher average historical payments and the majority of hospitals in these mandatory MSAs are participating in the CJR model. Since an analysis of low-volume and rural hospitals would also need to account for their ability to select to continue in the model, we do not include them in the analysis of the CJR hospitals in the 34 mandatory MSAs that were continuously required to participate throughout the entire model. Producing a separate impact estimate for this subgroup would be a challenge because it would require constructing an appropriate comparison group. This group of hospitals is small (39 low volume hospitals and 37 rural hospitals, with 20% opting to continue participation in PY3), and hospitals were incentivized by the CJR model to reclassify to rural by offering rural hospitals lower stop-loss limits (e.g., 5% in PY4 compared to 20% for all other hospitals). The CJR model may have influenced hospitals decision to reclassify as rural and this same incentive to reclassify to rural was not present in the control group.

Rather than producing separate impact estimates for this group, we performed sensitivity analyses (see Appendix E). Those analyses suggest that inclusion of the low volume and rural hospitals located in mandatory MSAs would not have changed our Medicare program savings results and conclusions. Those analyses result in point estimates of Medicare savings that lie within the range reported in our main analysis.

VI. Patient Survey

We developed the CJR patient survey to explore differences between CJR and control patients in functional status and pain, need of caregiver help, care experience, and overall satisfaction at the end of the episode. The patient-reported outcomes in the survey capture information that is not available from other data sources, such as claims or assessment data.

A. Survey sample

We administered the patient survey in two waves to a census of CJR and control patients who had inpatient LEJR surgery during the fourth performance year. Each wave covered two months of LEJR episodes, March or April 2019 and September or October 2019. Exhibit C-14 describes the patient survey waves.

Exhibit C-14: Patient survey sample by survey wave

Wave	Discharge date	CJR LEJR episodes	CJR patients sampled	Control LEJR episodes	Control patients sampled
5	March or April 2019	9,046	9,046	8,988	8,988
6	September or October 2019	8,144	8,144	8,326	8,326

Source: CJR evaluation team analysis of survey data for patients with discharge from LEJR surgery in March, April, September or October 2019.

Note: LEJR = lower extremity joint replacement.

In both waves, we selected all available CJR and control patients.

Prior survey waves included beneficiaries from mandatory CJR hospitals and corresponding control group hospitals with LEJR in March, April, August, and September 2018. These prior survey waves occurred during the Bundled Payments for Care Improvement (BPCI) Models 2-4. The survey waves presented in this report covered LEJR occurring in 2019 during the BPCI Advanced model. This resulted in several key changes to the underlying CJR and control populations between survey waves including: (1) hospitals participating in BPCI LEJR episodes were previously excluded from analysis, and with the end of BPCI are now included in both the CJR and control groups; (2) physician group practices participating in BPCI LEJR episodes were previously excluded from analysis, and with the end of BPCI are now included in both the CJR and control groups; (3) CJR takes precedence over BPCI Advanced. While no CJR episodes overlap with participation in BPCI Advanced LEJR episodes, control group episodes can overlap with BPCI Advanced (either through the control hospital itself, or physician groups practicing at control hospitals). Given these changes, we determined that it was conceptually preferable not to pool data from prior survey waves (2018) with our most recent waves of data (2019).

1. Survey administration

We mailed surveys to patients between 60 and 120 days after their LEJR discharge (94 days after discharge, on average). Reminder postcards were sent one week later. Four weeks after the initial mailing, we mailed non-respondents a second survey. Outbound telephone follow-up with non-

respondents began approximately eight weeks after the first mailing. On average, respondents returned the survey 121 days after hospital discharge. Sensitivity analysis did not find any evidence that average time between discharge and survey response differed between the CJR and control patients, nor did we find any evidence that results varied when we controlled for time between discharge and survey receipt.

2. Response rates and analytic samples

In the sample pooled across waves 5 and 6, the response rate was 65.6% for CJR patients and 68.0% for control patients, a statistically significant difference ($p<0.05$; Exhibit C-15). There were 11,273 surveys completed by CJR respondents with episodes during the four months covered by the two survey waves, including patients from 368 of the 388 mandatory CJR participant hospitals. There were 11,765 surveys completed by control respondents, including patients from 366 of 376 control hospitals.

The response rate for beneficiaries with a hip fracture was 41.3% for CJR patients and 44.0% for control patients (not statistically different between the two groups). There were 1,080 surveys completed by CJR patients with hip fractures, including patients from 285 of the 347 CJR mandatory hospitals where there was at least one hip fracture LEJR surgery during the sampling period. There were 1,088 completed surveys in the corresponding mandatory control hospitals, from 278 of the 327 hospitals where there was at least one hip fracture LEJR surgery during the sampling period. Sample sizes and response rates for the subpopulations with historically poorer access to care and health outcomes analysis are also provided in Exhibit C-15.

Exhibit C-15: Sample size and response rate overall, for patients with hip fracture, and subpopulations, waves 5 and 6 combined

Group	Patients surveyed (starting sample)		Survey responses received (analytic sample)		Response rate		
	CJR	Control	CJR	Control	CJR	Control	p-value
Overall[a]	17,190	17,314	11,273	11,765	65.6%	68.0%	p<0.05
Hip fracture	2,617	2,472	1,080	1,088	41.3%	44.0%	p=0.14
Black or African American	874	1,122	507	669	57.0%	57.4%	p=0.84
White	14,660	15,137	9,968	10,584	67.4%	69.1%	p=0.11
Dual eligible	1,735	1,483	680	672	39.2%	45.3%	p<0.05
Non-dual eligible	15,455	15,831	10,593	11,093	68.5%	70.1%	p<0.10

Source: CJR evaluation team analysis of survey data for patients with discharge from LEJR surgery in March, April, September, or October 2019.

Notes: Differences in CJR and control response rates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

LEJR = lower extremity joint replacement.

[a] Overall includes patients with elective inpatient LEJR or inpatient LEJR due to hip fracture.

B. Analytic methodology

This section describes the general analytic approach for the Wave 5 and 6 patient survey analyses.

1. Survey domains and measures

We analyzed 20 survey measures, organized in five domains (functional status and pain, caregiver help, care transitions, care management, and overall recovery), described in Exhibit C-16. The patient survey instrument is available in Appendix H.

Exhibit C-16: Patient survey domains and measures

Domain	Survey measures[a]	Description of survey measures
Functional status and pain[b]	Change in mobility	Ability to walk by yourself without resting
		Difficulty walking up or down 12 stairs
		Difficulty rising from sitting
		Difficulty standing
		Use of a mobility aid
	Change in toileting	Difficulty getting on/off the toilet
	Change in pain	Frequency that pain interferes with normal activities
	Change in medication	Medication use for pain in the joint you had replaced
Overall recovery	Satisfaction with overall recovery	Satisfaction with overall recovery since leaving the hospital
Care management	Composite measure of satisfaction with care management	Health care providers listened to preferences
		Satisfaction with discharge destination
		Satisfaction with care coordination
		Satisfaction with treatment instructions
Care transition	Discharged from the hospital at the right time	Discharged from the hospital at the right time
	Received the right amount of post-discharge care	Received the right amount of post-discharge care
	Had all the medical equipment needed at home	Had all the medical equipment needed at home
Caregiver help	Received any caregiver help	Received any caregiver help
	Composite measure of caregiver help	Help needed putting on or taking off clothes
		Help needed bathing
		Help needed using the toilet

Notes: LEJR = lower extremity joint replacement.

[a] Items regarding pain and medication refer directly to the joint that received surgery. All other items refer directly to the anchor hospitalization.

[b] For the eight functional status and pain measures, we modeled the change in functional status, where change was the difference between recalled status the week prior to the LEJR surgery, and reported status at the time the survey was completed.

Survey respondents were asked to recall their functional status and pain the week prior to their hospitalization, and to report their functional status and pain at the time of the survey, across eight

related measures of function and pain. Each measure consisted of a Likert scale with three, four, or five levels. For each of the eight measures, we calculated the change in functional status or pain as the difference between a beneficiary's level of function/pain at the time of the survey and their recalled level of function/pain. We converted differences in levels of the Likert scale to percentage terms by dividing them by the average recalled level among CJR respondents. That is, the percentage difference is the difference between CJR and control respondents in relation to CJR respondents' recalled level of function or pain prior to their hospitalization.

In the caregiver help domain, measures of activities of daily living consisted of a Likert scale with three levels. Measures of satisfaction with care management or recovery consisted of a Likert scale with five levels. Results in these domains were normalized so that the lowest response category (e.g., "very dissatisfied" or "complete help needed") yielded a score of 0, and the highest response category (e.g., "very satisfied" or "no help needed") yielded a score of 100.

There were three measures of care transition. The first measure, timing of discharge, included three response options (discharged too early, at the right time, or too late). The second measure, level of post-acute care received, included three response options (level of care during two weeks after surgery was more than respondent needed, about right, or not enough). The third measure, did the respondent have all the medical equipment he/she needed when sent home, had two response options (yes or no).

2. Composite measures

We created composite measures for two domains. Reliance on caregiver help, conditional on having any caregiver help, summarized responses to three questions. Satisfaction with care management summarized responses to four questions. To create the composite measure, we translated response items into numeric scores and set them so that zero represented "the most negative amount of the construct being measured" (e.g., most amount of caregiver help). Response categories were added, so that the composite measure for a given domain was the sum of scores for its individual questions. For example, the "caregiver help" measure summarized three survey questions that each had three possible answers (0 – 'complete help needed', 1 – 'some help needed', or 2 – 'no help needed'). The composite measure of "caregiver help" therefore ranges from zero (maximum help needed) to six (no help needed for any of the three tasks). Consistent with the Consumer Assessment of Health Providers and Systems (CAHPS) scoring, we re-scaled the composite items so that scores ranged from zero to 100, where zero indicated the least favorable outcome of the construct being measured (i.e., greatest reliance on caregiver help, and least satisfaction with care management).

Exploratory factor analysis of early returns from the first wave of the survey in PY1 (which comprised approximately 85% of the total wave 1 responses) indicated that the survey items we grouped into composites were internally consistent and, for each composite, reflected a single construct that we could be summarized with one number.

3. Weighting

We employed entropy balancing to address potential differences in key patient characteristics across the CJR and control patients, and to mitigate potential differences between our sample of respondents and the populations from which they were drawn. The entropy balance weights minimize differences between the CJR and control patients on key attributes (see domains 1-4 in Exhibit C-15), and minimize differences in observable patient characteristics between CJR or control respondents relative to the full CJR population. For the analysis focused on respondents with hip fractures, we weighted the sample of CJR and control respondents to reflect the CJR population of beneficiaries with hip fractures.

C. Results estimation

For each of the patient survey measures, we estimated the difference between CJR and control patients. We separately analyzed responses of beneficiaries who received LEJR surgery after a hip fracture.

For our analysis, we utilized the non-linear model listed below for beneficiaries *i*, hospitals *k*, and wave *t* using a general functional form:

$$Y_{i,k,t} = b_1 + b_2 \cdot CJR_i + X_{i,k,t}' \cdot B + u_{i,k,t} \quad \textbf{(1)}$$

Where:

- Coefficient b_2 captures the difference in outcomes between CJR and control episodes.
- $X_{i,k,t}$ indicates risk factors controlled for in our model.

We also explored the impact of the CJR model on subpopulations with historically poorer access to care and health outcomes using patient survey data. (Discussion of the subpopulations analyses using claims data is in Section III.C.1.d). We jointly estimated the impact of the CJR model on respondents in the subgroup (e.g. Black/African American) and respondents not in the subgroup (e.g. white). We calculated the difference between these two estimates to determine if the CJR model affected subgroups with social risk factors differently than subgroups without such risk factors. Estimates were risk-adjusted for patient/episode-level characteristics within each subgroup.

D. Risk adjustment to control for differences in patient demographics and clinical risk factors

All survey analyses controlled for potential differences in characteristics of the procedure, patient demographics and Medicare enrollment status, prior health conditions, and survey dimensions (first four domains in Exhibit C-17). We selected these 15 patient-level characteristics as covariates for all survey analyses, based on the factors most strongly correlated with patient experience on the prior Bundled Payments for Care Improvement initiative patient survey and conceptual considerations (i.e., factors predicted to be important based on theory).

Exhibit C-17: Risk adjustment to control for differences in patient demographics and clinical risk factors

Domain	Variables
Characteristics of the procedure	▪ Fracture ▪ Knee replacement procedure ▪ MS-DRG
Patient demographics and Medicare enrollment status	▪ Age ▪ Sex ▪ Dual Medicare/Medicaid eligibility ▪ Originally qualified for Medicare due to disability ▪ Assignment to ACO ▪ Self-reported race/ethnicity[a] ▪ Self-reported education[a] ▪ Self-reported pre-hospital functional status[a]
Prior health conditions	▪ HCC score ▪ Stay in skilled nursing facility or nursing home in six months prior to admission
Survey dimensions	▪ Wave of survey ▪ Proxy status (patient had help from someone else in responding to the survey)
Optional patient, hospital, and MSA-level covariates[b]	▪ Survey mode (phone/mail) ▪ Self-reported income ▪ Hospital size (staffed beds) ▪ Hospital academic affiliation ▪ Hospital ownership type ▪ Hospital prior BPCI experience (LEJR) ▪ Hospital prior BPCI experience (non-LEJR) ▪ PGP prior BPCI experience (LEJR) ▪ Hospital in BPCI-Advanced (non-LEJR) ▪ LEJR market competitiveness in MSA ▪ Medicare Advantage penetration in MSA (%)

Source: Risk adjustment variables were contructed from Medicare fee-for-service claims and beneficiary enrollment data, December 2016 POS, FY 2016 CMS Annual IPPS, CMS Master Data Management, 2015-2016 Area Health Resource Files, and Bundled Payments for Care Improvement initiative participant list.

Notes: ACO = accountable care organization, BPCI = Bundled Payment for Care Improvement initiative, HCC = hierarchical condition category, LEJR = lower extremity joint replacement, MSA = metropolitan statistical area, MS-DRG = Medicare Severity-Diagnosis Related Group, PGP = physician group practice

[a] For risk adjustment measures that are self-reported (i.e., pre-hospital functional status; race/ethnicity; education), we coded all missing responses as 0 and included an additional binary variable indicating "missing item" (e.g., missing race/ethnicity).

[b] While the first four domains acted as fixed covariates for our models, each measure's final risk-adjusted model included some unique combination of these optional variables, as well as squared and interaction terms.

In addition to these 15 fixed variables, which we controlled for in all regressions, we ran a stepwise selection procedure on each outcome to test for additional control variables. Potential new variables included squared and interaction terms among the 15 fixed variables, as well as optional patient-level variables (i.e., survey response mode and self-reported income), hospital-level variables (i.e., hospital size, academic affiliation, ownership type, prior Bundled Payments for Care Improvement LEJR experience, prior Bundled Payments for Care Improvement non-LEJR experience, prior physician group practice (PGP) Bundled Payments for Care Improvement LEJR

experience, and hospital participation in Bundled Payments for Care Improvement Advanced in a non-LEJR episode), and MSA-level variables (i.e., LEJR market competitiveness and Medicare Advantage Penetration) (last domain in Exhibit C-17).

E. Comparing recalled functional status in the week prior to hospitalization between CJR and control respondents

We analyzed changes in self-reported functional status, and whether this differed for CJR and control respondents. We defined change as the difference between recalled status the week prior to the LEJR surgery and reported status at the time the survey was completed. Although we controlled for recalled pre-hospital functional status, our results may still be biased if CJR and control respondents had substantially different functional status prior to surgery. For each of the eight pre-hospital functional status measures, we calculated the standardized difference in the unweighted mean between CJR and control respondents. Standardized differences for pre-hospital functional status between CJR and control respondents were all below 0.10 for the pooled wave 5 and 6 overall sample, which is a conservative threshold for identifying potentially problematic differences between two groups (Exhibit C-18).[35] Among hip fracture respondents, standardized differences exceeded 0.10 for three measures of pre-hospital function, including rising from sitting, using the toilet, and medication intensity, although differences were only slightly above 0.10.

Exhibit C-18: Summary statistics in pre-hospital functional status between CJR and control respondents, waves 5 and 6 combined

Measure	Response range[a]	Overall			Hip fracture		
		CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.
Walking without rest	-4 to 4	2.75	2.66	0.09	3.15	3.14	0.07
Going up or down stairs	-3 to 3	2.22	2.21	0.03	2.88	2.86	0.05
Rising from sitting	-4 to 4	2.74	2.74	0.01	3.92	3.97	0.11
Standing	-4 to 4	2.97	2.95	0.03	4.04	4.06	0.06
Use of a mobility device	-2 to 2	2.24	2.22	0.03	2.30	2.32	0.02
Getting on or off the toilet	-4 to 4	3.02	3.02	0.01	4.07	4.09	0.12
Pain limiting regular activities	-4 to 4	1.98	1.95	0.03	4.23	4.23	0.08
Medication intensity	-3 to 3	2.76	2.71	0.06	3.68	3.62	0.10

Source: CJR evaluation team analysis of patient survey data for LEJR episodes with discharge in March, April, September, or October 2019.

***Notes*:** Means and standardized differences are unweighted.

LEJR = lower extremity joint replacement.

[a] Difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status prior to the hospitalization.

[35] Austin, P. C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behav Res* 46(3): 399-424.

F. Comparing claims-based patient characteristics between CJR and control respondents

Differences in functional status and pain between CJR and control respondents were risk-adjusted for a number of measures, including a fixed set of claims-based patient and episode characteristics (Exhibit C-19). For each of these measures, we calculated the standardized difference in the unweighted mean between CJR and control respondents. Standardized differences were below 0.10 for all variables except ACO assignment.

Exhibit C-19: Summary statistics in claims-based patient characteristics between CJR and control respondents, waves 5 and 6 combined

	Overall			Hip fractures		
Measure	CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.
Hip Fracture	0.10	0.09	0.01	N/A	N/A	N/A
Knee procedure	0.50	0.49	0.02	0.00	0.00	0.00
MS-DRG 469	0.03	0.04	0.01	0.15	0.14	0.01
Age	73.9	73.5	0.05	81.3	80.8	0.06
Female	0.64	0.64	0.01	0.74	0.71	0.07
Eligible for Medicaid	0.06	0.06	0.01	0.09	0.08	0.04
Disability, no ESRD	0.11	0.11	0.01	0.09	0.08	0.02
Assignment to ACO	0.44	0.49	0.11	0.39	0.46	0.14
HCC score	1.38	1.34	0.03	2.32	2.39	0.05
Prior SNF stay[a]	0.03	0.03	0.01	0.09	0.09	0.00

Source: CJR evaluation team analysis of patient survey data for LEJR episodes with discharge in March, April, September, or October 2019.

***Notes*:** Means and standardized differences are unweighted.

ACO = accountable care organization, ESRD = end-stage renal disease, HCC = hierarchical condition category, LEJR = lower extremity joint replacement, MS-DRG = Medicare Severity-Diagnosis Related Group, SNF = skilled nursing facility.

[a] Stay in skilled nursing facility or nursing home in six months prior to admission.

G. Sensitivity analysis

Unlike CJR hospitals, CMS allowed control hospitals to enroll in BPCI Advanced for LEJR, a voluntary episode-based payment model, which could potentially introduce bias. If such bias existed, we would expect results to change if we dropped control episodes attributed to BPCI Advanced LEJR. Hospitals participating in the BPCI Advanced LEJR clinical episode initiated 7.3% of all control episodes, and 38.4% of control episodes were attributed to physician group practices participating in the BPCI Advanced LEJR clinical episode. As a sensitivity analysis, we excluded all of these episodes, which did not change our results (see Appendix I, Exhibits I-3 and I-4). This suggests that BPCI Advanced involvement within control hospitals did not bias our CJR patient survey results.

H. Limitations

The analyses have potential limitations related to the sample, timing of the survey, potential for recall bias, and differential characteristics of CJR and control respondents. Roughly one in three patients did not respond to the survey. Although we applied nonresponse weights to account for observable patient characteristics, to the extent non-respondents differed from respondents on unobservable factors correlated with our outcomes of interest, our results may not generalize to all patients in CJR. Since most survey measures focus on past events (e.g., recalled functional status a week prior to surgery, PAC received weeks or months prior to the survey), incorrect recall may lead to mismeasurement of outcomes. This type of measurement error would not change the results, on average, because the same recall issue applies to both intervention and control groups, but it would reduce the precision of the estimates (greater confidence intervals).

As reported in the third annual report, beginning in 2018, Medicare pays for TKA in the outpatient setting and CJR participant hospitals shifted a lower share of TKA to the hospital outpatient setting than control hospitals.[36] While CJR and control respondents had generally similar demographic and health characteristics in our overall analytic sample (Exhibit C-17 and C-18), we cannot completely exclude the possibility that differences in unobserved characteristics between the CJR and control respondents affected our results.

Although the availability of outpatient TKA does not influence the treatment of patients with hip fracture, we observed notable differences in pre-hospital functional status corresponding with two of the three measures for which we estimated significantly different changes in functional status between CJR and control respondents. We risk-adjusted estimates to control for differences in pre-hospital functional status, but there could be differences in unobserved patient characteristics that were not captured in our models (such as health conditions not included in the HCC score) that affected results for hip fracture patients.

[36] Centers for Medicare & Medicaid Services. Comprehensive care for joint replacement model - third annual report. https://innovation.cms.gov/data-and-reports/2020/cjr-thirdannrpt. 2020: 31-37.

VII. Impact of the CJR Model on Total Market Volume of Elective LEJR Discharges

We analyzed the impact of the CJR model on the volume of LEJR discharges in a market by testing whether MSAs selected to participate in the CJR model experienced larger or smaller increases in the LEJR discharge rate (discharges per 1,000 FFS population) than they would have otherwise.

We ran separate analyses for elective and fracture LEJR discharges, because CJR participant hospitals have more influence over elective episode volume than fracture episode volume.

A. Market definition

Markets were defined by the MSAs used in the design of the CJR model. For this analysis, we focused on MSAs that continued mandatory participation in PY3 and PY4 (n=34) and their respective control group MSAs (n=45). Further, we split very large MSAs into smaller metropolitan divisions following the methodology of the geographic payment adjustment used in the IPPS.[37]

B. Time periods

The analysis was at the market-quarter level and covered October 2007 to December 2019. We included indicators for the baseline period, the interim period, and two CJR intervention periods.

- **The CJR baseline period (October 2007 – June 2015)** begins the date the hospital IPPS switched to the MS-DRG system (the LEJR episode is defined by MS-DRG 469 and MS-DRG 470) and ends prior to the announcement of the Bundled Payments for Care Improvement initiative.
- **CJR interim period (July 2015 – March 2016)** begins the date that the CJR model was announced (July 2015) and ends the day before the model was implemented (March 31, 2016).
- **CJR PY1-PY2 (April 2016 – December 2017)** begins the date that the CJR model took effect (April 1, 2016) and ends with the end of PY2 (December 31, 2017).
- **CJR PY3-PY4 (January – December 2019)** begins the date that new changes to the CJR model were implemented (January 1, 2018) and ends with the end of PY4 (December 31, 2019).

C. Discharges per 1,000 FFS population

The discharge rate was calculated as the number of LEJR discharges per 1,000 FFS population in a given quarter. LEJR discharges included: 1) hospital inpatient LEJRs discharged under MS-DRG 469 or 470 in Part A IPPS claims; and 2) hospital outpatient total knee arthroplasty (TKA)

[37] Large MSAs that are split into smaller metropolitan divisions are Chicago, Dallas, Detroit, Los Angeles, Miami, New York, and Philadelphia resulting in 34 CJR MSAs represented by 40 markets and 45 control group MSAs represented by 53 markets.

procedures in calendar years 2018 and 2019, identified using CPT code 27447 in Part B institutional claims.[38] Counts of the Medicare FFS population within each quarter of the year were obtained from Medicare enrollment data.

D. Measures of CJR and Bundled Payments for Care Improvement "dose"

We analyzed the impact of the CJR model on the volume of elective LEJR discharges in a market by estimating the relationship between CJR "dose" and the change in the elective LEJR discharge rate (discharges per 1,000 Medicare FFS beneficiaries) in MSAs. CJR "dose" was the market share of LEJR discharges[39] for hospitals that ever participated in the CJR model (i.e., the number of LEJR discharges from hospitals that ever participated in the CJR model divided by total LEJR discharges in the market). Similarly, we measured Bundled Payments for Care Improvement "dose" as the market share of LEJR discharges for providers (hospitals, PGPs, SNFs, and HHAs) that were ever in the risk-bearing phase of the Bundled Payments for Care Improvement initiative for Models 2 through 4 for the LEJR clinical episode. The market share was calculated using the three-year period prior to the first Bundled Payments for Care Improvement intervention time period (October 2009 through September 2012). We measured market share using this period since market share in the intervention periods of the Bundled Payments for Care Improvement initiative and CJR model is endogenous to the model.

In the first four years of the model, there were 39 CJR mandatory hospitals located across nine markets that were formerly Bundled Payments for Care Improvement LEJR participants, and therefore their baseline market shares were included in both the CJR dose and the Bundled Payments for Care Improvement dose potentially overstating bundled payment penetration in these markets. If we did find a significant impact of the CJR model on LEJR discharge rates, then it could be difficult to disentangle whether the effect is due to the CJR model or other bundled payment models in the markets.

E. Statistical model

The impact of the CJR model and the Bundled Payments for Care Improvement initiative on LEJR volume was estimated using an OLS regression model, which incorporated market fixed effects, time fixed effects, and market-specific linear time trends:

$$[1]\quad V_{it} = b_0 + b_{1i} + b_{2t} + b_{3it} + (b_4 \cdot Z_{it}) + (b_5 \cdot CJRdose_i \cdot CJRInterim_1) + (b_6 \cdot CJRdose_i \cdot CJRPY1 - PY2_2) + (b_7 \cdot CJRdose_i \cdot CJRPY3 - PY4_3) + (b_8 \cdot BPCIdose_i \cdot BPCIPost_1) + (b_9 \cdot BPCIdose_i \cdot BPCIPost_2) + (b_{10} \cdot BPCIdose_i \cdot BPCIPost_3) + \varepsilon_{it}$$

Where:

- V_{it} is the LEJR discharge rate (the number of LEJR discharges per 1,000 FFS beneficiaries) in market *i* and quarter *t*;

[38] Outpatient TKA was removed from the inpatient only list starting January 1, 2018.

[39] The number of discharges can be slightly greater than the number of episodes due to the exclusion criteria applied during the episode creation algorithm.

- b_{1i} allows for market fixed effects that control for market-specific factors that are constant across time;
- b_{2t} allows for time fixed effects (measured in quarters) that control for time-specific factors that are common across markets;
- b_{3it} allows for markets to follow different linear time trends;
- Z_{it} controls for characteristics of the FFS population residing in market *i* in quarter *t* (age, sex, dual eligibility, disabled/not ESRD), the share of the Medicare beneficiary population enrolled in Medicare Advantage, and the share of the Medicare FFS beneficiary population aligned with ACOs;
- $CJRdose_i$ is the market share of ever-CJR participants in market *i* measured over a portion of the baseline time period (share of market discharges initiated by ever-CJR participant hospitals from October 2009 – September 2012).
- $CJRInterim_1$, $CJRPY1 – PY2_2$, *and* $CJRPY3 – PY4_3$ equal 1 during the interim period and each CJR intervention period, respectively (July 2015 – March 2016, April 2016 – December 2017, and January 2018 – December 2019)
- $BPCIdose_i$ is the market share of participants that ever participated in the Bundled Payments for Care Improvement initiative in market *i* measured over a portion of the baseline period (share of market discharges initiated by participants that ever participated in the Bundled Payments for Care Improvement initiative from October 2009 – September 2012);
- $BPCIPost_1$, $BPCIPost_2$, *and* $BPCIPost_3$ equal 1 during each Bundled Payments for Care Improvement intervention period, respectively (October 2012 – September 2013, October 2013 – September 2015, and October 2015 – September 2018).[40]

The impact of the CJR model on LEJR volume was captured by coefficients b_5, b_6, and b_7, which measured the average change in the LEJR discharge rate due to the CJR dose. The impact of the Bundled Payments for Care Improvement initiative on LEJR volume was captured by coefficients b_8, b_9, and b_{10}, which measured the average change in the LEJR discharge rate due to the Bundled Payments for Care Improvement dose as measured by Bundled Payments for Care Improvement market shares during the CJR baseline.

Standard errors were clustered at the market level to account for non-independence of observations within markets. We weighted the regression by the FFS beneficiary population in the market and the inverse probability of selection into the CJR model.[41] Finally, we tested

[40] *BPCI Post 1* is the Bundled Payments for Care Improvement initiative intervention period in which no awardees were in the risk-bearing phase of the initiative. *BPCI Post 2* is the Bundled Payments for Care Improvement initiative intervention period in which some awardees were in the risk-bearing phase of the initiative, some had not yet joined the initiative, and some had terminated participation. *BPCI Post 3* is the Bundled Payments for Care Improvement initiative intervention period in which all awardees were either in the risk-bearing phase of the initiative or had terminated participation.

[41] Angrist, Joshua D., and Jörn- Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press.

whether the CJR and control group discharge rates were significantly different at the CJR PY1/ PY2 and CJR PY3/PY4 time periods.

We included market-specific linear trends in our model. In application, there might be concern that market trends soak up the treatment effect if treatment assignment is correlated with market trend shifts, but CJR's initial randomized controlled trial design mitigates the concern for this. Furthermore, to investigate whether the inclusion of market trends was warranted, we used the Post-Double Selection LASSO (PDSLASSO)[42] technique for variable selection. PDSLASSO is a data driven method for model selection that takes into account both prediction and inference when selecting variables. When we used PDSLASSO for elective procedures, 69 out of the 92 (75%) market-level trends were selected as important to the model for prediction and inference. There is no theoretical justification for including only some market trends and not others, so we included them all.

F. Limitations

A limitation of our analysis was that the measurement of CJR and Bundled Payments for Care Improvement "dose" did not vary based on the duration of participation within the market, nor did it vary as hospitals switched from Bundled Payments for Care Improvement to CJR participation. In all MSAs, a market was assigned the same Bundled Payments for Care Improvement dose from a given provider whether the provider had yet to participate, dropped out, or continued to participate through the end of the initiative. In CJR-eligible MSAs, each dose included market share from Bundled Payments for Care Improvement and CJR participant hospitals even if they switched from Bundled Payments for Care Improvement to CJR participation during the intervention. This methodology can overestimate the Bundled Payments for Care Improvement dose in both control and CJR MSAs, and can lead to overlap between the two doses, either of which would bias estimates toward zero. However, constructing the measures in this way was necessary so that the doses were not endogenous to performance under the CJR and Bundled Payments for Care Improvement initiatives.

[42] Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 29-50.

VIII. Factors Associated with Receiving Reconciliation Payments under the CJR Model

We identified hospital and patient characteristics associated with the average reconciliation payment per episode. This analysis controlled for potential confounders (i.e. other variables that may be related to the characteristics and reconciliation payments).

A. Sample

Mandatory CJR hospitals were included if they were located in the 34 MSAs required to continue participation in the CJR model in PY3 and PY4. We excluded hospitals with less than 20 episodes in the PY to improve reliability of results. The threshold of 20 episodes was selected to be consistent with the minimum threshold used by the CMS CJR payment contractor to set quality-adjusted target prices. The sample included 244 hospitals with NPRA in PY1, 292 hospitals in PY2, 279 hospitals in PY3, and 313 hospitals in PY4.

In addition, we ran the regression model for opt-in hospitals located in the 33 voluntary MSAs. Similarly, we excluded hospitals with less than 20 episodes in the PY to improve reliability of results. The sample included 61 hospitals with NPRA in PY1, 71 hospitals in PY2, 73 hospitals in PY3, and 73 hospitals in PY4. Results from this secondary analysis are included in Appendix L.

B. Reconciliation payment per episode

The average reconciliation payment per episode was calculated dividing annual hospital reconciliation or repayment amount by the overall number of episodes. A positive value per episode indicates the hospital earned a reconciliation payment in the PY, while a negative value per episode indicates the hospital was required to repay CMS.[43]

C. Statistical model

The analysis was conducted at the hospital-performance year level. The relationship between the average reconciliation payment per episode and hospital and patient characteristics was estimated using an OLS regression model. These covariates were selected because they were identified as correlated with average reconciliation payment per episode in bivariate analyses and were also included in our risk-adjusted episode-level DiD models. The regression model took into account repeated observations (i.e., multiple observations or PYs per hospital) and clustering of hospitals within CJR MSAs. Results were considered statistically significant at $p<0.10$.

[2] $$NPRA_{h,t} = b_0 + b_1 \cdot Hopital_h + b_2 \cdot Patient_{ht} + b_3 \cdot Model_{ht} + \varepsilon_{it}$$

Where:

- Hospital covariates measured at baseline: Census region, bed count, ownership, Disproportionate Share Hospital (DSH) patient percentage, any affiliation with a

[43] Hospitals were not required by CMS to make repayments in PY1; we estimated potential repayment amounts for PY1 and included them in our analysis for consistency across PYs.

medical school, and ever participated in Bundled Payments for Care Improvement initiative for LEJR.

- Patient covariates measured in the PY: Average HCC score for patients and percent of the hospital's episodes that were: female, age 80 years or older, non-Hispanic Black or African American, MS-DRG 470 elective, dual eligible for Medicaid, disabled (not ESRD), and with an institutional PAC stay in the six months prior to the LEJR.
- Model-specific covariates measured in the PY: Hospital quality performance, average quarterly volume, and relationship between hospital historical average payments and PY quality-adjusted target price.

Median values were used to create binary variables of the continuous covariates, so we were able to compare financial performance for hospitals in the lower half of the distribution of the covariate to hospitals in the top half of the distribution (Exhibit C-20).

Exhibit C-20: Median values used to create binary versions of the continuous covariates for the average reconciliation payment per episode regression analysis

Covariate	Median value for mandatory CJR hospitals	Median value for voluntary opt-in CJR hospitals
Bed count	258	188
DSH patient percentage	24.2%	21.8%
Percent of episodes MS-DRG 470 elective	81.4%	89.0%
Percent of episodes female	65.6%	64.3%
Percent of episodes 80 years or older	26.9%	21.9%
Percent of episodes non-Hispanic Black or African American	3.9%	2.2%
Average HCC score	1.60	1.31
Percent of episodes dual eligible	11.2%	9.0%
Percent of episodes disabled (no ESRD)	15.1%	13.9%
Percent of episodes with a prior institutional PAC stay	4.9%	3.1%

Source: CJR evaluation team analysis of FY 2016 CMS Annual IPPS and Medicare claims and enrollment and quality-adjusted target price data for mandatory CJR participant hospital(s) in PY1 (episodes starting on or after April 2016 and ending on or before December 2016), PY2 (episodes ending between January and December 2017), PY3 (episodes ending between January and December 2018), and PY4 (episodes ending between January and December 2019).

***Note*:** DSH = disproportionate share hospital, ESRD = end stage renal disease, FY = fiscal year, HCC = hierarchical condition category, IPPS = Inpatient Prospective Payment System, MS-DRG = Medicare Severity Diagnosis Related Group, PAC = post-acute care, PY = performance year.

D. Limitations

The analysis examined PY1, PY2, PY3, and PY4 NPRA. PY1, PY2, and PY3 NPRA are considered final as of the writing of this report, while the PY4 results are preliminary and subject to change when they are finalized in spring 2021. We used average reconciliation payment per episode instead of total amount because the total reconciliation or repayment amount is highly driven by hospital LEJR volume. Finally, we excluded hospitals with very low volume (less than 20 episodes in the year), and as a result, these hospitals are not represented in the analysis. Results may not be generalizable to low volume providers participating in the CJR model. However, we ran a sensitivity test that included these low volume hospitals and results were generally consistent.

IX. Patient Selection/Patient Complexity Measure

A. Analyses of a composite measure of patient complexity

As multiple patient characteristics are related to patient complexity, we used total episode spending as a composite measure of patient complexity. The use of a composite measure allows us to better understand the relationship between the CJR model and patient complexity as a whole. We estimated how much of the relative change in total payments experienced over the intervention period was attributable to relative changes in patient mix using the Oaxaca decomposition method. Kröger and Hartmann (2020) developed and described the approach in detail.[44]

First, we conducted an OLS regression to estimate the relative difference in total payments between mandatory CJR and control hospitals over the intervention period, risk adjusting for hospital- and market-level covariates. This model did not risk adjust for patient-level covariates, as those variables were used later to analyze the impact of changes in patient mix. As such, the estimated relative difference captured both the impact of the model on total payments, similar to that captured in our total payments DiD approach, as well as the impact of any effects resulting in relative changes in patient characteristics across CJR hospitals and control hospitals.

Next, we decomposed the relative change in total payments from the baseline to the intervention period into separate impacts, each of which are a different type of effect that contributed to the overall relative changes in total payments. The impact of interest for this analysis was a bundle of patient characteristics.[45] It informed us of the degree to which the relative change in total payments resulted from relative changes in patient characteristics. The method allowed us to "turn off" any changes between CJR hospitals and control hospitals that were not due to changes in patient mix and isolate the effect of changes in patient mix on relative changes in total payments. The estimated impact of changes in patient mix was interpreted in per-episode units of total episode spending. We also reported this result as a percent of the total relative change (while only controlling for hospital and market covariates). Standard errors were clustered at the MSA level.

The Oaxaca decomposition designed for panel data followed the empirical strategy of the standard DiD approach, but in our application had a few notable differences. In our DiD analyses, we included patient characteristics as risk-adjusting covariates, which caused relative changes in patient characteristics to not influence the DiD impact estimate. Alternatively, the Oaxaca decomposition analyses did not include patient characteristics and any effect found represented entirely separate ways in which the CJR model impacted total payments, and thus should not be considered part of the DiD impact estimate. Likewise, as the Oaxaca decompositions found the effects of relative changes in patient characteristics on relative changes in total payments, independent of other changes, the DiD impact estimate should not be considered part of the

[44] Kröger, Hannes, and Jörg Hartmann. 2020. "Xtoaxaca - Extending the Kitagawa-oaxaca-blinder Decomposition Approach to Panel Data." SocArXiv. February 1. doi:10.31235/osf.io/egj79.

[45] Specifically, the following patient characteristics: HCC score, age, sex, race/ethnicity, disability status at Medicare enrollment (not ESRD), Medicaid eligibility status, obesity, hypertension, tobacco use, and prior utilization measures.

estimates produced by the Oaxaca decompositions. Instead, each estimate contributed additional and unique information, and when taken together, provided a more complete picture of the various ways in which the CJR model influenced relative changes in total payments.

Because hospitals have different quality-adjusted target prices by fracture status and MS-DRG, we performed this analysis separately for elective MS-DRG 470, elective MS-DRG 469, fracture MS-DRG 470, and fracture MS-DRG 469 episode groups. In order to gain a more complete understanding of the relationship between the CJR model and changes in patient mix, we ran our analyses using only inpatient LEJR episodes included in the CJR model.

We performed a sensitivity analysis accounting for the differential outpatient TKA rates, because outpatient TKA patterns can affect the composition of the inpatient LEJR patient population. Accounting for the differential outpatient TKA rates, changes in the patient population resulted in a \$137 ($p<0.01$) relative decrease in average CJR episode payments for elective MS-DRG 470 episodes. In other words, the differential outpatient TKA rates explained about \$33 of the \$ \$170 ($p<0.01$) per-episode decrease due to changes in patient mix for inpatient elective MS-DRG 470 episodes. Therefore, we conclude that differential outpatient TKA rates do not explain this reduction.

B. Analyses of patient characteristics

For our univariate analysis, we estimated DiD regressions[46] on various beneficiary characteristics. We used a variety of regression models, including logistic, multinomial logit, and OLS regressions, chosen to account for the data properties of each characteristic. This analysis did not contain risk-adjusting covariates, since the dependent variables are beneficiary characteristics that are included in our risk adjustment. Standard errors were clustered at the MSA level.

[46] See Section III.C for additional details about our DiD design.

X. Qualitative and Mixed Methods Analysis

A. Data collection

1. Provider telephone interviews

One interviewer and one note taker conducted the telephone interviews. Notes were taken during telephone interviews and, if the interviewee agreed, the interview was recorded. Recordings were used to verify and enhance interview notes. Notes from telephone interviews were organized and entered into ATLAS.ti software (version 8; Scientific Software Development GmbH, Berlin, Germany) for coding and analysis.

2. Orthopedic surgeon survey

We conducted cognitive interviews with a convenience sample of orthopedic surgeons to test clarity and wording of survey questions and response options. Twenty-four surgeons were contacted mostly via hospital administrators. We conducted five cognitive interviews with surgeons in a variety of geographic locations and practice settings (e.g., employment by physician practice versus employment by hospital), and interviewees were compensated for their time. We attempted to select surgeons who varied in the volume of LEJR surgeries performed on Medicare fee-for-service patients, the types of LEJR surgeries they perform (knee, elective hip, hip fracture), and the region of the country where they practice. The survey was adjusted in response to cognitive interviewee feedback, and was designed to be completed in about 15 minutes.

The survey was fielded for eleven weeks from to August 14 to October 31, 2019. The survey protocol was customized depending on availability of contact information. We used a multi-mode survey approach utilizing mail, phone, and web outreach and completion strategies. For the 88% of surgeons with email addresses, we initially sent an email with a link to an online survey designed for completion via computer or smartphone. Non-respondents were reminded of the request for survey completion via email, post-card, telephone calls, or with a mailed version of the survey (with return envelope) roughly every week during the fielding period. CMS also sent a final survey reminder email prior to the end of the survey administration period.

For the 12% of surgeons without email addresses in the IQVIA data, we initially sent a mailed survey packet with a cover letter on CMS letterhead and a postage paid return envelope. The letter included a URL and user ID for those who preferred to complete the survey online. Non-respondents were reminded of the request for survey completion via a second mailed survey packet, as well as postcard and telephone call reminders.

Outreach efforts included an email address and toll-free number for respondents to submit questions throughout the fielding period. A $100 incentive was offered to surgeons for completion of the survey. The response rate by outreach protocol is included in Exhibit C-21.

Exhibit C-21. Orthopedic surgeon survey response rate by outreach protocol

Status	Protocol	Response	N	%
Respondents	**Email + mail**	Web response	162	65.1%
		Telephone response	1	0.4%
		Paper response	58	23.3%
	Mail-only	Web response	14	5.6%
		Paper response	14	5.6%
		Total, respondents	249	
Non-respondents	**Email + mail**	No response	471	76.3%
		Refusal	51	8.3%
		Incomplete response	4	0.6%
		No longer with organization	14	2.3%
	Mail-only	No response	70	11.3%
		Unreachable	5	0.8%
		No longer with organization	2	0.3%
		Total, non-respondents	617	
		Sample total	866	

Source: CJR evaluation orthopedic surgeon survey, fielded between August and October 2019.

B. Analysis

1. Provider telephone interviews

We developed analytic codebooks including primary and sub-codes based on the telephone interview protocols. Coders used ATLAS.ti to apply codes and sub-codes to comprehensive interview notes, and ran queries to identify themes across interviews. All coders received systematic training, which included parallel coding and discussion of results with trainers until consistency was established. Throughout the analysis the codebooks were refined (i.e., codes were dropped, consolidated, added, or revised) to better capture patterns as they emerged.

2. Orthopedic surgeon survey

For survey measures, we calculated frequencies and summary statistics for all of the close-ended questions included in the survey. For open-text items, we reviewed responses and identified common themes. We completed a comparison of the characteristics of survey respondents and non-respondents using measures from claims data, IQVIA, and 2019 CJR Financial Arrangement Clinician Engagement (FACE) data provided by CMS. A descriptive analysis was completed for surgeons' responses to the survey questions, weighted to adjust for survey non-response. Reported percentages may not sum to 100% due to non-response.

a. Respondent characteristics

Of the 866 sampled surgeons, 249 surgeons responded to the survey (29% response rate), though the response rate by question varied. Over half of respondents (55%) indicated that they had over 20 years of experience as an orthopedic surgeon, excluding their training. About one quarter (27%)

of respondents had 11-20 years of experience, and 18% had 3-11 years of experience. Half (50%) of respondents reported that they did more than 150 LEJR procedures in the prior 12 months. Over half of respondents (53%) indicated they performed LEJR procedures at one hospital; 34% performed these procedures at two hospitals; 13% of surgeons performed the procedures at three or more hospitals. Nearly two-thirds (62%) of respondents reported they worked in a physician-owned practice. Almost equivalent proportions of respondents were either hospital employees or in a hospital or health-system owned practice (11% and 14%, respectively). Smaller percentages of respondents (6% and 7%) were in an academic department or practice or were independent contractors. Over half of respondents (52%) reported previously participating in a value-based payment model, including a Medicare ACO (25%), the Bundled Payments for Care Improvement initiative or Bundled Payments for Care Improvement Advanced initiatives (35%), or a model run by a commercial payer (21%).

b. Non-response analysis and weights

A non-response analysis was conducted to assess the generalizability of the surgeon survey respondents to the overall surgeon survey sample. Statistical significance of factors associated with survey response was determined using Pearson's chi-squared tests. Characteristics of respondents were proportionally similar to non-respondents regarding gender, age, number of CJR episodes, number of CJR hip replacement episodes, number of CJR episodes other than for THA or TKA, and number of hospitals at which they completed CJR episodes.

We adjusted for the following identified differences in our analyses by applying non-response weights. To calculate non-response weights, we used logistic regression to estimate the probability of responding to the survey. We then calculated the non-response weights as the inverse probability of responding to the survey.

- The distribution of survey respondents and non-respondents differed by the number of CJR knee replacement episodes completed ($p<0.10$). Respondents performed a higher volume of CJR knee replacements episodes (21-50 or >50 episodes) than non-respondents.
- The distribution of survey respondents and non-respondents differed by their employment setting ($p<0.01$). Respondents were more often employed by a physician practice or medical group.
- The distribution of respondents and non-respondents differed by the MSA in which they performed LEJR ($p=<0.05$).
- The distribution of survey respondents and non-respondents differed by the presence of a hospital-reported CJR model gainsharing agreement ($p=<0.05$) as noted in the FACE data. Respondents more often had a hospital-reported gainsharing arrangement

3. Limitations

The analysis of the telephone interview data provides a description of themes and patterns in response to the protocols, which may not include the full experience of PAC providers. Our sample

of SNFs and PTs was limited to 72 total interviews and may not be representative of all post-acute providers that received patients from CJR participating hospitals.

The response rate of the surgeon survey was 29%. Although responding surgeons were similar to non-respondents on most characteristics, and although we used non-response weights, respondents may differ from non-respondents in unobserved ways.

There is the possibility of recall bias, as survey respondents were asked to reflect solely about their Medicare fee-for-service patients when providing answers to survey questions, respondents practicing at multiple hospitals were asked to respond for the hospital where they work most, and respondents were also asked to reflect on the first post-operative appointment for their patients following TKA for information about short-term outcomes. Social desirability bias and recency bias may lead responding surgeons to provide more positive reports than their actual experience, or to forget changes implemented over the prior three years from survey response.

The survey did not include a comparison group; surgeons performing LEJR outside of CJR mandatory MSAs may have experienced similar changes over the past three years. Lastly, the study design was a cross-sectional survey and thus analyses conducted for the surgeon survey cannot inform statements about cause and effect.

C. Clinical Review Panel

Information obtained from ten Clinical Review Panels provided clinician insights into the impact of the CJR model on payments, utilization or patterns of care, and quality of care identified through quantitative data analysis or qualitative findings. Specific panel topics are listed in Exhibit C-22. Six panelists of various backgrounds and expertise participated in Clinical Review Panels. Panelists were identified through professional contacts and vetted by CMS. Specifically, the panel was comprised of a private practice orthopedic surgeon, an academic orthopedic surgeon, a physical therapist with home health expertise, a gerontological nurse practitioner, an academic nurse with care transition expertise, and a geriatrician with SNF expertise.

The objectives of the Clinical Review Panels were to:

- Review and comment on changes in patterns of care and quality outcomes identified in quarterly reports.
- Report on changes in clinical practice that may affect the CJR model.
- Present medical or provider community feedback on the CJR model.
- Raise questions for possible further analysis.
- Corroborate qualitative findings.
- Provide additional insight into utilization and quality patterns we might expect given the incentives of the program.
- Identify changes in practice patterns that may differentially impact subpopulations of Medicare patients.

- Aid in the identification of promising practices and unintended consequences.
- Assist in the detection of the CJR model's overlap with other Center for Medicare and Medicaid Innovation models and demonstrations.

All Clinical Review Panels were administered in the same manner and convened via webinar. Panelists received CMS-approved packet to review prior to each webinar. This packet consisted of relevant CJR model background information, an agenda, general expectations for the Clinical Review Panel, and presentation slides that included evaluation results and the probing questions for discussion. Dr. Christine LaRocca, a geriatric medicine physician and medical director at Telligen, led a discussion structured on questions based on the evaluation results to date. Each question was discussed and all participants were given an opportunity to answer. The meetings were recorded and transcribed to ensure accurate records of the discussions. Key takeaways from each Clinical Review Panel were used to inform future analyses and interpretations of results.

Exhibit C-22: Clinical Review Panel topics

Panel	Topics
1. July 2017	Introduction to the CJR model and Clinical Review Panel responsibilities
2. October 2017	Early findings from claims-based analysis and qualitative data
3. January 2018	Claims- and assessment-based findings for elective episodes
4. May 2018	Claims- and assessment-based findings for fracture episodes
5. August 2018	Selected qualitative findings: rehabilitation and discharge planning
6. January 2019	Key patient reported outcomes reported through patient surveys, and insights related to a provider survey
7. May 2019	Potential unintended consequences of the CJR model
8. September 2019	Removal of elective TKA from the Medicare inpatient only list and anesthesia practices
9. February 2020	Changes in clinical practice for inpatient and outpatient LEJR, medical and/or provider community feedback about the CJR model, and the impact of the October 2019 SNF payment changes (PDPM)
10. July 2020	CJR in the COVID-19 environment

***Note*:** COVID-19 = coronavirus disease 2019, LEJR = lower extremity joint replacement, PDPM = Patient Driven Payment Model, SNF = skilled nursing facility, TKA = total knee arthroplasty.

Appendix D: Payment, Utilization, Quality, and Activities of Daily Living Results

Mandatory CJR hospitals

Exhibit D-1: Risk-adjusted claims-based difference-in-differences results for payment, utilization, and quality metrics, mandatory CJR hospitals, LEJR episodes, PY1-4

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
Payments	Episode payments	153,813	179,291	$29,192	$26,379	$28,665	$27,363	-$1,511	-5.2%	p<0.01	[-$2,113 to -$909]
	IRF payments	153,813	179,291	$2,237	$1,199	$2,164	$1,720	-$593	-26.5%	p<0.01	[-$920 to -$267]
	SNF payments	153,813	179,291	$6,142	$4,212	$6,192	$5,105	-$843	-13.7%	p<0.01	[-$1,205 to -$481]
	HH payments[a]	153,813	179,291	$2,415	$2,477	$2,317	$2,314	$65	2.7%	p=0.61	[-$144 to $274]
	Readmission payments	153,813	179,291	$1,225	$1,056	$1,104	$1,088	-$153	-12.5%	p<0.05	[-$277 to -$29]
	Part B payments	153,813	179,291	$4,994	$4,895	$4,822	$4,795	-$72	-1.4%	p=0.22	[-$169 to $25]
	30-day PEP payments[a]	153,813	179,291	$1,483	$1,505	$1,489	$1,544	-$32	-2.2%	p=0.18	[-$72 to $7]
	Anchor payments	153,813	179,291	$12,190	$12,159	$12,194	$12,096	$67	0.6%	p=0.13	[-$6 to $141]
Utilization	First PAC IRF	153,813	179,291	13.8%	5.9%	13.2%	9.2%	-3.9	-28.1%	p<0.01	[-6.1 to -1.6]
	First PAC SNF	153,813	179,291	41.3%	29.4%	41.8%	32.5%	-2.7	-6.5%	p=0.103	[-5.4 to 0.0]
	First PAC HH	153,813	179,291	36.6%	49.4%	33.5%	38.8%	7.5	20.5%	p<0.05	[1.8 to 13.2]
	First PAC home without HH	153,813	179,291	8.3%	15.3%	11.5%	19.5%	-0.9	-11.3%	p=0.69	[-4.8 to 2.9]
	Any HH use[a]	153,813	179,291	72.1%	73.2%	69.0%	65.8%	4.3	6.0%	p=0.25	[-1.8 to 10.4]
	IRF days	10,193	14,375	11.7	11.8	11.6	11.8	-0.1	-0.9%	p=0.58	[-0.4 to 0.2]
	SNF days	47,589	49,180	27.1	22.1	27.2	24.8	-2.6	-9.5%	p<0.01	[-3.5 to -1.7]
	HH visits	111,846	119,003	16.9	15.6	16.7	16.3	-1.0	-6.1%	p<0.05	[-1.7 to -0.3]
	HH PT/OT visits	111,846	119,003	10.4	10.3	10.3	10.8	-0.6	-5.9%	p<0.10	[-1.2 to 0.0]
	Outpatient PT/OT visits[a]	105,763	127,094	13.0	14.7	13.2	14.5	0.4	3.2%	p=0.14	[-0.1 to 0.9]

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
Quality	Unplanned readmission rate	153,767	179,256	9.1%	8.7%	8.9%	8.8%	-0.3	-3.5%	p<0.10	[-0.6 to 0.0]
	ED use	153,767	179,256	13.1%	14.1%	12.7%	13.6%	0.1	1.0%	p=0.65	[-0.3 to 0.6]
	Mortality rate	157,040	175,587	2.6%	2.5%	2.7%	2.5%	0.0	1.0%	p=0.80	[-0.1 to 0.2]
	Complications[b]	131,738	158,862	3.3%	2.5%	3.0%	2.6%	-0.3	-7.9%	p<0.05	[-0.4 to -0.1]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the DiD estimate as a percent of the CJR baseline level.

The change in separate provider payments do not sum to the change in episode payments because separate models were estimated for episode payments and each component payment.

Because CJR participant hospitals shifted a lower share of TKAs to the hospital outpatient setting, the control group includes outpatient TKA episodes that have been weighted to balance the episode volume in the CJR hospitals.

CI = confidence interval, DiD = difference-in-differences, ED = emergency department, HH = home health, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, OT = occupational therapy, PAC = post-acute care, PEP = post-episode period, PT = physical therapy, PY = performance year, SNF = skilled nursing facility, TKA = total knee arthroplasty.

[a] We cannot be certain that there is no impact of the model because this outcome failed parallel trends tests (Appendix K). Parallel trends is an assumption that underlies our methodological approach. Please see Appendix C (Section III.C.1.c) for additional details.

[b] The complications measure only applies to elective episodes.

Exhibit D-2: Risk-adjusted assessment-based difference-in-differences results for activities of daily living metrics, mandatory CJR hospitals, LEJR episodes, April 2016-September 2019

First PAC setting	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
IRF	Average change in mobility score	7,520	10,598	10.5	11.0	10.0	10.8	-0.2	-2.0%	p=0.41	[-0.6 to 0.2]
SNF	Improved transfer, locomotion on unit, and walking in corridor	32,937	33,814	65.4%	67.4%	69.6%	71.0%	0.6	1.0%	p=0.64	[-1.6 to 2.9]
	Improved toilet use	32,971	33,733	42.7%	39.8%	46.3%	46.7%	-3.3	-7.8%	p=0.12	[-6.8 to 0.2]
	Without self-reported pain[a,b]	31,678	32,249	58.3%	76.0%	53.2%	67.0%	4.0	6.8%	p<0.01	[1.5 to 6.4]
HHA	Improved ambulation/ locomotion	55,989	52,908	90.2%	91.2%	90.3%	91.2%	0.2	0.2%	p=0.75	[-0.7 to 1.0]
	Improved bed transferring	55,755	52,659	85.1%	86.5%	84.8%	86.6%	-0.4	-0.5%	p=0.59	[-1.6 to 0.8]
	Reduced pain	55,731	52,717	75.1%	84.1%	74.7%	83.2%	0.5	0.6%	p=0.70	[-1.6 to 2.5]

Source: CJR evaluation team analysis of Medicare claims and enrollment data, MDS data, OASIS data, and IRF-PAI data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by September 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the DiD estimate as a percent of the CJR baseline level.

CI = confidence interval, DiD = difference-in-differences, HHA = home health agency, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, SNF = skilled nursing facility.

[a] We cannot be certain that this result is an impact of the model, because this outcome failed parallel trends tests (Appendix K). Parallel trends is an assumption that underlies our methodological approach. Please see Appendix C (Section III.C.1.c) for additional details.

[b] The pain measure for those initially discharged to a SNF was not risk adjusted following the specifications of the MDS 3.0 Quality Measure for short-stay patients used in the CMS Nursing Home Five-Star Rating System.

Opt-in hospitals

Exhibit D-3: Risk-adjusted claims-based difference-in-differences results for payment, utilization, and quality metrics, opt-in CJR hospitals, LEJR episodes, PY1-4

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
Payments	Episode payments	52,813	59,179	$23,451	$21,525	$23,751	$22,577	-$752	-3.2%	p<0.01	[-$1,107 to -$397]
	IRF payments	52,813	59,177	$398	$202	$924	$719	$9	2.4%	p=0.93	[-$179 to $198]
	SNF payments	52,813	59,179	$4,338	$2,638	$4,523	$3,279	-$456	-10.5%	p<0.05	[-$788 to -$124]
	HH payments	52,813	59,179	$1,739	$1,551	$1,508	$1,385	-$65	-3.7%	p=0.44	[-$203 to $73]
	Readmission payments	52,813	59,179	$785	$784	$746	$729	$17	2.2%	p=0.63	[-$42 to $76]
	Part B payments	52,813	59,179	$4,099	$4,150	$4,242	$4,274	$20	0.5%	p=0.69	[-$62 to $102]
	30-day PEP payments	52,810	59,166	$993	$1,073	$1,021	$1,074	$27	2.8%	p=0.37	[-$23 to $78]
	Anchor payments	52,813	59,179	$12,060	$12,017	$12,167	$12,230	-$106	-0.9%	p=0.15	[-$227 to $15]
Utilization	First PAC IRF	52,813	59,179	2.3%	0.8%	6.1%	3.9%	0.7	28.2%	p=0.50	[-0.9 to 2.3]
	First PAC SNF	52,813	59,179	35.8%	21.9%	33.4%	21.6%	-2.1	-5.8%	p=0.25	[-5.0 to 0.9]
	First PAC HH	52,813	59,179	37.7%	36.9%	26.9%	27.6%	-1.5	-4.0%	p=0.56	[-5.8 to 2.7]
	First PAC home without HH	52,813	59,179	24.2%	40.4%	33.6%	46.9%	2.9	12.1%	p=0.24	[-1.2 to 7.0]
	Any HH use	52,813	59,179	59.0%	51.4%	48.6%	41.8%	-0.9	-1.5%	p=0.75	[-5.3 to 3.6]
	IRF days	556	1,346	11.6	11.8	12.6	12.3	0.5	4.5%	p=0.20	[-0.2 to 1.2]
	SNF days	12,001	12,623	23.3	18.9	22.7	21.2	-2.8	-12.2%	p<0.01	[-3.9 to -1.8]
	HH visits	26,925	26,166	13.4	12.8	14.5	14.8	-0.9	-6.6%	p<0.05	[-1.5 to -0.3]
	HH PT/OT visits	26,925	26,166	8.9	9.1	9.6	10.0	-0.2	-2.4%	p=0.47	[-0.7 to 0.3]
	Outpatient PT/OT visits	37,709	41,729	11.8	13.0	12.3	13.2	0.2	2.1%	p=0.13	[0.0 to 0.5]

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
Quality	Unplanned readmission rate	52,804	59,166	6.2%	6.3%	6.4%	6.3%	0.1	1.6%	p=0.67	[-0.3 to 0.5]
	ED use	52,804	59,166	13.6%	13.8%	13.3%	14.3%	-0.8	-5.9%	p<0.05	[-1.4 to -0.2]
	Mortality rate	53,454	56,008	1.7%	1.4%	1.7%	1.5%	0.0	0.8%	p=0.91	[-0.2 to 0.2]
	Complications[a]	48,475	54,582	2.6%	2.0%	2.4%	1.8%	0.0	-0.5%	p=0.93	[-0.2 to 0.2]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the DiD estimate as a percent of the CJR baseline level.

The change in separate provider payments do not sum to the change in episode payments because separate models were estimated for episode payments and each component payment.

Because CJR participant hospitals shifted a lower share of TKAs to the hospital outpatient setting, the matched control group includes outpatient TKA episodes that have been weighted to balance the episode volume in the CJR hospitals.

CI = confidence interval, DiD = difference-in-differences, ED = emergency department, HH = home health, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, OT = occupational therapy, PAC = post-acute care, PEP = post-episode period, PT = physical therapy, PY = performance year, SNF = skilled nursing facility, TKA = total knee arthroplasty.

[a] The complications measure only applies to elective episodes.

Exhibit D-4: Risk-adjusted assessment-based difference-in-differences results for activities of daily living metrics, opt-in CJR hospitals, LEJR episodes, April 2016-September 2019

First PAC setting	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
IRF	Average change in mobility score	348	966	10.4	10.4	10.0	11.3	-1.3	-12.0%	p<0.01	[-1.9 to -0.7]
SNF	Improved transfer, locomotion on unit, and walking in corridor	8,511	8,894	73.1%	69.8%	70.1%	72.2%	-5.4	-7.4%	p<0.01	[-8.5 to -2.2]
	Improved toilet use	8,476	8,855	51.0%	45.9%	51.5%	51.7%	-5.3	-10.3%	p<0.01	[-8.2 to -2.3]
	Without self-reported pain[a]	8,004	8,464	52.1%	63.4%	51.7%	64.0%	-1.0	-1.9%	p=0.68	[-5.0 to 3.0]
HHA	Improved ambulation/ locomotion	13,878	11,756	91.8%	91.6%	91.1%	91.9%	-0.9	-1.0%	p=0.22	[-2.1 to 0.3]
	Improved bed transferring	13,749	11,665	84.7%	84.9%	84.7%	87.0%	-2.2	-2.6%	p<0.05	[-3.8 to -0.5]
	Reduced pain	13,838	11,704	75.0%	83.2%	73.1%	81.1%	0.2	0.3%	p=0.88	[-2.0 to 2.4]

Source: CJR evaluation team analysis of Medicare claims and enrollment data, MDS data, OASIS data, and IRF-PAI data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by September 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the DiD estimate as a percent of the CJR baseline level.

CI = confidence interval, DiD = difference-in-differences, HHA = home health agency, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, SNF = skilled nursing facility.

[a] The pain measure for those initially discharged to a SNF was not risk adjusted following the specifications of the MDS 3.0 Quality Measure for short-stay patients used in the CMS Nursing Home Five-Star Rating System.

Non-opt-in hospitals

Exhibit D-5: Risk-adjusted claims-based difference-in-differences results for payment, utilization, and quality metrics, non-opt-in CJR hospitals, LEJR episodes, PY1-4

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
Payments	Episode payments	77,400	83,767	$25,851	$24,366	$25,700	$24,576	-$361	-1.4%	p<0.10	[-$666 to -$56]
	IRF payments	77,163	83,767	$937	$663	$1,129	$861	-$5	-0.5%	p=0.96	[-$181 to $170]
	SNF payments	77,400	83,767	$5,285	$3,945	$5,492	$4,179	-$28	-0.5%	p=0.86	[-$288 to $232]
	HH payments	77,400	83,767	$1,999	$1,874	$1,740	$1,724	-$110	-5.5%	p<0.10	[-$211 to -$9]
	Readmission payments[a]	77,400	83,767	$950	$944	$921	$918	-$3	-0.3%	p=0.93	[-$53 to $47]
	Part B payments	77,400	83,767	$4,364	$4,466	$4,324	$4,456	-$29	-0.7%	p=0.53	[-$105 to $47]
	30-day PEP payments[a]	77,400	83,767	$1,218	$1,248	$1,188	$1,238	-$20	-1.7%	p=0.43	[-$62 to $22]
	Anchor payments	77,400	83,767	$12,255	$12,237	$12,208	$12,220	-$30	-0.2%	p=0.49	[-$100 to $41]
Utilization	First PAC IRF	77,400	83,767	5.7%	3.3%	6.1%	4.0%	-0.2	-4.3%	p=0.75	[-1.5 to 1.0]
	First PAC SNF	77,400	83,767	37.9%	27.4%	39.4%	27.5%	1.5	3.9%	p=0.24	[-0.6 to 3.6]
	First PAC HH	77,400	83,767	38.7%	39.6%	32.9%	36.8%	-3.1	-8.0%	p=0.11	[-6.3 to 0.1]
	First PAC home without HH	77,400	83,767	17.7%	29.6%	21.6%	31.6%	1.9	10.6%	p=0.32	[-1.2 to 5.0]
	Any HH use	77,400	83,767	63.9%	58.9%	56.7%	52.6%	-0.9	-1.4%	p=0.66	[-4.4 to 2.5]
	IRF days	2,924	2,999	11.3	11.2	12.0	11.8	0.0	0.4%	p=0.83	[-0.3 to 0.4]
	SNF days	22,068	24,991	26.2	22.3	25.4	23.7	-2.2	-8.3%	p<0.01	[-2.8 to -1.5]
	HH visits	45,201	41,209	14.8	13.9	14.4	14.3	-0.8	-5.3%	p<0.01	[-1.2 to -0.4]
	HH PT/OT visits	45,201	41,209	9.7	9.7	9.5	10.1	-0.6	-6.3%	p<0.01	[-0.9 to -0.4]
	Outpatient PT/OT visits	51,584	56,268	11.6	12.8	12.2	13.4	0.1	0.8%	p=0.63	[-0.2 to 0.4]

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
Quality	Unplanned readmission rate	77,386	83,749	7.7%	7.8%	7.4%	7.4%	0.1	1.3%	p=0.64	[-0.2 to 0.4]
	ED use	77,386	83,749	14.3%	15.2%	15.0%	15.8%	0.1	0.5%	p=0.81	[-0.4 to 0.6]
	Mortality rate	78,942	85,329	2.4%	2.2%	2.3%	2.2%	-0.1	-3.7%	p=0.39	[-0.3 to 0.1]
	Complications[b]	67,562	73,968	3.2%	2.6%	2.8%	2.4%	-0.2	-6.1%	p=0.15	[-0.4 to 0.0]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the DiD estimate as a percent of the CJR baseline level.

The change in separate provider payments do not sum to the change in episode payments because separate models were estimated for episode payments and each component payment.

CI = confidence interval, DiD = difference-in-differences, ED = emergency department, HH = home health, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, OT = occupational therapy, PAC = post-acute care, PEP = post-episode period, PT = physical therapy, PY = performance year, SNF = skilled nursing facility.

[a] We cannot be certain that there is no impact of the model because this outcome failed parallel trends tests (Appendix K). Parallel trends is an assumption that underlies our methodological approach. Please see Appendix C (Section III.C.1.c) for additional details.

[b] The complications measure only applies to elective episodes.

Exhibit D-6: Multi-period difference-in-differences results for payment, utilization, and quality metrics, non-opt-in CJR hospitals, LEJR episodes, PY1-4

Domain	Measure	PY 1-2			PY 3-4		
		DiD	p-value	90% CI	DiD	p-value	90% CI
Payments	Episode payments	-$440	p<0.05	[-$750 to -$131]	-$286	p=0.18	[-$639 to $67]
	IRF payments	$5	p=0.95	[-$144 to $155]	-$20	p=0.85	[-$195 to $156]
	SNF payments	-$140	p=0.34	[-$380 to $101]	$134	p=0.41	[-$135 to $403]
	HH payments	-$32	p=0.59	[-$130 to $66]	-$172	p<0.05	[-$292 to -$53]
	Readmission payments[a]	$26	p=0.47	[-$32 to $83]	-$26	p=0.47	[-$87 to $34]
	Part B payments	-$53	p=0.21	[-$122 to $16]	-$12	p=0.84	[-$108 to $84]
	30-day PEP payments[a]	-$20	p=0.49	[-$69 to $28]	-$22	p=0.48	[-$74 to $29]
	Anchor payments	-$67	p=0.20	[-$153 to $20]	$0	p=0.99	[-$68 to $67]
Utilization	First PAC IRF	-0.1	p=0.90	[-1.1 to 1]	-0.3	p=0.72	[-1.5 to 1.0]
	First PAC SNF	0.3	p=0.82	[-1.6 to 2.1]	2.3	p<0.10	[0.2 to 4.4]
	First PAC HH	-0.9	p=0.64	[-4.0o 2.2]	-5.3	p<0.05	[-9.1 to -1.6]
	First PAC home without HH	0.7	p=0.72	[-2.6 to 4.0]	3.3	p=0.14	[-0.4 to 7.1]
	Any HH use	1.4	p=0.46	[-1.8 to 4.6]	-2.9	p=0.25	[-6.9 to 1.2]
	IRF days	-0.3	p=0.17	[-0.7 to 0.1]	0.4	p=0.14	[0.0 to 0.7]
	SNF days	-2.2	p<0.01	[-2.9 to -1.5]	-2.1	p<0.01	[-2.9 to -1.4]
	HH visits	-0.8	p<0.01	[-1.1 to -0.4]	-0.8	p<0.01	[-1.2 to -0.4]
	HH PT/OT visits	-0.6	p<0.01	[-0.8 to -0.3]	-0.7	p<0.01	[-0.9 to -0.4]
	Outpatient PT/OT visits	-0.1	p=0.54	[-0.4 to 0.2]	0.3	p=0.22	[-0.1 to 0.6]
Quality	Unplanned readmission rate	0.2	p=0.37	[-0.2 to 0.6]	0.0	p=0.99	[-0.4 to 0.4]
	ED use	0.2	p=0.68	[-0.5 to 0.8]	0.0	p=0.94	[-0.6 to 0.7]
	Mortality rate	0.0	p=0.86	[-0.2 to 0.2]	-0.2	p=0.16	[-0.4 to 0.0]
	Complications[b]	-0.2	p=0.25	[-0.4 to 0.1]	-0.2	p=0.25	[-0.5 to 0.1]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The change in separate provider payments do not sum to the change in episode payments because separate models were estimated for episode payments and each component payment.

CI = confidence interval, DiD = difference-in-differences, ED = emergency department, HH = home health, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, OT = occupational therapy, PAC = post-acute care, PEP = post-episode period, PT = physical therapy, PY = performance year, SNF = skilled nursing facility.

[a] We cannot be certain that there is no impact of the model because this outcome failed parallel trends tests (Appendix K). Parallel trends is an assumption that underlies our methodological approach. Please see Appendix C (Section III.C.1.c) for additional details.

[b] The complications measure only applies to elective episodes.

Exhibit D-7: Risk-adjusted assessment-based difference-in-differences results for activities of daily living metrics, non-opt-in CJR hospitals, LEJR episodes, April 2016-September 2019

First PAC setting	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90% CI
		Intervention episodes (N)	Intervention episodes (N)	Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average				
IRF	Average change in mobility score	2,280	2,381	10.3	11.0	9.5	10.2	0.0	0.1%	p=0.98	[-0.4 to 0.5]
SNF	Improved transfer, locomotion on unit, and walking in corridor	15,618	17,532	72.2%	69.5%	70.9%	71.9%	-3.7	-5.1%	p<0.05	[-6.2 to -1.2]
	Improved toilet use	15,529	17,454	50.5%	45.7%	51.1%	50.7%	-4.5	-8.8%	p<0.05	[-7.4 to -1.5]
	Without self-reported pain[a]	14,695	16,622	50.1%	64.3%	49.7%	62.6%	1.3	2.7%	p=0.38	[-1.2 to 3.9]
HHA	Improved ambulation/ locomotion	24,219	21,521	92.0%	91.8%	91.4%	92.1%	-0.9	-1.0%	p=0.12	[-1.8 to 0.0]
	Improved bed transferring	24,020	21,328	86.4%	86.7%	86.1%	88.2%	-1.8	-2.1%	p<0.05	[-3.2 to -0.4]
	Reduced pain[b]	24,128	21,431	76.8%	84.5%	75.6%	82.7%	0.7	0.9%	p=0.50	[-1.0 to 2.4]

Source: CJR evaluation team analysis of Medicare claims and enrollment data, MDS data, OASIS data, and IRF-PAI data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by September 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the **1%**, **5%**, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the DiD estimate as a percent of the CJR baseline level.

CI = confidence interval, DiD = difference-in-differences, HHA = home health agency, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility, Patient Assessment Instrument, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, SNF = skilled nursing facility.

[a] The pain measure for those initially discharged to a SNF was not risk adjusted following the specifications of the MDS 3.0 Quality Measure for short-stay patients used in the CMS Nursing Home Five-Star Rating System.

[b] We cannot be certain that there is no impact of the model because this outcome failed parallel trends tests (Appendix K). Parallel trends is an assumption that underlies our methodological approach. Please see Appendix C (Section III.C.1.c) for additional details.

Exhibit D-8: Multi-period difference-in-differences results for activities of daily living metrics, non-opt-in CJR hospitals, LEJR episodes, PY1-4

First PAC setting	Measure	PY 1-2			PY 3-4		
		DiD	p-value	90% CI	DiD	p-value	90% CI
IRF	Average change in mobility score	-0.2	p=0.48	[-0.7 to 0.3]	0.2	p=0.58	[-0.4 to 0.9]
SNF	Improved transfer, locomotion on unit, and walking in corridor	-4.5	p<0.01	[-6.6 to -2.3]	-2.9	p=0.13	[-6.1 to 0.2]
	Improved toilet use	-5.7	p<0.01	[-8.3 to -3.1]	-3.1	p=0.15	[-6.6 to 0.5]
	Without self-reported pain[a]	0.4	p=0.81	[-2.1 to 2.9]	1.9	p=0.26	[-0.9 to 4.7]
HHA	Improved ambulation/Locomotion	-0.8	p=0.17	[-1.8 to 0.2]	-1.0	p=0.16	[-2.2 to 0.2]
	Improved bed transferring	-1.5	p<0.10	[-2.9 to -0.1]	-2.1	p<0.10	[-4.0 to -0.2]
	Reduced pain[b]	0.9	p=0.35	[-0.7 to 2.5]	0.1	p=0.91	[-1.8 to 2.0]

Source: CJR evaluation team analysis of Medicare claims and enrollment data, MDS data, OASIS data, and IRF-PAI data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by September 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, DiD = difference-in-differences, HHA = home health agency, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, PY = performance year, SNF = skilled nursing facility.

[a] The pain measure for those initially discharged to a SNF was not risk adjusted following the specifications of the MDS 3.0 Quality Measure for short-stay patients used in the CMS Nursing Home Five-Star Rating System.

[b] We cannot be certain that there is no impact of the model because this outcome failed parallel trends tests (Appendix K). Parallel trends is an assumption that underlies our methodological approach. Please see Appendix C (Section III.C.1.c) for additional details.

Performance-year specific results

Exhibit D-9: Performance-year specific difference-in-differences results for average episode payments, mandatory hospitals, opt-in hospitals, and non-opt-in hospitals, LEJR episodes, PY1-4

Measure	Performance Year	Mandatory			Opt-in			Non-opt-in		
		DiD	p-value	90% CI	DiD	p-value	90% CI	DiD	p-value	90% CI
Episode Payments	PY1	-$1,431	p<0.01	[-$2,053 to -$810]	-$520	p<0.10	[-$1,030 to -$11]	-$133	p=0.57	[-$515 to $249]
	PY2	-$1,618	p<0.01	[-$2,199 to -$1,037]	-$821	p<0.01	[-$1,181 to -$461]	-$583	p<0.01	[-$900 to -$265]
	PY3	-$1,330	p<0.01	[-$1,995 to -$665]	-$517	p<0.05	[-$947 to -$86]	-$203	p=0.37	[-$572 to $166]
	PY4	-$1,263	p<0.01	[-$1,977 to -$549]	-$985	p<0.01	[-$1,484 to -$486]	-$372	p=0.14	[-$789 to $45]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

Because mandatory and opt-in CJR hospitals shifted a lower share of TKAs to the hospital outpatient setting, the respective control groups include outpatient TKA episodes that have been weighted to balance the episode volume in the CJR hospitals.

CI = confidence interval, DiD = difference-in-differences, LEJR = lower extremity joint replacement, PY = performance year, TKA = total knee arthroplasty.

Appendix E: Medicare Program Savings Sensitivity Analyses

In our main analysis, we present estimates of Medicare Program savings separately for mandatory CJR hospitals, opt-in CJR hospitals, and non-opt-in CJR hospitals, as well as a combined estimate of Medicare savings for all CJR hospitals. Those estimates do not include hospitals in mandatory MSAs designated as low volume or rural under the CJR model (LVR).[1] After the first two years of the model, LVR hospitals in mandatory MSAs were given the one-time option to continue participation in the CJR model. Fourteen hospitals categorized as rural, and one hospital categorized as low volume, opted to remain in the CJR model.

We do not include LVR hospitals in the analysis of opt-in and non-opt-in CJR hospitals in voluntary MSAs because LVR hospitals differ in important ways that would make them less comparable to the voluntary control group. First, by construction, the LVR hospitals are located in MSAs with higher average historical payments. Second, LVR hospitals are located in MSAs in which most hospitals are participating in CJR. We also do not include LVR hospitals in the analysis of mandatory CJR hospitals since the ability of LVR hospitals to select to continue in the model would make them less comparable to the mandatory control group.

In this section we provide a sensitivity analysis to determine whether our overall conclusions regarding Medicare Program savings would change if we included LVR hospitals. The potential impact of adding LVR hospitals to our analysis of total Medicare Program savings depends on their overall episode volume, the extent to which LVR hospitals reduced payments, and their net reconciliation payments under the CJR model. Exhibit E-1 reports descriptive statistics for episode volume and net reconciliation payments (NPRA) among LVR hospitals.

[1] Hospitals are defined as low volume if they had less than 20 episodes over a three-year historical period (2012 to 2014) and defined as rural based if they have section 401 status in FY 2019 IPPS data or an address located in a rural Census tract (identified using https://data.hrsa.gov/tools/rural-health). A hospital must apply for section 401 status.

Exhibit E-1: Descriptive statistics for LVR hospitals in mandatory MSAs and non-LVR hospitals by performance year

	LVR hospitals			Non-LVR hospitals[a]		
	Number of episodes	Total NPRA	NPRA per episode	Number of episodes	Total NPRA	NPRA per episode
PY1	4,765	$4,116,475	$864	39,807	$31,141,677	$782
PY2	11,247	$12,694,789	$1,129	83,865	$78,370,686	$934
PY3	6,905	$12,052,067	$1,745	54,075	$48,933,697	$905
PY4	7,148	$14,684,075	$2,054	63,191	$71,915,687	$1,138
Total	30,065	$43,547,406	$1,448	240,938	$230,361,747	$956

Source: CJR evaluation team analysis of CJR payment contractor data for CJR participating hospitals in PY1 (episodes starting on or after April 2016 and ending on or before December 2016), PY2 (episodes ending between January and December 2017), PY3 (episodes ending between January and December 2018), and PY4 (episodes ending between January and December 2019).

***Notes*:** LVR = low volume and rural, MSA = metropolitan statistical area, NPRA = net payment reconciliation amount, PY = performance year.

[a] Non-LVR hospitals include mandatory CJR hospitals and voluntary opt-in and non-opt-in CJR hospitals.

In total, LVR hospitals in mandatory MSAs accounted for slightly over 30,000 episodes and $43.5 million in net reconciliation payments over the first four performance years. Over the same timeframe, the average reconciliation payment per episode was $1,448. Over the first four performance years, low volume hospitals accounted for only 417 episodes and approximately $54,200 in net reconciliation payments, averaging $130 in reconciliation per episode. Rural hospitals accounted for 98.6% of all LVR episodes and 99.9% of total net reconciliation among LVR hospitals. For scale, non-LVR hospitals in the CJR model accounted for 240,938 episodes and $230.4 million in net reconciliation payments over the first four performance years, with an average reconciliation payment per episode of $956.

We performed a sensitivity analysis to determine whether our overall conclusions regarding Medicare Program savings would change if we included LVR hospitals. To obtain a range of possible estimates, we assume that LVR hospitals reduced payments by no more or no less than 20% of the average payment reductions for mandatory hospitals or voluntary hospitals. Specifically, we report the point estimate of net Medicare Program savings including LVR hospitals under four assumptions:

1. LVR hospitals (opt-in and non-opt-ins) reduced spending by 20% more than mandatory hospitals in each PY.
2. LVR hospitals (opt-in and non-opt-ins) reduced spending by 20% less than mandatory hospitals in each PY.
3. LVR opt-in hospitals reduced spending by 20% more than voluntary opt-in hospitals in each PY, and LVR non-opt-in hospitals reduced spending by 20% more than voluntary non-opt-in hospitals in each PY.

4. LVR opt-in hospitals reduced spending by 20% less than voluntary opt-in hospitals in each PY, and LVR non-opt-in hospitals reduced spending by 20% less than voluntary non-opt-in hospitals in each PY.

Exhibit E-2 reports the savings estimates resulting from each of these assumptions. Assumptions 1 and 2 implicitly suppose that the geographic similarities between LVR hospitals and mandatory hospitals are the dominant factor in determining payment reductions, while Assumptions 3 and 4 implicitly suppose that the similarity in the voluntary nature of the participation decision between LVR hospitals and voluntary hospitals is the dominant factor. The additional buffer of 20% in either direction raises the likelihood that the true impact of including LVR hospitals in our analysis is bracketed by the sensitivity analysis estimates.

Exhibit E-2: Sensitivity analyses of Medicare program savings including low volume and rural hospitals

LVR Hospital Assumption	Assumption	PY1	PY2	PY3	PY4	Cumulative
Main analysis (LVR hospitals excluded)	N/A	$3,182,422 [-$11,038,235 to $17,403,078]	$15,688,707 [-$12,452,537 to $43,829,952]	$7,134,244 [-$20,214,466 to $34,482,954]	-$4,567,437 [-$40,218,041 to $31,083,167]	$21,437,936 [-$75,033,091 to $117,908,963]
Savings relative to mandatory hospitals	1 (+20%)	$7,034,280	$24,202,972	$5,474,963	-$9,834,479	$26,877,736
	2 (-20%)	$4,378,169	$17,133,287	$2,010,701	-$12,973,490	$10,548,667
Savings relative to hospitals in voluntary MSAs	3 (+20%)	$966,755	$12,457,457	-$1,225,907	-$11,073,898	$1,124,406
	4 (-20%)	$333,153	$9,302,944	-$2,456,546	-$13,799,769	-$6,620,219

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention) and CJR payment contractor data for CJR participant hospitals in PY1-PY4.

***Notes*:** Savings numbers are totals for all hospitals including LVR hospitals, based on Assumptions 1 through 4 above. For the main analysis with LVR hospitals excluded, we report savings numbers and 90% confidence intervals.

LVR = low volume and rural, N/A = not applicable, PY = performance year

The main estimate of total Medicare program savings is $21.4 million with a range of -$75.0 million to $117.9 million. By including LVR hospitals under the previously described assumptions the point estimate of total Medicare program savings could range from a loss of $6.6 million to a savings of $26.9 million. Therefore, all four sensitivity analyses result in point estimates of Medicare savings that lie within the range reported in our main analysis.

Appendix F: Outcome Definitions

Exhibit F-1: Claims-based outcome definitions

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
Medicare payments	Total Medicare standardized allowed amounts per episode[1]	The sum of Medicare payment and beneficiary out-of-pocket amounts for related items and services covered by Medicare Part A and Part B[2] performed during the LEJR hospitalization (anchor hospitalization) through the 90-day post-discharge period that are included in the episode.	Anchor hospitalization through 90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
	Medicare standardized allowed amount for the anchor hospitalization per episode	The sum of Medicare payment and beneficiary out-of-pocket amounts for the LEJR anchor hospitalization (MS-DRG 469 or 470 for inpatient episodes covered under Medicare Part A; CPT 27447 for outpatient TKA episodes covered under Medicare Part B).	Anchor hospitalization	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
	Medicare Part A standardized allowed amounts per episode, by service	The sum of Medicare payment and beneficiary out-of-pocket amounts for readmissions, IRF, and SNF services covered under Medicare Part A. Includes all costs incurred during the 90 days following discharge.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.

[1] Standardized payments remove wage adjustments and other Medicare payment adjustments (e.g., GME, IME, and DSH). Allowed amounts include beneficiary cost sharing.

[2] Episode-related items and services paid under Medicare Part A or Part B, after exclusions are applied, include: physician services; inpatient hospital services (including readmissions with certain exceptions discussed in the Final Rule); inpatient psychiatric facility (IPF) services; LTCH services; IRF services; SNF services; HHA services; hospital outpatient services; outpatient therapy services; clinical laboratory services; DME; Part B drugs; and hospice.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
Medicare payments (cont'd)	Medicare standardized allowed amounts for HHA services per episode	The sum of Medicare payment and beneficiary out-of-pocket amounts for HHA services covered under Medicare Part A or Part B HHA.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
	Medicare Part B standardized allowed amounts per episode	The sum of Medicare payment and beneficiary out-of-pocket amounts for related items and services covered under Medicare Part B (except HHA services) including physician evaluation and management services, outpatient therapy services (speech, occupation, and physical therapy), imaging and lab services, procedures, DME, all other non-institutional services, and other institutional services.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
	Medicare standardized allowed amounts for services provided in the 30 days post-episode per episode	The sum of Medicare payment and beneficiary out-of-pocket amounts for all health care services covered under Medicare Part A or B performed during the 30-day post-episode period	30-day post-episode period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
Utilization	First discharge to IRF	The percent of all episodes with beneficiaries initially discharged to an IRF. The first PAC setting is an IRF (a freestanding facility or a distinct unit within an acute hospital) if admission to the IRF occurred within the first five days of hospital discharge and no other PAC use occurred prior to IRF admission. If the beneficiary is directly transferred to another ACH after the anchor hospitalization, then the first PAC setting was defined within five days of the transfer discharge.	1st to 5th day after discharge from the anchor/ transfer hospitalization	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
Utilization (cont'd)	First discharge to SNF	The percent of all episodes with beneficiaries initially discharged to a SNF. The first PAC setting is a SNF if admission to the SNF occurred within the first five days of hospital discharge and no other PAC use occurred prior to SNF admission. If the beneficiary is directly transferred to another ACH after the anchor hospitalization, then the first PAC setting was defined within five days of the transfer discharge.	1st to 5th day after discharge from the anchor/ transfer hospitalization	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
	First discharge to HHA	The percent of all episodes with beneficiaries initially discharged to an HHA. The first PAC setting is an HHA if admission to the HHA occurred within 14 days of hospital discharge and no other PAC use occurred prior to HHA admission. If the beneficiary is directly transferred to another ACH after the anchor hospitalization, then the first PAC setting was defined within 14 days of the transfer discharge.	1st to 14th day after discharge from the anchor/ transfer hospitalization	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
	First discharge to home without HHA	The percent of all episodes with beneficiaries initially discharged to home without HHA services. The first PAC setting is home without HHA if the beneficiary is not admitted to a SNF or IRF within 5 days of hospital discharge and is not admitted to an HHA within 14 days of hospital discharge. If the beneficiary is directly transferred to another ACH after the anchor hospitalization, then the first PAC setting was defined within 14 days of the transfer discharge.	1st to 14th day after discharge from the anchor/ transfer hospitalization	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
Utilization (cont'd)	Any HH use	The percent of all episodes with beneficiaries using any HHA services during the 90-day post-discharge period, as indicated by non-zero Medicare payment and beneficiary out-of-pocket amounts for HHA services covered under Medicare Part A or Part B.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have non-missing Medicare standardized allowed payment information for the episode.
	Number of IRF days	The average number of IRF days of care during the 90-day post-discharge period.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have at least one IRF day during this period; 6) have non-missing Medicare standardized allowed payment information for the episode.
	Number of SNF days	The average number of SNF days of care during the 90-day post-discharge period.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have at least one SNF day during this period; 6) have non-missing Medicare standardized allowed payment information for the episode.
	Number of HHA visits	The average number of HHA visits during the 90-day post-discharge period.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have at least one HHA visit during this period; 6) have non-missing Medicare standardized allowed payment information for the episode.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
Utilization (cont'd)	Number of HHA PT/OT visits	The average number of HHA physical therapy and occupational therapy visits during the 90-day post-discharge period.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have at least one HHA visit during this period; 6) have non-missing Medicare standardized allowed payment information for the episode.
	Number of outpatient PT/OT visits	The average number of outpatient physical therapy and occupational therapy (PT/OT) visits during the 90-day post-discharge period.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) have at least one outpatient PT/OT visit during this period; 6) have non-missing Medicare standardized allowed payment information for the episode.
Quality	Unplanned readmission rate	The proportion of episodes with one or more unplanned readmissions for any eligible condition. This measure was based on specifications for the NQF-endorsed all-cause unplanned readmission measure (NQF measure 1789).[3] Following these specifications, we excluded planned admissions, based on AHRQ Clinical Classification System Procedure and Diagnoses codes.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) are discharged from the anchor hospital hospitalization in accordance with medical advice; 6) have non-missing Medicare standardized allowed payment information for the episode.

[3] Updated specification documents were released by CMS in March 2019 for the unplanned readmission measure, and the measure was revised accordingly. Available at: https://www.qualitynet.org/inpatient/measures/readmission/methodology

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
Quality (cont'd)	Emergency department visit rate	The proportion of episodes with one or more ED visits during the 90-day post-discharge period for which the beneficiary required medical treatment but was not admitted to the hospital. Eligible ED visits are outpatient claims with a code indicating the beneficiary used the emergency department but was not admitted to the hospital.	90-day post-discharge period	Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have a measurement period that ends on or before December 31, 2019; 5) are discharged from the anchor hospital hospitalization in accordance with medical advice; 6) have non-missing Medicare standardized allowed payment information for the episode.
	All-cause mortality rate	Death from any cause during the anchor hospitalization or 90-day post-discharge period.	Anchor hospitalization and 90-day post-discharge period	Under the CJR model, death during the anchor hospitalization or 90-day PDP cancels the episode. Therefore, this analysis includes CJR and control group episodes as well as beneficiaries at CJR participant and control group hospitals that would have been identified as episodes if they had not died during the episode of care. Beneficiaries who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age <115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) have not received hospice care in the six months prior to admission; 5) have a measurement period that ends on or before December 31, 2019; 6) are discharged from the anchor hospital hospitalization in accordance with medical advice; 7) have non-missing Medicare standardized allowed payment information for the episode.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
	Incidence of any complications	The proportion of elective episodes with incidence (during the anchor hospitalization or a readmission) of: AMI, pneumonia, or sepsis/septicemia within the 7-day PDP; or surgical site bleeding or pulmonary embolism within the 30-day PDP; or mechanical complications, periprosthetic joint infection, or wound infection within the 90-day PDP. This measure was based on specifications for the NQF-endorsed THA/TKA complications measure (NQF measure 1550).[4] Death in the 30 days after discharge is part of the technical definition, but is not included in our analysis because beneficiaries who died during the anchor hospitalization or in the 90-day PDP are excluded from the CJR model.	90-day post-discharge period	Beneficiaries who: 1) have an elective procedure (non-fracture); 2) have a complete FFS enrollment history six months prior to the anchor hospitalization; 3) have consistent, reliable sex and age data (age <115); 4) maintain Parts A and B enrollment throughout the measurement period; 5) have a measurement period that ends on or before December 31, 2019; 6) are discharged from the anchor hospital hospitalization in accordance with medical advice; 7) have non-missing Medicare standardized allowed payment information for the episode.

Notes: ACH = acute care hospital, AHRQ = Agency for Healthcare Research and Quality, AMI = acute myocardial infarction, CPT = current procedural terminology, DME = durable medical equipment, DSH = disproportionate share hospital, ED = emergency department, FFS = fee-for-service, GME = graduate medical education, HH = home health, HHA = home health agency, IME = indirect medical education, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, LTCH = long-term care hospital, MS-DRG = Medicare Severity-Diagnosis Related Group, NQF = National Quality Forum, OT = occupational therapy, PAC = post-acute care, PDP = post-discharge period, PT = physical therapy, SNF = skilled nursing facility, THA = total hip arthroplasty, TKA = total knee arthroplasty.

[a] The eligible sample column notes the inclusion criteria for episodes as defined by the Final Rule and additional measure-specific inclusion criteria required for the evaluation.

[4] Updated specification documents were released by CMS in March 2019 for the THA/TKA complications measure, and the measure was revised accordingly. Available at: https://www.qualitynet.org/inpatient/measures/complication/methodology

Exhibit F-2: Assessment-based quality outcome definitions

First PAC setting	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
HHA	Improved ambulation/ locomotion	Percent of patients who improve status in ambulation/locomotion over the measurement period (i.e., change in performance score that was negative).	From start or resumption of HH care to HHA discharge, if HHA discharge is within 90 days of hospital discharge. Else, from start or resumption of HH care to the 60-day recertification assessment.	Beneficiaries whose first PAC setting is HHA who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age<115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) had a valid start or resumption of care assessment and at least one follow-up OASIS assessment within 90 days of hospital discharge; 5) were not transferred from HH care to an inpatient facility during the HHA episode or at discharge; 6) could not perform the ADL independently (had pain) at start or resumption of care; 7) had no missing data used to calculate the performance score.
	Improved bed transferring	Percent of patients who improve status in bed transferring over the measurement period (i.e., change in performance score that was negative).		
	Reduced pain	Percent of patients whose frequency of pain when moving around reduced.		
SNF	Improved transfer, locomotion on unit, and walking in corridor	Percent of patients whose cumulative status in transfer, locomotion on unit, and walk in corridor improved over the measurement period (i.e., change in performance score that was negative).	SNF admission to SNF discharge, if SNF discharge is within 90 days of hospital discharge. Else, from SNF admission to the most recent MDS PPS assessment within 90 days of hospital discharge.	Beneficiaries whose first PAC setting is a SNF who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age<115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) had a valid 5-day MDS assessment and at least one follow-up MDS assessment within 90 days of hospital discharge; 5) were not indicated as comatose, whose life expectancy was greater than six months, and were not in hospice as of the 5-day MDS assessment; 6) were not independent in all three ADLs (for the first measure) and dressing (for the second measure) at the 5-day MDS assessment; 7) had no missing data used to calculate the performance score.
	Improved toilet use	Percent of patients with improved status in toilet use over the measurement period (i.e., change in performance score that was negative).		
	Without self-reported pain	Percent of patients who did not self-report moderate to severe pain in the first five days of their SNF hospitalization. (This measure is not risk-adjusted.)	Measured once within five days of SNF admission.	Beneficiaries whose first PAC setting is a SNF who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age<115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) had a valid 5-day MDS assessment, with the pain assessment interview and pain presence item completed and, if any pain was indicated, the pain frequency and pain intensity items were completed and valid.

First PAC setting	Outcome name	Definition	Measurement period(s)	Eligible sample[a]
IRF	Average change in mobility score	Average change in a composite mobility score over the measurement period. The composite score ranges from 4 (worst) to 28 (best).	From IRF admission to IRF discharge	Beneficiaries whose first PAC setting is an IRF who: 1) have a complete FFS enrollment history six months prior to the anchor hospitalization; 2) have consistent, reliable sex and age data (age<115); 3) maintain Parts A and B enrollment throughout the measurement period; 4) had a valid IRF-PAI assessment with discharge at or before 90 days after hospital discharge; 5) were not diagnosed with the following conditions on the IRF-PAI assessment: coma, persistent vegetative state, complete tetraplegia, locked-in syndrome, severe anoxic brain damage, cerebral edema, or compression of brain; 6) were not independent in mobility (for the first measure) and lower body dressing (for the second measure) at the time of admission; 7) had a length of hospitalization longer than three days; 8) were not discharged from the IRF against medical advice; 9) had no missing data used to calculate the performance score.

Note: ADL = activities of daily living, FFS = fee-for-service, HH = home health, HHA = home health agency, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility-Patient Assessment Instrument, MDS = Minimum Data Set, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, PPS = prospective payment system, SNF = skilled nursing facility.

[a] The eligible sample column notes the inclusion criteria for episodes as defined by the Final Rule and additional measure-specific inclusion criteria required for the evaluation.

Appendix G: Definitions of Patient Characteristics

Exhibit G-1: Patient characteristic variable definitions

Variable	Definition	Source
Age	Percent of patients by age category; 20 to 64, 65 to 79, 80 and above.	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Enrollment Database
Ambulation/locomotion	Among those first discharged to HHA, percent of patients who were bedfast, chairfast, or able to walk only with supervision or assistance at all times. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Bathing	Among those first discharged to HHA, percent of patients who were: able to bathe in shower or tub but required presence of another person throughout; unable to use shower or tub but able to bathe independently at the sink, in chair, or on commode; or bathed totally by another person. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Bathing poor	Among those first discharged to SNF, percent of patients who either needed physical help or were totally dependent in taking a full-body bath or shower, sponge bath. Measured at admission to the SNF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) MDS
Cognition not intact	Among those first discharged to SNF, percent of patients who were considered cognitively not intact if they completed the BIMS and scored lower than 13, out of 15. Measured at admission to the SNF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) MDS
Cognitive functioning	Among those first discharged to HHA, percent of patients who required assistance and some direction in specific situations or consistently required low stimulus environment due to distractibility; required considerable assistance in routine situations; or were totally dependent due to disturbances. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Cognitive index	Among those first discharged to IRF, cognitive score (a composite functional status measure built from the FIM instrument) with higher scores indicating better cognition. Measured at admission to the IRF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) IRF-PAI
Confusion	Among those first discharged to HHA, percent of patients who were nonresponsive, confused constantly, or confused during the day and evening, but not constantly, in the last 14 days. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Congestive heart failure	Percent of patients with congestive heart failure (HCC flag #85).	July 2011 – December 2014 (baseline) and April 2015 – December 2019 (intervention) Medicare Claims

Variable	Definition	Source
Dementia	Percent of patients with dementia (with and without complications; HCC flags #51 and #52).	July 2011 – December 2014 (baseline) and April 2015 – December 2019 (intervention) Medicare Claims
Diabetes	Percent of patients with diabetes.	July 2010 – December 2014 (baseline) and April 2014 – December 2019 (intervention) Medicare Claims
Disability, not due to ESRD	Percent disabled, based on Medicare eligibility status (not including ESRD).	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Enrollment Database
Early-loss ADL score	Among those first discharged to SNF, early loss ADL score (dressing, personal hygiene) with higher values indicating more severe limitations. Measured at admission to the SNF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) MDS
Eligible for Medicaid	Percent eligible for Medicaid based on Medicare enrollment file.	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Enrollment Database
Fracture status	Percent of patients with hip fractures at the anchor hospitalization based on ICD codes provided by CMMI on the CJR model website (https://innovation.cms.gov/Files/worksheets/cjr-icd10hipfracturecodes.xlsx).	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Claims
Grooming	Among those first discharged to HHA, percent of patients who needed someone to assist them with grooming self or depended entirely upon someone else for grooming needs. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
HCC score	Average CMS-HCC score that corresponds to the HCCs present during the one year prior to the anchor hospitalization. HCC scores of less than 1.0 indicate the patient is healthier than the average Medicare beneficiary, while scores greater than 1.0 indicate a patient is unhealthier than the average Medicare beneficiary.	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Claims
Hypertension	Percent of patients with hypertension.	July 2011 – December 2014 (baseline) and April 2015 – December 2019 (intervention) Medicare Claims
Impaired decision-making	Among those first discharged to HHA, percent of patients who failed to perform usual ADLs or IADLs, were unable to appropriately stop activities, or jeopardized safety through actions at least once a week. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Late-loss ADL score	Among those first discharged to SNF, late loss ADL score (bed mobility, eating, transfers, and toilet use) with higher values indicating more severe limitations. Measured at admission to the SNF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) MDS

Variable	Definition	Source
Lower body dressing	Among those first discharged to HHA, the percent of patients who were entirely dependent or someone had to help them put on undergarments, slacks, socks, nylons, or shoes. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Memory deficit	Among those first discharged to HHA, percent of patients who failed to recognize familiar persons/places, were unable to recall events of past 24 hours, or had significant memory loss so that supervision is required at least once a week. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Mid-loss ADL score	Among those first discharged to SNF, mid-loss ADL score (motion, transfer, locomotion, and walking in corridor) based on MDS section G with higher values indicating more severe limitations. Measured at admission to the SNF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) MDS
Mobility index	Among those first discharged to IRF, composite mobility score with higher scores indicating better mobility. Measured at admission to the IRF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) IRF-PAI
MS-DRG 469	Percent of patients discharged under MS-DRG 469 (major joint replacement or reattachment of lower extremity with major complications or comorbidities) for the anchor hospitalization.	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Claims
Obesity	Percent of patients obese or with a BMI of greater than 30.	July 2010 – December 2014 (baseline) and April 2014 – December 2019 (intervention) Medicare Claims
Prior ACH stay	Percent of patients with one or more inpatient acute care hospitalizations during the six months prior to the anchor hospitalization.	July 2011 - December 2014 (baseline) and October 2015 – December 2019 (intervention) Medicare Claims
Prior care use	Percent of patients with any care use (inpatient, SNF, IRF, HHA, or LTCH) during the six months prior to anchor hospitalization.	July 2011 - December 2014 (baseline) and October 2015 – December 2019 (intervention) Medicare Claims
Prior IRF use	Percent of patients with one or more IRF stays during the six months prior to the anchor hospitalization.	July 2011 - December 2014 (baseline) and October 2015 – December 2019 (intervention) Medicare Claims
Prior SNF use	Percent of patients with one or more SNF stays during the six months prior to the anchor hospitalization.	July 2011 - December 2014 (baseline) and October 2015 – December 2019 (intervention) Medicare Claims
Prior HHA use	Percent of patients with one or more instances of HHA use during the six months prior to the anchor hospitalization.	July 2011 - December 2014 (baseline) and October 2015 – December 2019 (intervention) Medicare Claims

Variable	Definition	Source
Race/ethnicity	Percent of patients by race/ethnicity: White, Black or African-American, Hispanic, Other race, or Unknown.	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Enrollment Database
Self-care index	Among those first discharged to IRF, self-care score (a composite functional status measure built from the FIM instrument) with higher scores indicating better self-care. Measured at admission to the IRF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) IRF-PAI
Severe cognitive impairment	Among those first discharged to SNF, percent of patients who were considered to have a severe cognitive impairment based on either a BIMS score of less than 7 (out of 15), or their cognitive skills for daily decision making were severely impaired *and* they had a short-term memory problem. Measured at admission to the SNF.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) MDS
Sex	Percent of female patients.	January 2012 - December 2014 (baseline) and April 2016 – December 2019 (intervention) Medicare Enrollment Database
Toilet transferring	Among those first discharged to HHA, percent of patients who were unable to get to and from the toilet or were totally dependent in toileting. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Toileting hygiene	Among those first discharged to HHA, percent of patients who were totally dependent or needed someone to help them maintain toileting hygiene and/or adjust clothing. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Transferring	Among those first discharged to HHA, percent of patients who were unable to transfer self from bed to chair or were bedfast. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS
Upper body dressing	Among those first discharged to HHA, percent of patients who were entirely dependent or needed someone to help them put on upper body clothing. Measured at start of HHA care.	January 2012 – December 2014 (baseline) and April 2016 – September 2019 (intervention) OASIS

Note: ACH = acute care hospital, ADL = activities of daily living, BIMS = brief interview for mental status, BMI = body mass index, CMS = Centers for Medicare & Medicaid Services, ESRD = end-stage renal disease, FIM = functional independence measure, HCC = hierarchical condition category, HHA = home health agency, IADL = instrumental activities of daily living, ICD = International Classification of Diseases, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility-Patient Assessment Instrument, LTCH = long-term care hospital, MDS = minimum data set, MS-DRG = Medicare Severity-Diagnosis Related Group, OASIS = outcome and assessment information set, SNF = skilled nursing facility.

Appendix H: Patient Survey Questions

1. Who is completing this survey?
   - ○ Person named in the cover letter
   - ○ Person named in the cover letter, with help from a family member, friend or caregiver
   - ○ A family member, friend, or caregiver of the person named in the cover letter
   - ○ If the person to whom this was mailed cannot complete the survey, and there is no one else who can do it for him or her, please mark this response and return the blank survey

Section 1. Before the Hospital

We would like to know how you were doing before you went to the hospital listed in the cover letter to have your joint replaced.

2. Did you have any sessions with a physical therapist for the joint you had replaced in the two weeks or so before your joint replacement surgery?
   - ○ Yes
   - ○ No
   - ○ Don't know/Don't remember

The next questions ask about the week before your joint replacement surgery.

3. Thinking about the week before your joint replacement surgery, how often did pain in the joint that you had replaced interfere with your normal activities?
   - ○ All of the time
   - ○ Most of the time
   - ○ Some of the time
   - ○ A little of the time
   - ○ None of the time
   - ○ Don't know/Don't remember

4. Thinking about the week before your joint replacement surgery, were you taking any of the following types of medications specifically for pain in the joint that you had replaced?
   - ○ Prescription pain medication only
   - ○ Over the counter pain medication only
   - ○ Both prescription and over the counter pain medications
   - ○ No medication for pain in the joint that was replaced
   - ○ Don't know/Don't remember

5. Thinking about the week before your joint replacement surgery, what best describes your use of a mobility aid such as a wheelchair, scooter, walker, or cane?
   - ○ I never used a mobility aid
   - ○ I sometimes used a mobility aid
   - ○ I always used a mobility aid
   - ○ Don't know/Don't remember

6. Thinking about the week before your joint replacement surgery, what best describes your ability to walk by yourself without resting? That is, walk without the help of another person or the help of a mobility aid.
   - ○ I could walk more than several blocks by myself without resting
   - ○ I could walk several blocks by myself without resting
   - ○ I could walk one block by myself without resting
   - ○ I could walk from one room to another by myself without resting
   - ○ I was not able to walk by myself without resting
   - ○ Don't know/Don't remember

7. Thinking about the week before your joint replacement surgery, how much difficulty did you have walking up or down 12 stairs?
   - ○ I had no difficulty walking up or down 12 stairs
   - ○ I had some difficulty walking up or down 12 stairs
   - ○ I had a lot of difficulty walking up or down 12 stairs
   - ○ I was not able to walk up or down 12 stairs
   - ○ Don't know/Don't remember

8. Thinking about the week before your joint replacement surgery, how much difficulty did you have rising from sitting?
   - ○ Extreme
   - ○ Severe
   - ○ Moderate
   - ○ Mild
   - ○ None
   - ○ Don't know/Don't remember

9. Thinking about the week before your joint replacement surgery, how much difficulty did you have standing?
   - ○ Extreme
   - ○ Severe
   - ○ Moderate
   - ○ Mild
   - ○ None
   - ○ Don't know/Don't remember

10. Thinking about the week before your joint replacement surgery, how much difficulty did you have getting on/off the toilet?
    - ○ Extreme
    - ○ Severe
    - ○ Moderate
    - ○ Mild
    - ○ None
    - ○ Don't know/Don't remember

Section 2. After the Hospital

Now we'd like to learn about your experience after you left the hospital listed in the cover letter, and the weeks immediately after.

11. Thinking about when you left the hospital for your joint replacement surgery, would you say that you were…
    - ○ Discharged too early
    - ○ Discharged at the right time or
    - ○ Discharged too late
    - ○ Don't know/Don't remember

12. Thinking about the care you received – in the two weeks after your joint replacement surgery – from doctors, nurses and therapists, at home, in a doctor or therapist's office or in a medical facility – how would you rate the level of care overall?
    - ○ Level of care during two weeks after surgery was more than I needed
    - ○ Level of care during two weeks after surgery was about right
    - ○ Level of care during two weeks after surgery was not enough
    - ○ Don't know/Don't remember

13. Do you live in your own home, in someone else's home, or in an assisted living facility?
    - ○ Yes
    - ○ No, **Go To Section 3 on page I-5**

14. When you went home after your joint replacement surgery, did you have all the medical equipment you needed (for example, walker, elevated commode, grabber, shower chair, device to help put on socks)?
    - ○ Yes
    - ○ No
    - ○ Don't know/Don't remember

We would like to learn about the help you received from other people when you went home after your joint replacement surgery, or to someone else's home or an assisted living facility.

15. Thinking back to the people who helped you, who was your main caregiver, that is, the person who helped you the most after your joint replacement surgery?
    - ○ Spouse/partner
    - ○ Adult child
    - ○ Another relative
    - ○ Paid caregiver
    - ○ Friend, neighbor, or someone else
    - ○ No help at home after joint replacement surgery

16. When you went home after joint replacement surgery, how much help did you need from your main caregiver with putting on or taking off your clothes?
    - ○ No help needed
    - ○ Some help needed
    - ○ Complete help needed
    - ○ Don't know/Don't remember

17. When you went home after joint replacement surgery, how much help did you need from your main caregiver with using the toilet?
    - ○ No help needed
    - ○ Some help needed
    - ○ Complete help needed
    - ○ Don't know/Don't remember

18. When you went home after joint replacement surgery, how much help did you need from your main caregiver with bathing?
    - ○ No help needed
    - ○ Some help needed
    - ○ Complete help needed
    - ○ Don't know/Don't remember

Section 3. Health Care Experiences in-Hospital and After

We want to learn about your experiences while you were in the hospital listed in the cover letter and any other place where you received medical care following that hospitalization.

In the following questions, the term "health care providers" means doctors, nurses, physical or occupational therapists and any other medical professionals who helped take care of you during your time in the hospital and afterwards, in other facilities or at home in any capacity.

Please think of all these types of providers and locations when rating your level of satisfaction in the next few questions.

19. How satisfied or dissatisfied were you with the extent to which health care providers listened to your thoughts and preferences about your medical treatment?
    - ○ Very dissatisfied
    - ○ Somewhat dissatisfied
    - ○ Neither satisfied nor dissatisfied
    - ○ Somewhat satisfied
    - ○ Very satisfied

20. How satisfied or dissatisfied were you with the place you were sent after you left the hospital, for example, home, rehabilitation facility, nursing home, long-term care hospital?
    - ○ Very dissatisfied
    - ○ Somewhat dissatisfied
    - ○ Neither satisfied nor dissatisfied
    - ○ Somewhat satisfied
    - ○ Very satisfied

21. How satisfied or dissatisfied were you with the coordination of your care among doctors, nurses, and therapists in the hospital and after discharge?
    - ○ Very dissatisfied
    - ○ Somewhat dissatisfied
    - ○ Neither satisfied nor dissatisfied
    - ○ Somewhat satisfied
    - ○ Very satisfied
    - ○ Don't know

22. How satisfied or dissatisfied were you with the instructions you received from doctors, nurses, and therapists about your treatment?
    - ○ Very dissatisfied
    - ○ Somewhat dissatisfied
    - ○ Neither satisfied nor dissatisfied
    - ○ Somewhat satisfied
    - ○ Very satisfied

23. How satisfied or dissatisfied were you with your overall recovery from joint replacement surgery since you left the hospital?
    - ○ Very dissatisfied
    - ○ Somewhat dissatisfied
    - ○ Neither satisfied nor dissatisfied
    - ○ Somewhat satisfied
    - ○ Very satisfied

Section 4. How are you Feeling Today?

24. In the past week, how much does pain in the joint that you had replaced currently interfere with your normal activities?
    - ○ All of the time
    - ○ Most of the time
    - ○ Some of the time
    - ○ A little of the time
    - ○ None of the time
    - ○ Don't know/Don't remember

25. Thinking about the past week, have you been taking any of the following types of medications specifically for pain in the joint you had replaced?
    - ○ Prescription pain medication only
    - ○ Over the counter pain medication only
    - ○ Both prescription and over the counter pain medications
    - ○ No medication for pain in the joint that was replaced
    - ○ Don't know/Don't remember

26. What best describes your use of a mobility aid over the past week, such as a wheelchair, scooter, walker or cane?
    - ○ I never use a mobility aid
    - ○ I sometimes use a mobility aid
    - ○ I always use a mobility aid
    - ○ Don't know/Don't remember

27. What best describes your current ability to walk by yourself without resting? That is, without the help of another person or the help of a mobility aid?
    - ○ I can walk more than several blocks by myself without resting
    - ○ I can walk several blocks by myself without resting
    - ○ I can walk one block by myself without resting
    - ○ I can walk from one room to another by myself without resting
    - ○ I am not able to walk by myself without resting
    - ○ Don't know/Don't remember

28. How much difficult do you currently have walking up or down 12 stairs?
    - ○ I have no difficulty walking up or down 12 stairs
    - ○ I have some difficulty walking up or down 12 stairs
    - ○ I have a lot of difficulty walking up or down 12 stairs
    - ○ I am not able to walk up or down 12 stairs
    - ○ Don't know/Don't remember

29. Continuing to think about the past week, how much difficulty did you have rising from sitting?
    - ○ Extreme
    - ○ Severe
    - ○ Moderate
    - ○ Mild
    - ○ None
    - ○ Don't know/Don't remember

30. Continuing to think about the past week, how much difficulty did you have standing?
    - ○ Extreme
    - ○ Severe
    - ○ Moderate
    - ○ Mild
    - ○ None
    - ○ Don't know/Don't remember

31. Continuing to think about the past week, how much difficulty did you have getting on/off toilet?
    - ○ Extreme
    - ○ Severe
    - ○ Moderate
    - ○ Mild
    - ○ None
    - ○ Don't know/Don't remember

Section 5. About You

32. What is the highest grade or level of school that you completed?
    - ○ Some high school, but did not graduate
    - ○ High school graduate or GED
    - ○ Some college or 2-year degree
    - ○ 4-year college degree
    - ○ More than 4-year college degree
    - ○ I prefer not to answer

33. What was your total household income before taxes during the past 12 months?
    - ○ Less than $12,500
    - ○ $12,500-$19,999
    - ○ $20,000-$29,999
    - ○ $30,000-$49,999
    - ○ $50,000-$75,000
    - ○ Greater than $75,000
    - ○ I prefer not to answer

34. Are you of Hispanic, Latino, or Spanish origin?
- ○ No, not of Hispanic, Latino, or Spanish origin
- ○ Yes, of Hispanic, Latino, or Spanish origin
- ○ I prefer not to answer

35. What is your race? **Choose all that apply**.
- ○ White
- ○ Black or African American
- ○ American Indian or Alaska Native
- ○ Asian
- ○ Native Hawaiian or Other Pacific Islander
- ○ I prefer not to answer

Appendix I: Patient Survey Results

Exhibit I-1: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, LEJR patients discharged from mandatory hospitals

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Change in functional status and pain[a]	Ability to walk by yourself without resting	-4 to 4	10,650	11,142	0.76	0.74	0.02 (0.9%)	p=0.19
	Difficulty walking up or down 12 stairs	-3 to 3	10,063	10,558	0.75	0.75	-0.01 (-0.2%)	p=0.72
	Difficulty rising from sitting	-4 to 4	10,843	11,348	1.22	1.21	0.00 (0.1%)	p=0.90
	Difficulty standing	-4 to 4	10,867	11,381	1.17	1.18	-0.01 (-0.4%)	p=0.52
	Use of a mobility aid	-2 to 2	10,581	11,031	0.16	0.16	0.00 (0.1%)	p=0.93
	Difficulty getting on/off the toilet	-4 to 4	10,819	11,328	1.36	1.37	-0.01 (-0.2%)	p=0.74
	Frequency that pain interferes with normal activities	-4 to 4	10,905	11,410	1.98	1.97	0.01 (0.4%)	p=0.67
	Medication use for pain in the joint you had replaced	-3 to 3	10,539	11,047	0.60	0.57	0.03 (1.1%)	p<0.10
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	11,000	11,499	80.7	80.4	0.3	p=0.68
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	10,570	11,048	83.0	83.0	0.0	p=0.99
	Health care providers listened to preferences	0 to 100	10,931	11,419	78.6	78.7	-0.1	p=0.86
	Satisfaction with discharge destination	0 to 100	10,906	11,410	82.4	81.9	0.5	p=0.49
	Satisfaction with care coordination	0 to 100	10,965	11,437	82.4	82.6	-0.3	p=0.67
	Satisfaction with treatment instructions	0 to 100	10,971	11,457	84.2	84.4	-0.3	p=0.65
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	10,873	11,329	88.5	88.5	0.0	p=1.00
	Received the right amount of post-discharge care	0 to 100	10,956	11,411	84.8	85.2	-0.4	p=0.48
	Had all the medical equipment needed at home	0 to 100	10,255	10,630	91.5	92.0	-0.5	p=0.32

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Caregiver help	Received any caregiver help[c]	0 to 100	10,581	10,977	95.2	95.9	-0.6	p=0.25
	Composite measure of caregiver help[d]	0 to 100	10,013	10,333	67.9	69.1	-1.3	p<0.05
	Help needed putting on or taking off clothes[d]	0 to 100	10,544	10,946	59.8	61.5	-1.7	p<0.10
	Help needed bathing[d]	0 to 100	10,481	10,894	65.0	66.0	-1.1	p=0.101
	Help needed using the toilet[d]	0 to 100	10,093	10,433	80.3	81.4	-1.1	p<0.05

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

LEJR = lower extremity joint replacement.

[a] The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Estimated changes, and the difference between changes in the CJR and control group, are reported in "level" terms (that is, levels of the Likert scale for each measure). Percentage differences are equal to the difference between CJR and control groups divided by the average CJR recalled status prior to the hospitalization.

[b] Satisfaction outcomes are scaled from 0 to 100 points, where 0 = very dissatisfied, 25 = dissatisfied, 50 = neutral, 75 = satisfied, and 100 = very satisfied. The composite summarizes the level of satisfaction across the four measures of care management. Differences between CJR and control outcomes are reported in point terms.

[c] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

[d] Respondents were only asked about the amount of help needed with a given activity of daily living if they indicated that they received caregiver help. Measures of caregiver help required among respondents who received any help are scaled from 0 to 100 points, where 0 = complete help needed, 50 = some help needed, and 100 = no help needed. The composite summarizes the amount of help needed across all three activities of daily living. Differences between CJR and control outcomes are reported in point terms.

Exhibit I-2: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, LEJR patients with hip fractures discharged from mandatory hospitals

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Change in functional status and pain[a]	Ability to walk by yourself without resting	-4 to 4	922	924	-0.72	-0.71	-0.01 (-0.4%)	p=0.86
	Difficulty walking up or down 12 stairs	-3 to 3	846	858	-0.58	-0.50	-0.08 (-3.0%)	p=0.30
	Difficulty rising from sitting	-4 to 4	941	964	-0.37	-0.29	-0.08 (-2.2%)	p<0.05
	Difficulty standing	-4 to 4	948	969	-0.34	-0.20	-0.14 (-3.6%)	p<0.01
	Use of a mobility aid	-2 to 2	930	934	-0.61	-0.60	-0.01 (-0.6%)	p=0.64
	Difficulty getting on/off the toilet	-4 to 4	930	962	-0.17	0.03	-0.20 (-5.0%)	p<0.01
	Frequency that pain interferes with normal activities	-4 to 4	930	961	-0.39	-0.37	-0.02 (-0.4%)	p=0.81
	Medication use for pain in the joint you had replaced	-3 to 3	904	931	-0.31	-0.32	0.01 (0.1%)	p=0.93
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	1,015	1,036	73.2	75.4	-2.2	p=0.17
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	970	994	75.4	76.9	-1.6	p=0.37
	Health care providers listened to preferences	0 to 100	1,017	1,036	72.4	73.8	-1.3	p=0.42
	Satisfaction with discharge destination	0 to 100	1,026	1,038	73.2	73.1	0.1	p=0.96
	Satisfaction with care coordination	0 to 100	1,002	1,043	75.8	75.7	0.1	p=0.95
	Satisfaction with treatment instructions	0 to 100	1,020	1,041	76.8	79.3	-2.5	p<0.10
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	958	972	84.7	85.9	-1.2	p=0.37
	Received the right amount of post-discharge care	0 to 100	1,018	1,009	76.6	79.6	-3.0	p<0.05
	Had all the medical equipment needed at home	0 to 100	896	907	87.2	87.4	-0.2	p=0.88

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Caregiver help	Received any caregiver help[c]	0 to 100	931	934	96.5	96.5	-0.0	p=0.99
	Composite measure of caregiver help[d]	0 to 100	872	871	52.4	58.0	-5.6	p<0.01
	Help needed putting on or taking off clothes[d]	0 to 100	917	930	48.1	53.5	-5.3	p<0.05
	Help needed bathing[d]	0 to 100	907	924	45.8	52.1	-6.2	p<0.01
	Help needed using the toilet[d]	0 to 100	882	888	64.0	70.2	-6.2	p<0.01

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

LEJR = lower extremity joint replacement.

[a] The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Estimated changes, and the difference between changes in the CJR and control group, are reported in "level" terms (that is, levels of the Likert scale for each measure). Percentage differences are equal to the difference between CJR and control groups divided by the average CJR recalled status prior to the hospitalization.

[b] Satisfaction outcomes are scaled from 0 to 100 points, where 0 = very dissatisfied, 25 = dissatisfied, 50 = neutral, 75 = satisfied, and 100 = very satisfied. The composite summarizes the level of satisfaction across the four measures of care management. Differences between CJR and control outcomes are reported in point terms.

[c] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

[d] Respondents were only asked about the amount of help needed with a given activity of daily living if they indicated that they received caregiver help. Measures of caregiver help required among respondents who received any help are scaled from 0 to 100 points, where 0 = complete help needed, 50 = some help needed, and 100 = no help needed. The composite summarizes the amount of help needed across all three activities of daily living. Differences between CJR and control outcomes are reported in point terms.

Exhibit I-3: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, LEJR patients from mandatory hospitals – excluding BPCI Advanced episodes from the control group

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Change in functional status and pain[a]	Ability to walk by yourself without resting	-4 to 4	10,650	5,936	0.76	0.72	0.04 (1.5%)	p<0.10
	Difficulty walking up or down 12 stairs	-3 to 3	10,063	5,614	0.74	0.75	0.00 (-0.1%)	p=0.88
	Difficulty rising from sitting	-4 to 4	10,843	6,054	1.21	1.21	0.01 (0.3%)	p=0.79
	Difficulty standing	-4 to 4	10,867	6,065	1.17	1.17	-0.01 (-0.2%)	p=0.78
	Use of a mobility aid	-2 to 2	10,581	5,884	0.16	0.15	0.00 (0.2%)	p=0.84
	Difficulty getting on/off the toilet	-4 to 4	10,819	6,042	1.36	1.36	0.00 (0.1%)	p=0.91
	Frequency that pain interferes with normal activities	-4 to 4	10,905	6,076	1.98	1.96	0.02 (1.0%)	p=0.39
	Medication use for pain in the joint you had replaced	-3 to 3	10,539	5,889	0.60	0.56	0.04 (1.5%)	p<0.05
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	11,000	6,154	80.7	80.4	0.3	p=0.73
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	10,570	5,927	82.7	82.8	-0.1	p=0.93
	Health care providers listened to preferences	0 to 100	10,931	6,122	78.7	78.6	0.1	p=0.87
	Satisfaction with discharge destination	0 to 100	10,906	6,100	82.5	81.4	1.1	p=0.12
	Satisfaction with care coordination	0 to 100	10,965	6,122	82.4	82.8	-0.4	p=0.52
	Satisfaction with treatment instructions	0 to 100	10,971	6,134	84.2	84.6	-0.4	p=0.52
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	10,873	6,040	88.4	88.2	0.2	p=0.85
	Received the right amount of post-discharge care	0 to 100	10,956	6,094	84.7	85.5	-0.8	p=0.13
	Had all the medical equipment needed at home	0 to 100	10,255	5,675	91.4	91.7	-0.3	p=0.66

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Caregiver help	Received any caregiver help[c]	0 to 100	10,581	5,853	95.2	95.6	-0.4	p=0.55
	Composite measure of caregiver help[d]	0 to 100	10,013	5,509	68.0	69.1	-1.2	p<0.10
	Help needed putting on or taking off clothes[d]	0 to 100	10,544	5,838	59.8	61.4	-1.6	p<0.10
	Help needed bathing[d]	0 to 100	10,481	5,805	64.9	66.3	-1.4	p<0.10
	Help needed using the toilet[d]	0 to 100	10,093	5,569	80.2	81.4	-1.1	p<0.10

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates excluded any control LEJR performed by a hospital or physician group practice participating in the Major Joint Replacement of the Lower Extremity bundle of the BPCI Advanced model. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

BPCI = Bundled Payments for Care Improvement.

LEJR = lower extremity joint replacement.

[a] The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Estimated changes, and the difference between changes in the CJR and control group, are reported in "level" terms (that is, levels of the Likert scale for each measure). Percentage differences are equal to the difference between CJR and control groups divided by the average CJR recalled status prior to the hospitalization.

[b] Satisfaction outcomes are scaled from 0 to 100 points, where 0 = very dissatisfied, 25 = dissatisfied, 50 = neutral, 75 = satisfied, and 100 = very satisfied. The composite summarizes the level of satisfaction across the four measures of care management. Differences between CJR and control outcomes are reported in point terms.

[c] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

[d] Respondents were only asked about the amount of help needed with a given activity of daily living if they indicated that they received caregiver help. Measures of caregiver help required among respondents who received any help are scaled from 0 to 100 points, where 0 = complete help needed, 50 = some help needed, and 100 = no help needed. The composite summarizes the amount of help needed across all three activities of daily living. Differences between CJR and control outcomes are reported in point terms.

Exhibit I-4: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, LEJR patients with hip fractures discharged from mandatory hospitals – excluding BPCI Advanced episodes from the control group

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Change in functional status and pain[a]	Ability to walk by yourself without resting	-4 to 4	922	587	-0.73	-0.73	-0.00 (-0.1%)	p=0.98
	Difficulty walking up or down 12 stairs	-3 to 3	846	546	-0.58	-0.52	-0.07 (-2.4%)	p=0.45
	Difficulty rising from sitting	-4 to 4	941	611	-0.37	-0.32	-0.06 (-1.5%)	p=0.28
	Difficulty standing	-4 to 4	948	611	-0.34	-0.21	-0.12 (-3.2%)	p<0.05
	Use of a mobility aid	-2 to 2	930	600	-0.61	-0.62	0.01 (0.6%)	p=0.75
	Difficulty getting on/off the toilet	-4 to 4	930	607	-0.13	0.02	-0.15 (-3.7%)	p<0.10
	Frequency that pain interferes with normal activities	-4 to 4	930	606	-0.39	-0.41	0.03 (0.6%)	p=0.72
	Medication use for pain in the joint you had replaced	-3 to 3	904	595	-0.32	-0.36	0.04 (1.2%)	p=0.44
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	1,015	659	73.4	76.0	-2.6	p=0.16
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	970	636	75.4	77.6	-2.2	p=0.21
	Health care providers listened to preferences	0 to 100	1,017	668	72.8	74.4	-1.6	p=0.48
	Satisfaction with discharge destination	0 to 100	1,026	659	73.5	72.9	0.6	p=0.74
	Satisfaction with care coordination	0 to 100	1,002	666	76.2	77.4	-1.2	p=0.46
	Satisfaction with treatment instructions	0 to 100	1,020	667	76.8	81.8	-5.0	p<0.01
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	958	616	84.8	83.6	1.2	p=0.49
	Received the right amount of post-discharge care	0 to 100	1,018	639	76.7	78.6	-1.9	p=0.31
	Had all the medical equipment needed at home	0 to 100	896	579	86.5	88.1	-1.6	p=0.34

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Caregiver help	Received any caregiver help[c]	0 to 100	931	592	96.4	95.5	0.9	p=0.38
	Composite measure of caregiver help[d]	0 to 100	872	556	53.0	58.8	-5.8	p<0.01
	Help needed putting on or taking off clothes[d]	0 to 100	917	591	48.7	52.9	-4.2	p<0.10
	Help needed bathing[d]	0 to 100	907	586	46.4	52.5	-6.1	p<0.01
	Help needed using the toilet[d]	0 to 100	882	568	64.5	71.0	-6.5	p<0.01

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates excluded any control LEJR performed by a hospital or physician group practice participating in the Major Joint Replacement of the Lower Extremity bundle of the BPCI Advanced model. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

BPCI = Bundled Payments for Care Improvement.

LEJR = lower extremity joint replacement.

[a] The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Estimated changes, and the difference between changes in the CJR and control group, are reported in "level" terms (that is, levels of the Likert scale for each measure). Percentage differences are equal to the difference between CJR and control groups divided by the average CJR recalled status prior to the hospitalization.

[b] Satisfaction outcomes are scaled from 0 to 100 points, where 0 = very dissatisfied, 25 = dissatisfied, 50 = neutral, 75 = satisfied, and 100 = very satisfied. The composite summarizes the level of satisfaction across the four measures of care management. Differences between CJR and control outcomes are reported in point terms.

[c] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

[d] Respondents were only asked about the amount of help needed with a given activity of daily living if they indicated that they received caregiver help. Measures of caregiver help required among respondents who received any help are scaled from 0 to 100 points, where 0 = complete help needed, 50 = some help needed, and 100 = no help needed. The composite summarizes the amount of help needed across all three activities of daily living. Differences between CJR and control outcomes are reported in point terms.

Exhibit I-5: Risk-adjusted survey-based results for change in functional status, decomposed by amount of change, LEJR patients with hip fractures discharged from mandatory participating hospitals

Change in response category from the week prior to hospitalization through the time of the survey		Difficulty rising from sitting			Difficulty standing			Difficulty getting on or off the toilet		
		CJR	Control	Difference (pp)	CJR	Control	Difference (pp)	CJR	Control	Difference (pp)
Declined	-4	1.38	1.17	0.20	1.51	1.16	0.35	1.69	1.17	0.53
	-3	3.23	2.80	0.43	3.25	2.56	0.69	2.56	1.82	0.75
	-2	13.67	12.38	1.30	12.35	10.32	2.03	9.75	7.39	2.36
	-1	24.17	23.36	0.82	22.34	20.60	1.74	18.18	15.52	2.66
Regained prior status or improved	0	38.60	39.75	-1.15	44.06	46.51	-2.45	47.48	49.72	-2.24
	1	11.65	12.46	-0.81	8.75	9.76	-1.01	10.09	11.66	-1.57
	2	4.73	5.17	-0.44	4.62	5.28	-0.65	6.62	7.95	-1.33
	3	1.92	2.16	-0.24	1.97	2.36	-0.39	2.27	2.89	-0.62
	4	0.64	0.75	-0.11	1.14	1.44	-0.30	1.37	1.89	-0.52

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional ordered logistic regression model, weighted for sampling and nonresponse.

The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Each of the three functional status measures in this exhibit were based on items with five response categories. Respondents could therefore improve or decline up to four categories of function (e.g., -4 represents a change from no difficulty with the activity to unable to perform the activity). Each cell indicates the proportion of CJR and control respondents who reported a given amount of change from the week prior to hospitalization through the time of the survey, and the difference in proportions between CJR and control respondents.

LEJR = lower extremity joint replacement, pp = percentage point.

Exhibit I-6: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, LEJR patients with hip fractures discharged from mandatory participating hospitals in *preceding* survey waves

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Change in functional status and pain[a]	Ability to walk by yourself without resting	-4 to 4	645	717	-0.67	-0.65	-0.03 (-0.9%)	p=0.60
	Difficulty walking up or down 12 stairs	-3 to 3	586	656	-0.48	-0.50	0.02 (0.8%)	p=0.62
	Difficulty rising from sitting	-4 to 4	666	747	-0.27	-0.21	-0.06 (-1.7%)	p=0.27
	Difficulty standing	-4 to 4	670	750	-0.27	-0.17	-0.10 (-2.5%)	p<0.10
	Use of a mobility aid	-2 to 2	652	716	-0.56	-0.55	-0.01 (-0.5%)	p=0.76
	Difficulty getting on/off the toilet	-4 to 4	662	734	-0.06	-0.02	-0.04 (-1.0%)	p=0.39
	Frequency that pain interferes with normal activities	-4 to 4	657	750	-0.29	-0.24	-0.06 (-1.4%)	p=0.40
	Medication use for pain in the joint you had replaced	-3 to 3	637	736	-0.27	-0.29	0.02 (0.5%)	p=0.73
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	716	811	74.9	75.3	-0.4	p=0.81
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	673	767	75.8	76.7	-0.9	p=0.63
	Health care providers listened to preferences	0 to 100	711	803	71.6	73.6	-2.0	p=0.30
	Satisfaction with discharge destination	0 to 100	729	816	74.0	72.9	1.1	p=0.64
	Satisfaction with care coordination	0 to 100	725	814	74.6	75.9	-1.3	p=0.52
	Satisfaction with treatment instructions	0 to 100	706	800	77.0	77.8	-0.7	p=0.66
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	678	772	84.8	85.4	-0.7	p=0.67
	Received the right amount of post-discharge care	0 to 100	704	806	78.0	81.6	-3.6	p<0.10
	Had all the medical equipment needed at home	0 to 100	643	708	89.5	91.3	-1.9	p=0.20

Domain	Measure	Range	CJR respondents (N)	Control respondents (N)	CJR risk-adjusted average	Control risk-adjusted average	Estimated difference	p-value
Caregiver help	Received any caregiver help[c]	0 to 100	662	728	93.8	95.5	-1.72	p=0.18
	Composite measure of caregiver help[d]	0 to 100	607	666	52.9	55.7	-2.75	p<0.10
	Help needed putting on or taking off clothes[d]	0 to 100	655	721	49.4	51.8	-2.4	p=0.18
	Help needed bathing[d]	0 to 100	653	716	46.6	49.2	-2.7	p=0.13
	Help needed using the toilet[d]	0 to 100	623	671	64.9	67.8	-3.0	p=0.13

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, August, or September 2018.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively. These waves occurred during the Bundled Payments for Care Improvement (BPCI) Models 2-4, and as specified in the Final Rule, episodes attributed to BPCI Models 2-4 were excluded from the analysis.

LEJR = lower extremity joint replacement.

[a] The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Estimated changes, and the difference between changes in the CJR and control group, are reported in "level" terms (that is, levels of the Likert scale for each measure). Percentage differences are equal to the difference between CJR and control groups divided by the average CJR recalled status prior to the hospitalization.

[b] Satisfaction outcomes are scaled from 0 to 100 points, where 0 = very dissatisfied, 25 = dissatisfied, 50 = neutral, 75 = satisfied, and 100 = very satisfied. The composite summarizes the level of satisfaction across the four measures of care management. Differences between CJR and control outcomes are reported in point terms.

[c] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

[d] Respondents were only asked about the amount of help needed with a given activity of daily living if they indicated that they received caregiver help. Measures of caregiver help required among respondents who received any help are scaled from 0 to 100 points, where 0 = complete help needed, 50 = some help needed, and 100 = no help needed. The composite summarizes the amount of help needed across all three activities of daily living. Differences between CJR and control outcomes are reported in point terms.

Appendix J: Change in Patient Characteristics

Exhibit J-1: Change in patient complexity measures, LEJR patients whose first PAC setting was an inpatient rehabilitation facility, PY1-4

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90% CI
		Baseline average (N=19,802)	*Intervention average (N=9,172)*	*Baseline average (N=21,278)*	*Intervention average (N=12,548)*				
IRF	Age 80+	42.7%	49.2%	39.7%	41.7%	4.6	10.7%	p<0.05	[1.4 to 7.7]
	Female	71.1%	68.9%	71.0%	70.4%	-1.6	-2.3%	p<0.10	[-3.2 to -0.1]
	Black or African American[a]	7.9%	6.9%	8.3%	7.9%	-0.7	-8.4%	p=0.38	[-1.9 to 0.6]
	Hispanic[a]	7.1%	6.9%	3.1%	2.8%	0.2	2.3%	p=0.78	[-0.8 to 1.1]
	Eligible for Medicaid	16.2%	15.4%	12.1%	10.7%	0.7	4.2%	p=0.49	[-0.9 to 2.3]
	Disability, no ESRD	16.1%	16.0%	16.4%	15.1%	1.2	7.4%	p=0.16	[-0.2 to 2.6]
	HCC Score	1.8	2.2	1.8	2.0	0.1	0.1%	p<0.01	[0.1 to 0.2]
	Obesity	16.3%	23.6%	18.4%	28.0%	-2.3	-14.0%	p=0.16	[-5.0 to 0.4]
	Diabetes	34.1%	32.7%	32.4%	31.3%	-0.3	-0.9%	p=0.71	[-1.8 to 1.1]
	Hypertension	79.1%	79.4%	80.3%	81.0%	-0. 4	-0.5%	p=0.59	[-1.7 to 0.9]
	Dementia	8.8%	11.7%	9.2%	10.2%	1.9	21.0%	p<0.01	[0.7 to 3.0]
	CHF	20.1%	22.2%	20.0%	21.4%	0.6	2.8%	p=0.61	[-1.3 to 2.4]
	Prior ACH stay	15.3%	17.7%	16.7%	17.6%	1.6	10.6%	p<0.05	[0.4 to 2.8]
	Prior IRF stay	5.3%	6.6%	6.2%	7.1%	0.4	7.5%	p=0.34	[-0.3 to 1.1]
	Prior SNF stay	3.7%	4.7%	3.7%	4.2%	0.5	13.0%	p=0.20	[-0.1 to 1.1]
	Prior HH use	16.8%	19.9%	17.4%	19.8%	0.7	4.2%	p=0.50	[-1.0 to 2.4]
	Any prior care	35.6%	41.0%	37.8%	41.2%	1.9	5.5%	p<0.10	[0.2 to 3.7]
	MS-DRG 469	8.0%	12.4%	8.3%	10.1%	2.5	31.6%	p<0.01	[1.4 to 3.7]
	Hip fracture	32.2%	51.3%	31.5%	40.1%	10.5	32.8%	p<0.01	[6.0 to 15.1]
	Mobility index[b]	8.1	7.3	8.2	7.8	-0.3	-4.2%	p<0.05	[-0.6 to -0.1]
	Self-care index[b]	20.6	19.1	21.0	19.9	-0.4	-2.1%	p=0.25	[-1.1 to 0.2]
	Cognitive index[b]	26.1	24.1	25.4	23.8	-0.2	-0.9%	p=0.62	[-1.1 to 0.6]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 and IRF-PAI data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by September 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of calculating the net DiD of the unadjusted baseline and intervention averages for the CJR and control groups. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the net differences estimate as a percent of the CJR baseline level.

ACH = acute care hospital, CHF = congestive heart failure, CI = confidence interval, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument, LEJR = lower extremity joint replacement, MS-DRG = Medicare Severity-Diagnosis Related Group, PAC = post-acute care, PY = performance year, SNF = skilled nursing facility.

[a] The number of episodes for these measures is lower because episodes were dropped to account for missing observations. These measures are based on 19,760 CJR baseline episodes, 9,119 CJR intervention episodes, 21,243 control group baseline episodes, and 12,471 control group intervention episodes.

[b] The number of episodes for these measures is lower because the intervention period is one quarter shorter than it is for the claims-based analyses because of the longer time needed for PAC assessment data to become available. Further, not all beneficiary stays were matched to an IRF-PAI assessment. These measures are based on 17,609 CJR baseline episodes, 7,547 CJR intervention episodes, 19,781 control group baseline episodes, and 10,631 control group intervention episodes.

Exhibit J-2: Change in patient complexity measures, LEJR patients whose first PAC setting was a skilled nursing facility, PY1-4

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90% CI
		Baseline average (N=61,057)	Intervention average (N=44,600)	Baseline average (N=68,716)	Intervention average (N=45,573)				
SNF	Age 80+	40.5%	45.1%	38.4%	41.4%	1.6	3.9%	p<0.10	[0.0 to 3.2]
	Female	72.8%	73.4%	73.0%	73.5%	0.1	0.1%	p=0.91	[-0.9 to 1.0]
	Black or African American [a]	6.8%	6.7%	7.6%	8.3%	-0.8	-11.9%	p=0.12	[-1.7 to 0.0]
	Hispanic [a]	6.5%	6.0%	2.6%	2.8%	-0.7	-10.6%	p<0.10	[-1.4 to -0.0]
	Eligible for Medicaid	18.6%	18.4%	13.7%	14.3%	-0.8	-4.4%	p=0.29	[-2.1 to 0.5]
	Disability, no ESRD	14.9%	15.4%	14.9%	15.5%	-0.2	-1.1%	p=0.77	[-1.1 to 0.8]
	HCC Score	1.7	2.0	1.6	1.9	0.0	0.0%	p=0.44	[-0.0 to 0.1]
	Obesity	16.3%	28.6%	16.8%	29.8%	-0.7	-4.4%	p=0.67	[-3.4 to 2.0]
	Diabetes	33.5%	34.5%	29.4%	29.8%	0.7	2.1%	p=0.28	[-0.4 to 1.8]
	Hypertension	78.9%	80.8%	77.9%	79.6%	0.2	0.2%	p=0.77	[-0.8 to 1.1]
	Dementia	12.9%	15.4%	12.4%	13.8%	1.1	8.2%	p=0.25	[-0.5 to 2.6]
	CHF	18.9%	20.9%	17.3%	19.1%	0.2	1.1%	p=0.80	[-1.1 to 1.5]
	Prior ACH stay	15.6%	17.4%	14.8%	16.3%	0.3	1.7%	p=0.57	[-0.5 to 1.1]
	Prior IRF stay	1.1%	1.3%	1.1%	1.3%	-0.1	-7.0%	p=0.69	[-0.4 to 0.2]
	Prior SNF stay	9.5%	10.8%	9.0%	10.3%	0.0	-0.3%	p=0.91	[-0.5 to 0.4]
	Prior HH use	16.4%	18.1%	15.3%	17.7%	-0.7	-4.1%	p=0.51	[-2.4 to 1.0]
	Any prior care	35.9%	39.6%	35.3%	39.4%	-0.4	-1.0%	p=0.72	[-2.1 to 1.3]
	MS-DRG 469	7.2%	9.5%	6.7%	8.6%	0.4	5.6%	p=0.56	[-0.7 to 1.6]
	Hip fracture	24.9%	32.8%	21.3%	27.6%	1.6	6.5%	p=0.42	[-1.7 to 4.9]
	Bathing poor [b]	87.9%	87.2%	80.9%	79.9%	0.2	0.3%	p=0.91	[-3.3 to 3.8]
	Cognition not intact [b]	20.1%	21.6%	19.7%	20.8%	0.4	2.2%	p=0.72	[-1.6 to 2.4]
	Severe cognitive impairment [b]	6.3%	7.8%	6.0%	7.2%	0.2	3.6%	p=0.63	[-0.6 to 1.0]
	Early-loss ADL score [b]	5.1	5.1	4.7	4.6	0.1	1.0%	p=0.39	[-0.0 to 0.1]
	Mid-loss ADL score (i.e., motion score) [b]	8.4	8.4	7.8	7.7	0.1	1.5%	p=0.22	[-0.0 to 0.3]

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90% CI
		Baseline average (N=61,057)	Intervention average (N=44,600)	Baseline average (N=68,716)	Intervention average (N=45,573)				
SNF (cont'd)	Late-loss ADL score[b]	13.5	13.5	12.7	12.6	0.1	0.9%	p=0.37	[-0.1 to 0.4]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 and MDS data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by September 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of calculating the net difference-in-differences of the unadjusted baseline and intervention averages for the CJR and control groups. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the net differences estimate as a percent of the CJR baseline level.

ACH = acute care hospital, ADL = activities of daily living, CHF = congestive heart failure, CI = confidence interval, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, MS-DRG = Medicare Severity-Diagnosis Related Group, PAC = post-acute care, PY = performance year, SNF = skilled nursing facility.

[a] The number of episodes for these measures is lower because episodes were dropped to account for missing observations. These measures are based on 60,860 CJR baseline episodes, 44,232 CJR intervention episodes, 68,557 control group baseline episodes, and 45,278 control group intervention episodes.

[b] The number of episodes for these measures is lower because the intervention period is one quarter shorter than it is for the claims-based analyses because of the longer time needed for PAC assessment data to become available. Further, not all beneficiary stays were matched to a MDS admission assessment. These measures are based on 57,809 CJR baseline episodes, 39,362 CJR intervention episodes, 64,657 control group baseline episodes, and 39,397 control group intervention episodes.

Exhibit J-3: Change in patient complexity measures, LEJR patients whose first PAC setting was home with home health agency care, PY1-4

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90% CI
		Baseline average (N=53,401)	Intervention average (N=76,487)	Baseline average (N=64,134)	Intervention average (N=70,018)				
HHA	Age 80+	14.5%	16.1%	12.7%	14.0%	0.3	2.0%	p=0.56	[-0.5 to 1.1]
	Female	59.2%	61.9%	59.1%	61.4%	0.4	0.6%	p=0.53	[-0.6 to 1.3]
	Black or African American[a]	5.8%	5.5%	7.3%	7.1%	-0.1	-1.2%	p=0.88	[-0.9 to 0.7]
	Hispanic[a]	5.3%	5.0%	3.2%	2.9%	0.0	0.4%	p=0.96	[-0.7 to 0.7]
	Eligible for Medicaid	10.1%	8.6%	9.4%	8.0%	-0.1	-0.8%	p=0.91	[-1.3 to 1.2]
	Disability, no ESRD	16.5%	14.4%	17.5%	15.4%	0.0	0.2%	p=0.96	[-1.2 to 1.3]
	HCC Score	1.1	1.2	1.1	1.2	0.0	0.0%	p=0.72	[-0.0 to 0.1]
	Obesity	14.9%	32.3%	16.1%	32.8%	0.8	5.3%	p=0.76	[-3.5 to 5.0]
	Diabetes	25.3%	26.6%	24.9%	24.1%	2.1	8.5%	p<0.01	[1.1 to 3.2]
	Hypertension	71.3%	73.1%	72.5%	72.9%	1.5	2.0%	p<0.05	[0.4 to 2.5]
	Dementia	2.2%	2.1%	2.2%	2.1%	0.0	-2.2%	p=0.80	[-0.4 to 0.3]
	CHF	9.5%	10.3%	9.7%	9.9%	0.6	6.1%	p=0.16	[-0.1 to 1.3]
	Prior ACH stay	10.0%	9.5%	10.1%	9.7%	-0.2	-1.8%	p=0.60	[-0.7 to 0.4]
	Prior IRF stay	0.4%	0.4%	0.4%	0.5%	0.0	-6.7%	p=0.70	[-0.1 to 0.1]
	Prior SNF stay	1.1%	0.9%	1.0%	1.0%	-0.2	-15.4%	p<0.10	[-0.3 to -0.0]
	Prior HH use	10.3%	9.5%	9.8%	9.7%	-0.6	-6.1%	p=0.28	[-1.6 to 0.3]
	Any prior care	23.3%	23.2%	23.4%	23.8%	-0.5	-2.2%	p=0.63	[-2.3 to 1.2]
	MS-DRG 469	1.6%	1.6%	1.7%	1.7%	0.0	-2.4%	p=0.83	[-0.3 to 0.3]
	Hip fracture	2.4%	2.6%	2.3%	2.6%	-0.1	-4.8%	p=0.74	[-0.7 to 0.4]
	Toilet transferring[b]	14.8%	28.0%	13.9%	29.0%	-1.9	-12.7%	p=0.49	[-6.4 to 2.7]
	Transferring[b]	30.0%	72.6%	29.8%	72.0%	0.4	1.2%	p=0.90	[-4.5 to 5.2]
	Ambulation / locomotion[b]	49.7%	82.0%	52.1%	81.9%	2.6	5.1%	p=0.26	[-1.2 to 6.3]
	Lower body dressing[b]	89.1%	95.0%	88.3%	94.7%	-0.5	-0.6%	p=0.60	[-2.1 to 1.1]
	Upper body dressing[b]	26.4%	48.6%	24.4%	44.1%	2.5	9.5%	p=0.24	[-1.0 to 6.0]

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90% CI
		Baseline average (N=53,401)	Intervention average (N=76,487)	Baseline average (N=64,134)	Intervention average (N=70,018)				
HHA cont'd	Bathing[b]	75.8%	89.5%	78.0%	88.5%	3.1	4.2%	p<0.10	[0.3 to 6.0]
	Toileting hygiene[b]	31.4%	65.0%	29.2%	61.2%	1.5	4.9%	p=0.59	[-3.2 to 6.3]
	Grooming[b]	21.3%	42.4%	20.6%	40.3%	1.4	6.5%	p=0.59	[-2.8 to 5.6]
	Cognitive functioning[b]	1.3%	1.7%	1.6%	1.9%	0.0	3.7%	p=0.89	[-0.5 to 0.6]
	Confusion[b]	1.2%	1.3%	1.3%	1.4%	-0.0	-3.0%	p=0.89	[-0.4 to 0.4]
	Memory deficit[b]	2.1%	2.1%	2.7%	2.5%	0.3	13.2%	p=0.38	[-0.2 to 0.8]
	Impaired decision-making[b]	4.2%	4.8%	4.8%	5.5%	-0.2	-3.7%	p=0.84	[-1.5 to 1.1]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 and OASIS data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by September 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of calculating the net DiD of the unadjusted baseline and intervention averages for the CJR and control groups. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The relative change from CJR baseline is calculated as the net differences estimate as a percent of the CJR baseline level.

ACH = acute care hospital, CHF = congestive heart failure, CI = confidence interval, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, HHA = home health agency, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, MS-DRG = Medicare Severity-Diagnosis Related Group, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, PY = performance year, SNF = skilled nursing facility.

[a] The number of episodes for these measures is lower because episodes were dropped to account for missing observations. These measures are based on 53,098 CJR baseline episodes, 75,040 CJR intervention episodes, 63,785 control group baseline episodes, and 68,892 control group intervention episodes.

[b] The number of episodes for these measures is lower because the intervention period is one quarter shorter than it is for the claims-based analyses because of the longer time needed for PAC assessment data to become available. Further, not all beneficiary stays were matched to an OASIS start of care assessment. These measures are based on 44,113 CJR baseline episodes, 40,259 CJR intervention episodes, 50,779 control group baseline episodes, and 37,896 control group intervention episodes.

Exhibit J-4: Change in patient complexity measures, elective MS-DRG 470 patients, PY1-4

Patient characteristics		CJR		Control group		Net differences (pp)	Net differences % of baseline	p-value	90% CI
		Baseline Average (N=119,239)	*Intervention Average (N=146,644)*	*Baseline Average (N=128,560)*	*Intervention Average (N=136,781)*				
Age	20-64	8.6%	6.7%	8.9%	7.1%	-0.1	-1.3%	p=0.76	[-0.7 to 0.5]
	65-79	70.1%	74.1%	71.7%	75.1%	0.7	1.0%	p=0.24	[-0.3 to 1.6]
	80+	21.3%	19.2%	19.4%	17.8%	-0.6	-2.7%	p=0.30	[-1.5 to 0.3]
Sex	Female	64.6%	63.7%	64.5%	63.7%	-0.1	-0.2%	p=0.67	[-0.7 to 0.4]
Race/ethnicity	White[a]	84.2%	85.6%	87.7%	87.8%	1.2	1.5%	p<0.05	[0.3 to 2.2]
	Black or African American[a]	7.0%	6.3%	7.8%	7.6%	-0.5	-7.5%	p=0.16	[-1.1 to 0.1]
	Hispanic[a]	6.1%	5.3%	3.1%	3.0%	-0.7	-12.2%	p=0.10	[-1.5 to 0.0]
	Other[b]	2.7%	2.9%	1.5%	1.6%	0.0	0.8%	p=0.90	[-0.3 to 0.3]
Medicaid	Eligible for Medicaid	12.9%	10.3%	10.2%	8.8%	-1.3	-9.8%	p<0.05	[-2.1 to -0.4]
Disability	Disability, no ESRD	16.6%	15.0%	16.9%	15.5%	-0.2	-1.2%	p=0.68	[-1.0 to 0.6]
Health Status	HCC score	1.3	1.3	1.2	1.3	-0.0	-0.0%	p=0.26	[-0.0 to 0.0]
	Obesity	17.6%	34.1%	18.1%	33.9%	0.7	4.2%	p=0.74	[-2.9 to 4.3]
	Diabetes	29.5%	28.5%	27.3%	25.8%	0.5	1.6%	p=0.22	[-0.2 to 1.1]
	Hypertension	75.2%	74.9%	75.3%	74.7%	0.3	0.4%	p=0.52	[-0.5 to 1.0]
	Dementia	3.2%	2.7%	3.2%	2.8%	-0.0	-0.7%	p=0.87	[-0.2 to 0.2]
	CHF	12.4%	11.7%	11.7%	11.4%	-0.3	-2.8%	p=0.32	[-0.9 to 0.2]
Utilization in the six months prior to the anchor hospitalization	ACH stay	11.2%	10.4%	11.2%	10.7%	-0.3	-2.5%	p=0.22	[-0.7 to 0.1]
	IRF stay	1.1%	0.7%	1.1%	0.9%	-0.2	-17.0%	p=0.14	[-0.4 to 0.0]
	SNF stay	3.6%	2.6%	3.2%	2.7%	-0.4	-10.6%	p<0.01	[-0.6 to -0.2]
	HH use	10.6%	9.3%	9.8%	9.1%	-0.7	-6.3%	p=0.18	[-1.5 to 0.1]
	Any prior care	26.4%	25.1%	26.2%	25.9%	-1.0	-3.7%	p<0.05	[-1.7 to -0.3]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of calculating the DiD of the unadjusted baseline and intervention averages for the CJR and control groups (net differences). Estimates that are significant at the 1%, 5%, or 10% significance levels are indicated by red, orange, or yellow shaded cells, respectively.

The MS-DRG 469 is assigned at the anchor hospitalization discharge for major joint replacement or reattachment of lower extremity *with* MCC, while MS-DRG 470 is *without* MCC.

Fracture is defined based on ICD codes for hip fracture provided by the CMMI on the CJR model website: https://innovation.cms.gov/Files/worksheets/cjr-icd10hipfracturecodes.xlsx.

ACH = acute care hospital, CHF = congestive heart failure, CI = confidence interval, CMMI = Center for Medicare & Medicaid Innovation, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, ICD = International Classification of Diseases, IRF = inpatient rehabilitation facility, MCC = major complications or comorbidities, MS-DRG = Medicare Severity Diagnosis Related Group, pp = percentage point, PY = performance year, SNF = skilled nursing facility.

[a] The number of episodes for these measures is lower because episodes were dropped to account for missing observations. These measures are based on 118,676 CJR baseline episodes, 146,020 CJR intervention episodes, 126,299 control group baseline episodes, and 134,750 control group intervention episodes.

[b] Other includes beneficiaries identified as "Asian," "American Indian/Alaska Native," or "Other."

Exhibit J-5: Change in patient complexity measures, elective MS-DRG 469 patients, PY1-4

Patient characteristics		CJR		Control group		Net difference s(pp)	Net differences % of baseline	p-value	90% CI
		Baseline Average (N=3,453)	*Intervention Average (N=4,086)*	*Baseline Average (N=3,218)*	*Intervention Average (N=3,747)*				
Age	20-64	12.0%	10.2%	11.1%	10.1%	-0.8	-6.6%	p=0.40	[-2.3 to 0.8]
	65-79	54.4%	58.2%	57.3%	63.2%	-2.1	-3.9%	p=0.18	[-4.7 to 0.5]
	80+	33.7%	31.6%	31.6%	26.7%	2.9	8.7%	p<0.10	[0.2 to 5.6]
Sex	Female	62.1%	62.6%	61.2%	61.1%	0.6	1.0%	p=0.72	[-2.1 to 3.3]
Race/ethnicity	White[a]	82.7%	83.6%	86.4%	86.5%	0.9	1.1%	p=0.44	[-1.0 to 2.8]
	Black or African American[a]	8.6%	7.5%	8.9%	9.2%	-1.4	-16.1%	p=0.11	[-2.8 to 0.0]
	Hispanic[a]	6.2%	6.5%	3.5%	2.8%	1.0	15.5%	p=0.20	[-0.3 to 2.2]
	Other[ab]	2.6%	2.4%	1.2%	1.5%	-0.5	-18.1%	p=0.20	[-1.1 to 0.1]
Medicaid	Eligible for Medicaid	22.1%	16.4%	16.4%	14.3%	-3.6	-16.2%	p<0.01	[-5.5 to -1.6]
Disability	Disability, no ESRD	23.1%	22.4%	21.3%	22.1%	-1.5	-6.6%	p=0.33	[-4.1 to 1.0]
Health Status	HCC score	2.1	2.2	2.0	2.1	0.0	0.0%	p=0.72	[-0.1 to 0.1]
	Obesity	21.7%	37.7%	23.2%	38.1%	1.1	5.0%	p=0.65	[-2.8 to 5.0]
	Diabetes	38.8%	36.0%	36.5%	33.6%	0.1	0.2%	p=0.97	[-3.0 to 3.2]
	Hypertension	82.2%	82.6%	81.9%	83.0%	-0.7	-0.8%	p=0.70	[-3.7 to 2.3]
	Dementia	8.9%	7.6%	9.3%	8.0%	0.1	1.3%	p=0.92	[-1.7 to 1.9]
	CHF	27.5%	26.5%	27.0%	24.7%	1.4	4.9%	p=0.38	[-1.2 to 3.9]
Utilization in the six months prior to the anchor hospitalization	ACH stay	20.2%	19.9%	18.5%	18.8%	-0.6	-2.9%	p=0.68	[-2.9 to 1.8]
	IRF stay	2.9%	2.6%	2.4%	2.5%	-0.5	-15.7%	p=0.34	[-1.2 to 0.3]
	SNF stay	8.7%	9.1%	7.9%	7.8%	0.5	6.1%	p=0.63	[-1.3 to 2.4]
	HH use	18.0%	19.5%	18.1%	18.2%	1.4	7.7%	p=0.50	[-2.0 to 4.7]
	Any prior care	40.6%	41.4%	39.9%	40.3%	0.4	1.0%	p=0.83	[-2.7 to 3.5]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of calculating the DiD of the unadjusted baseline and intervention averages for the CJR and control groups (net differences). Estimates that are significant at the 1%, 5%, or 10% significance levels are indicated by red, orange, or yellow shaded cells, respectively.

The MS-DRG 469 is assigned at the anchor hospitalization discharge for major joint replacement or reattachment of lower extremity *with* MCC, while MS-DRG 470 is *without* MCC.

Fracture is defined based on ICD codes for hip fracture provided by the CMMI on the CJR model website: https://innovation.cms.gov/Files/worksheets/cjr-icd10hipfracturecodes.xlsx.

ACH = acute care hospital, CHF = congestive heart failure, CI = confidence interval, CMMI = Center for Medicare & Medicaid Innovation, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, ICD = International Classification of Diseases, IRF = inpatient rehabilitation facility, MCC = major complications or comorbidities, MS-DRG = Medicare Severity Diagnosis Related Group, pp = percentage point, PY = performance year, SNF = skilled nursing facility.

[a] The number of episodes for these measures is lower because episodes were dropped to account for missing observations. These measures are based on 3,441 CJR baseline episodes, 4,075 CJR intervention episodes, 3,191 control group baseline episodes, and 3,704 control group intervention episodes.

[b] Other includes beneficiaries identified as "Asian," "American Indian/Alaska Native," or "Other."

Exhibit J-6: Change in patient complexity measures, fracture MS-DRG 470 patients, PY1-4

Patient characteristics		CJR		Control group		Net differences (pp)	Net differences % of baseline	p-value	90% CI
		Baseline Average (N=19,902)	*Intervention Average (N=19,931)*	*Baseline Average (N=18,100)*	*Intervention Average (N=16,810)*				
Age	20-64	2.9%	2.9%	3.2%	3.0%	0.1	5.1%	p=0.56	[-0.3 to 0.6]
	65-79	29.5%	33.1%	31.3%	35.5%	-0.7	-2.3%	p=0.33	[-1.8 to 0.5]
	80+	67.5%	64.0%	65.5%	61.5%	0.5	0.8%	p=0.46	[-0.6 to 1.7]
Sex	Female	75.2%	72.8%	75.1%	73.5%	-0.7	-1.0%	p=0.13	[-1.5 to 0.1]
Race/ethnicity	White[a]	88.4%	88.5%	91.7%	91.9%	-0.0	-0.0%	p=0.97	[-1.1 to 1.1]
	Black or African American[a]	3.5%	3.5%	4.9%	4.5%	0.4	11.1%	p=0.19	[-0.1 to 0.9]
	Hispanic[a]	5.1%	5.0%	2.0%	2.3%	-0.5	-8.9%	p=0.32	[-1.2 to 0.3]
	Other[ab]	2.9%	3.0%	1.3%	1.3%	0.1	3.3%	p=0.76	[-0.4 to 0.6]
Medicaid	Eligible for Medicaid	20.2%	18.4%	16.4%	14.4%	0.2	0.9%	p=0.76	[-0.8 to 1.2]
Disability	Disability, no ESRD	9.9%	11.1%	10.6%	11.4%	0.4	4.2%	p=0.35	[-0.3 to 1.1]
Health Status	HCC score	2.3	2.4	2.3	2.4	0.0	0.8%	p=0.43	[-0.0 to 0.1]
	Obesity	4.1%	9.2%	4.7%	9.6%	0.1	2.5%	p=0.89	[-1.1 to 1.3]
	Diabetes	28.9%	28.5%	26.4%	25.3%	0.8	2.7%	p=0.26	[-0.4 to 1.9]
	Hypertension	74.6%	75.9%	74.4%	75.7%	-0.1	-0.2%	p=0.85	[-1.4 to 1.1]
	Dementia	29.2%	27.1%	30.6%	27.4%	1.1	3.8%	p<0.10	[0.0 to 2.2]
	CHF	23.6%	22.6%	23.0%	21.9%	0.2	0.9%	p=0.78	[-1.0 to 1.4]
Utilization in the six months prior to the anchor hospitalization	ACH stay	20.9%	21.0%	21.3%	20.4%	1.0	4.8%	p=0.16	[-0.2 to 2.2]
	IRF stay	2.5%	2.8%	3.3%	3.2%	0.4	16.8%	p=0.20	[-0.1 to 1.0]
	SNF stay	11.5%	10.7%	11.3%	10.9%	-0.4	-3.1%	p=0.39	[-1.0 to 0.3]
	HH use	24.9%	23.6%	24.9%	24.2%	-0.7	-2.6%	p=0.55	[-2.4 to 1.1]
	Any prior care	46.6%	46.9%	48.2%	47.8%	0.7	1.6%	p=0.45	[-0.9 to 2.3]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of calculating the DiD of the unadjusted baseline and intervention averages for the CJR and control groups (net differences). Estimates that are significant at the 1%, 5%, or 10% significance levels are indicated by red, orange, or yellow shaded cells, respectively.

The MS-DRG 469 is assigned at the anchor hospitalization discharge for major joint replacement or reattachment of lower extremity *with* MCC, while MS-DRG 470 is *without* MCC.

Fracture is defined based on ICD codes for hip fracture provided by the CMMI on the CJR model website: https://innovation.cms.gov/Files/worksheets/cjr-icd10hipfracturecodes.xlsx.

ACH = acute care hospital, CHF = congestive heart failure, CI = confidence interval, CMMI = Center for Medicare & Medicaid Innovation, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, ICD = International Classification of Diseases, IRF = inpatient rehabilitation facility, MCC = major complications or comorbidities, MS-DRG = Medicare Severity Diagnosis Related Group, pp = percentage point, PY = performance year, SNF = skilled nursing facility.

[a] The number of episodes for these measures is lower because episodes were dropped to account for missing observations. These measures are based on 19,865 CJR baseline episodes, 19,903 CJR intervention episodes, 18,013 control group baseline episodes, and 16,740 control group intervention episodes.

[b] Other includes beneficiaries identified as "Asian," "American Indian/Alaska Native," or "Other."

Exhibit J-7: Change in patient complexity measures, fracture MS-DRG 469 patients, PY1-4

Patient characteristics		CJR		Control group		Net differences (pp)	Net differences % of baseline	p-value	90% CI
		Baseline Average (N=3,750)	*Intervention Average (N=3,859)*	*Baseline Average (N=3,935)*	*Intervention Average (N=3,593)*				
Age	20-64	3.2%	2.8%	3.5%	3.0%	0.1	3.6%	p=0.83	[-0.8 to 1.0]
	65-79	25.6%	29.5%	28.3%	32.1%	0.1	0.5%	p=0.93	[-2.3 to 2.6]
	80+	71.1%	67.6%	68.2%	64.9%	-0.2	-0.3%	p=0.87	[-2.7 to 2.2]
Sex	Female	67.9%	66.4%	66.5%	67.0%	-2.0	-2.9%	p=0.24	[-4.8 to 0.8]
Race/ethnicity	White[a]	87.6%	86.9%	90.5%	91.9%	-2.1	-2.4%	p<0.05	[-3.6 to -0.7]
	Black or African American[a]	4.0%	5.0%	5.2%	4.9%	1.3	32.8%	p<0.10	[0.1 to 2.5]
	Hispanic[a]	5.4%	5.0%	2.5%	2.0%	0.1	1.6%	p=0.87	[-0.8 to 1.0]
	Other[ab]	3.0%	3.1%	1.8%	1.1%	0.7	24.9%	p<0.05	[0.2 to 1.3]
Medicaid	Eligible for Medicaid	24.0%	23.2%	19.2%	16.1%	2.3	9.6%	p=0.14	[-0.3 to 4.9]
Disability	Disability, no ESRD	11.8%	12.9%	12.6%	12.7%	0.9	7.9%	p=0.43	[-1.0 to 2.9]
Health Status	HCC score	3.0	3.1	2.9	3.1%	-0.0	-0.0%	p=0.64	[-0.1 to 0.1]
	Obesity	5.0%	10.0%	7.2%	11.1%	1.0	20.6%	p=0.36	[-0.8 to 2.9]
	Diabetes	33.5%	30.9%	30.0%	29.0%	-1.5	-4.6%	p=0.39	[-4.5 to 1.4]
	Hypertension	78.8%	79.8%	79.2%	78.7%	1.5	2.0%	p=0.19	[-0.4 to 3.5]
	Dementia	34.9%	35.3%	34.4%	33.0%	1.8	5.3%	p=0.21	[-0.6 to 4.3]
	CHF	38.6%	36.8%	37.0%	35.4%	-0.1	-0.4%	p=0.92	[-2.3 to 2.0]
Utilization in the six months prior to the anchor hospitalization	ACH stay	30.2%	28.9%	28.9%	29.0%	-1.5	-4.8%	p=0.27	[-3.6 to 0.7]
	IRF stay	3.5%	3.2%	3.1%	4.2%	-1.4	-38.6%	p<0.05	[-2.3 to -0.4]
	SNF stay	15.9%	16.1%	15.4%	14.5%	1.1	6.9%	p=0.29	[-0.6 to 2.8]
	HH use	32.1%	29.7%	31.4%	31.2%	-2.3	-7.1%	p=0.24	[-5.4 to 0.9]
	Any prior care	56.5%	55.7%	56.1%	57.3%	-2.1	-3.7%	p=0.20	[-4.8 to 0.6]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of calculating the DiD of the unadjusted baseline and intervention averages for the CJR and control groups (net differences). Estimates that are significant at the 1%, 5%, or 10% significance levels are indicated by red, orange, or yellow shaded cells, respectively.

The MS-DRG 469 is assigned at the anchor hospitalization discharge for major joint replacement or reattachment of lower extremity *with* MCC, while MS-DRG 470 is *without* MCC.

Fracture is defined based on ICD codes for hip fracture provided by the CMMI on the CJR model website: https://innovation.cms.gov/Files/worksheets/cjr-icd10hipfracturecodes.xlsx.

ACH = acute care hospital, CHF = congestive heart failure, CI = confidence interval, CMMI = Center for Medicare & Medicaid Innovation, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, ICD = International Classification of Diseases, IRF = inpatient rehabilitation facility, MCC = major complications or comorbidities, MS-DRG = Medicare Severity Diagnosis Related Group, pp = percentage point, PY = performance year, SNF = skilled nursing facility.

[a] The number of episodes for these measures is lower because episodes were dropped to account for missing observations. These measures are based on 3,742 CJR baseline episodes, 3,854 CJR intervention episodes, 3,919 control group baseline episodes, and 3,586 control group intervention episodes.

[b] Other includes beneficiaries identified as "Asian," "American Indian/Alaska Native," or "Other.

Appendix K: Parallel Trends

Results of parallel trends tests

Exhibit K-1: Linear and joint tests of parallel trends for payment, utilization, and quality metrics, mandatory, opt-in and non-opt-in CJR hospitals, inpatient LEJR episodes, baseline

Domain	Measure	Mandatory		Opt-In		Non-Opt-In	
		Joint test	Linear test	Joint test	Linear test	Joint test	Linear test
Payments	Episode payments	p=0.64	p=0.88	p=0.45	p=0.70	p<0.05	p=0.12
	IRF payments	p=0.59	p=0.72	p=0.86	p=0.75	p=0.15	p=0.23
	SNF payments	p=0.56	p=0.94	p=0.40	p=0.90	p=0.79	p=0.30
	HH payments	p<0.01	p<0.01	p=0.74	p=0.72	p=0.50	p=0.26
	Readmission payments	p=0.42	p=0.29	p<0.10	p=0.20	p<0.05	p<0.10
	Part B payments	p=0.22	p=0.12	p=0.30	p=0.56	p=0.85	p=0.98
	30-day PEP payments	p<0.01	p<0.05	p=0.96	p=0.63	p<0.10	p<0.10
	Anchor payments	p<0.05	p=0.24	p=0.56	p=0.76	p=0.52	p=0.83
Utilization	First PAC IRF	p=0.57	p=0.75	p=0.63	p=0.53	p=0.18	p=0.31
	First PAC SNF	p=0.34	p=0.45	p=0.54	p=0.33	p=0.87	p=0.44
	First PAC HH	p=0.11	p<0.10	p=0.78	p=0.99	p=0.32	p=0.49
	First PAC home without HH	p=0.58	p=0.60	p=0.36	p=0.67	p=0.28	p<0.10
	Any HH use	p<0.05	p<0.05	p=0.29	p=0.44	p=0.43	p=0.16
	IRF days	p=0.60	p=0.50	p=0.39	p=0.17	p=0.85	p=0.49
	SNF days	p=0.72	p=0.18	p=0.16	p=0.14	p=0.24	p=0.75
	HH visits	p=0.42	p=0.21	p=0.93	p=0.87	p=0.91	p=0.35
	HH PT/OT visits	p=0.31	p=0.35	p=0.19	p=0.21	p=0.47	p=0.36
	Outpatient PT/OT visits	p<0.10	p<0.05	p<0.05	p=0.12	p=0.26	p=0.42

Domain	Measure	Mandatory		Opt-In		Non-Opt-In	
		Joint test	Linear test	Joint test	Linear test	Joint test	Linear test
Quality	Unplanned readmission rate	p=0.32	p=0.82	p=0.27	p=0.93	p=0.18	p=0.24
	ED use	p=0.56	p=0.62	p=0.79	p=0.93	p=0.13	p<0.10
	Mortality rate	p=0.70	p=0.82	p=0.93	p=0.46	p=0.19	p<0.10
	Complications[a]	p=0.92	p=0.72	p=0.11	p=0.86	p=0.98	p=0.85

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline).

***Notes*:** The p-values in this exhibit are the result of risk-adjusted regression models analyzing if the respective CJR and control groups followed parallel trends during the baseline period. For the joint test, we report the p-value of an F-test that tests if the differential between the CJR and control group are jointly equal across annual time periods. For the linear test, we report the p-value of a linear slope coefficient of the quarterly difference between the CJR and control group. We consider outcomes to *fail* parallel trends if we reject the null hypothesis of seemingly parallel trends for both tests at the 10% significance level. P-values of estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

ED = emergency department, HH = home health, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, OT = occupational therapy, PAC = post-acute care, PEP = post-episode period, PT = physical therapy, SNF = skilled nursing facility.

[a] The complications measure only applies to elective episodes.

Exhibit K-2: Linear and joint tests of parallel trends for activities of daily living metrics, mandatory, opt-in and non-opt-in CJR hospitals, inpatient LEJR episodes, baseline

First PAC setting	Measure	Mandatory		Opt-In		Non-Opt-In	
		Joint test	Linear test	Joint test	Linear test	Joint test	Linear test
IRF	Average change in mobility score	p=0.59	p=0.23	p=0.14	p=0.53	p=0.41	p=0.22
SNF	Improved transfer, locomotion on unit, and walking in corridor	p=0.20	p=0.19	p=0.60	p=0.48	p<0.05	p=0.41
	Improved toilet use	p=0.41	p=0.34	p=0.46	p=0.28	p=0.24	p=0.23
	Without self-reported pain[a]	p<0.10	p<0.05	p=0.72	p=0.57	p=0.18	p<0.10
HHA	Improved ambulation/locomotion	p=0.74	p=0.84	p=0.59	p=0.29	p=0.19	p=0.15
	Improved bed transferring	p=0.29	p=0.51	p=0.38	p=0.20	p=0.33	p=0.19
	Reduced pain	p=0.88	p=0.92	p=0.28	p=0.80	p<0.05	p<0.01

Source: CJR evaluation team analysis of Medicare claims and enrollment data, MDS data, OASIS data, and IRF-PAI data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline).

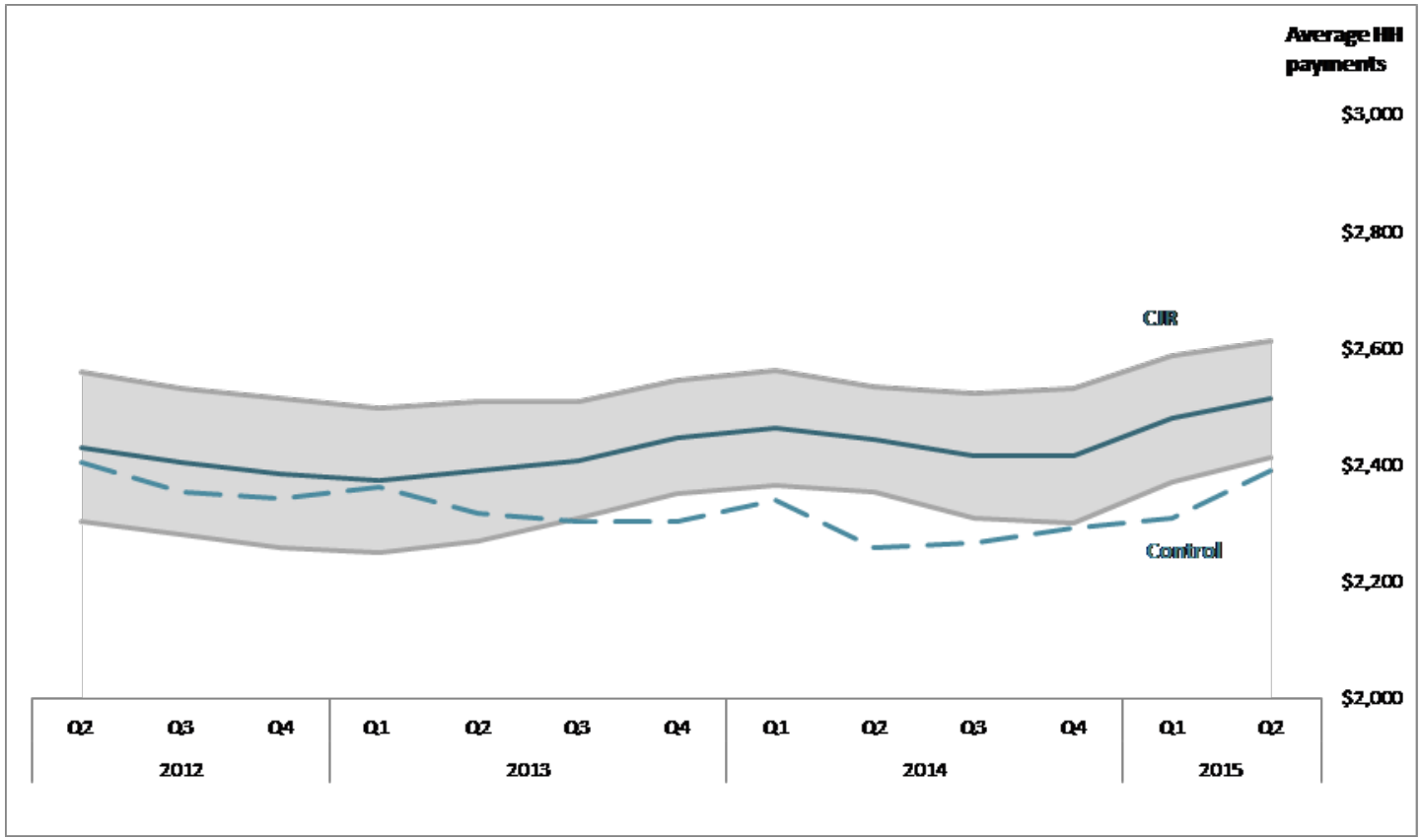
***Notes*:** The p-values in this exhibit are the result of risk-adjusted regression models analyzing if the respective CJR and control groups followed parallel trends during the baseline period. For the joint test, we report the p-value of an F-test that tests if the differential between the CJR and control group are jointly equal across annual time periods. For the linear test, we report the p-value of a linear slope coefficient of the quarterly difference between the CJR and control group. We consider outcomes to *fail* parallel trends if we reject the null hypothesis of seemingly parallel trends for both tests at the 10% significance level. P-values of estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

HHA = home health agency, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, OASIS = Outcome and Assessment Information Set, PAC = post-acute care, SNF = skilled nursing facility.

[a] The pain measure for those initially discharged to a SNF was not risk adjusted following the specifications of the MDS 3.0 Quality Measure for short-stay patients used in the CMS Nursing Home Five-Star Rating System.

Trends for outcomes that failed parallel trends

Exhibit K-3: Risk-adjusted baseline trends for HH payments, LEJR episodes at mandatory hospitals



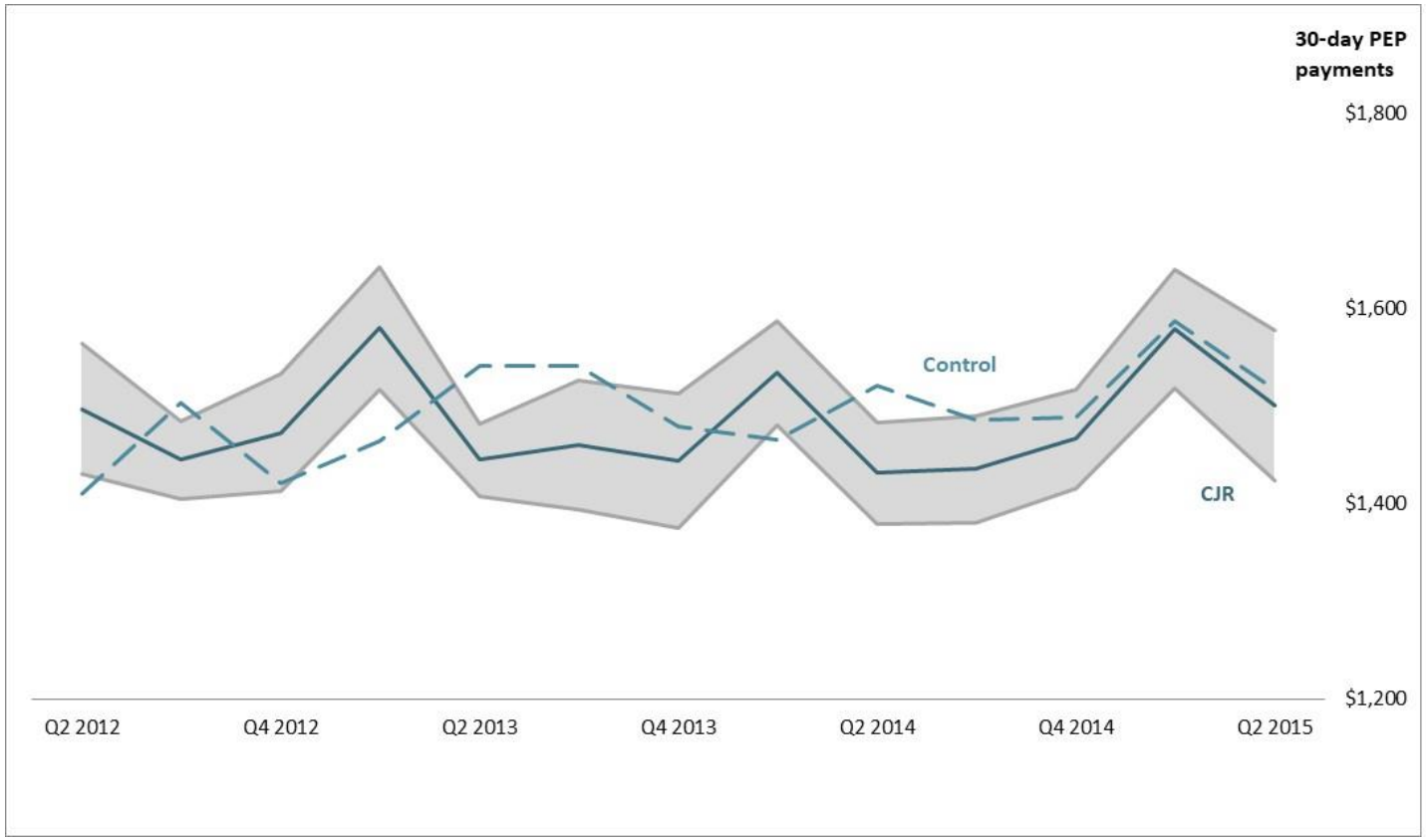
Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

HH = home health, LEJR = lower extremity joint replacement.

Exhibit K-4: Risk-adjusted baseline trends for 30-day post-episode payments, LEJR episodes at mandatory hospitals



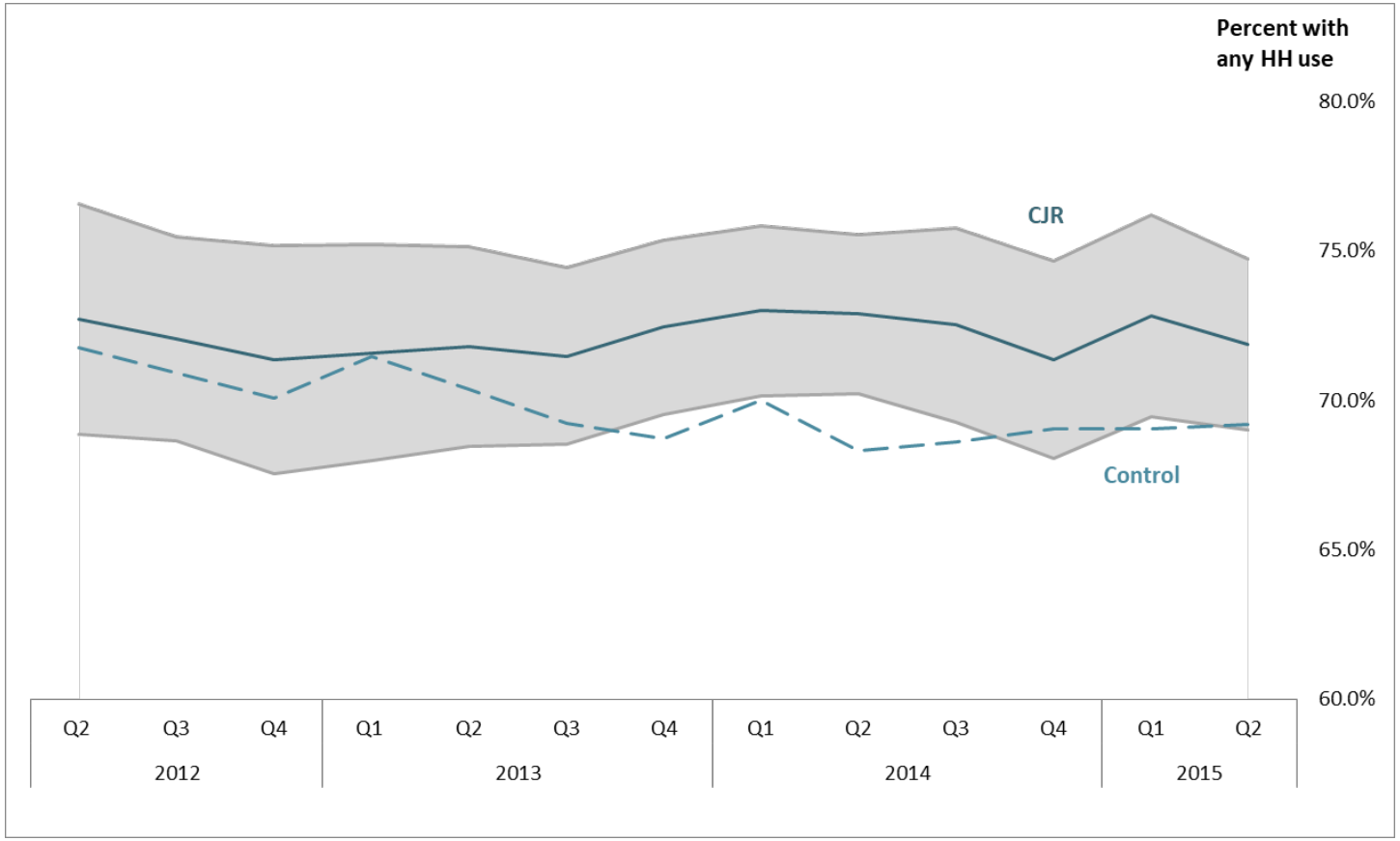
Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

LEJR = lower extremity joint replacement, PEP = post-episode period.

Exhibit K-5: Risk-adjusted baseline trends for any HH use, LEJR episodes at mandatory hospitals



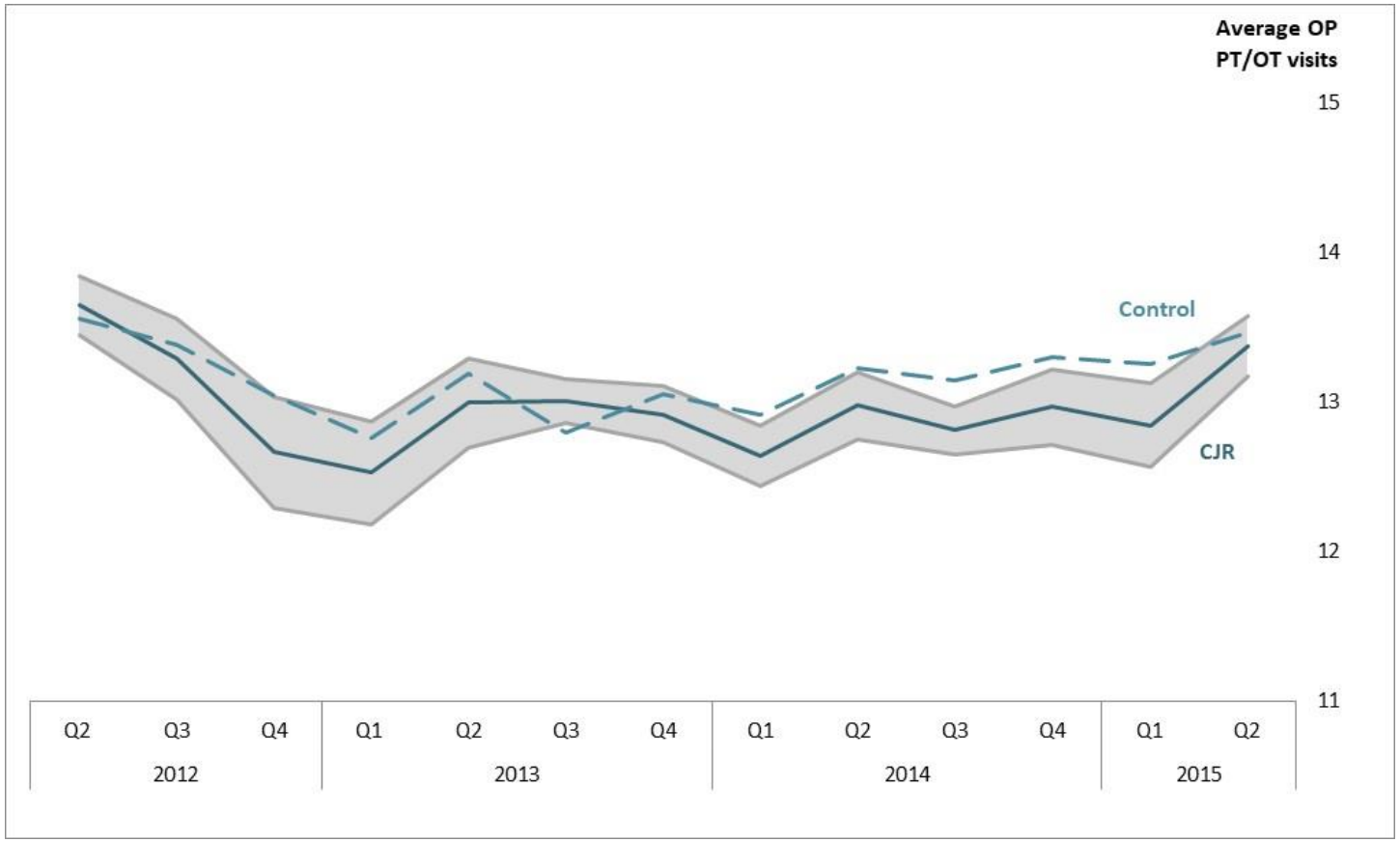
Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

HH = home health, LEJR = lower extremity joint replacement.

Exhibit K-6: Risk-adjusted baseline trends for number of outpatient PT/OT visits, among outpatient PT/OT users, LEJR episodes at mandatory hospitals



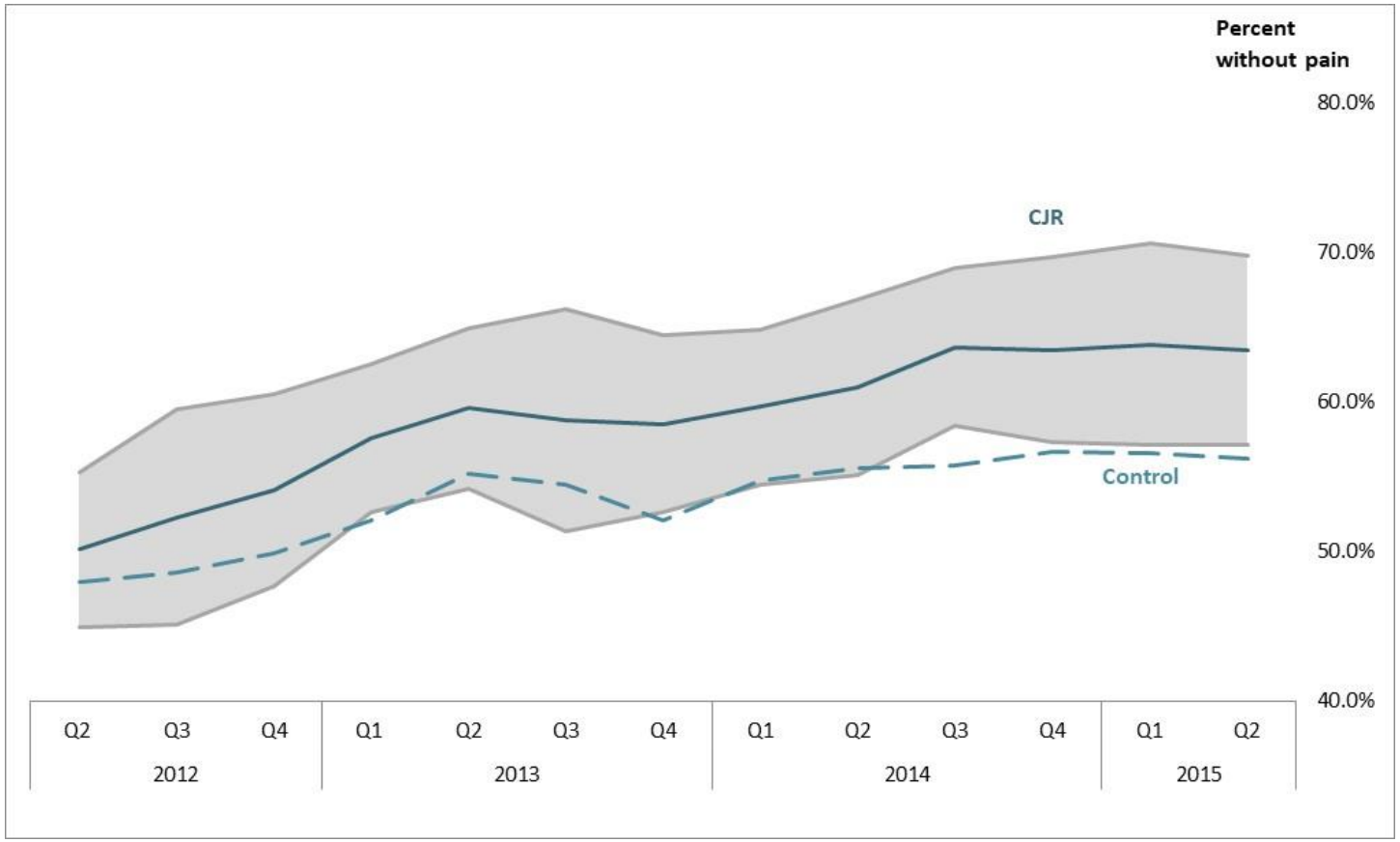
Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

LEJR = lower extremity joint replacement, OT = occupational therapy, PT = physical therapy.

Exhibit K-7: Unadjusted baseline trends for percent of patients without moderate to severe pain at admission, LEJR episodes at mandatory hospitals discharged to SNF



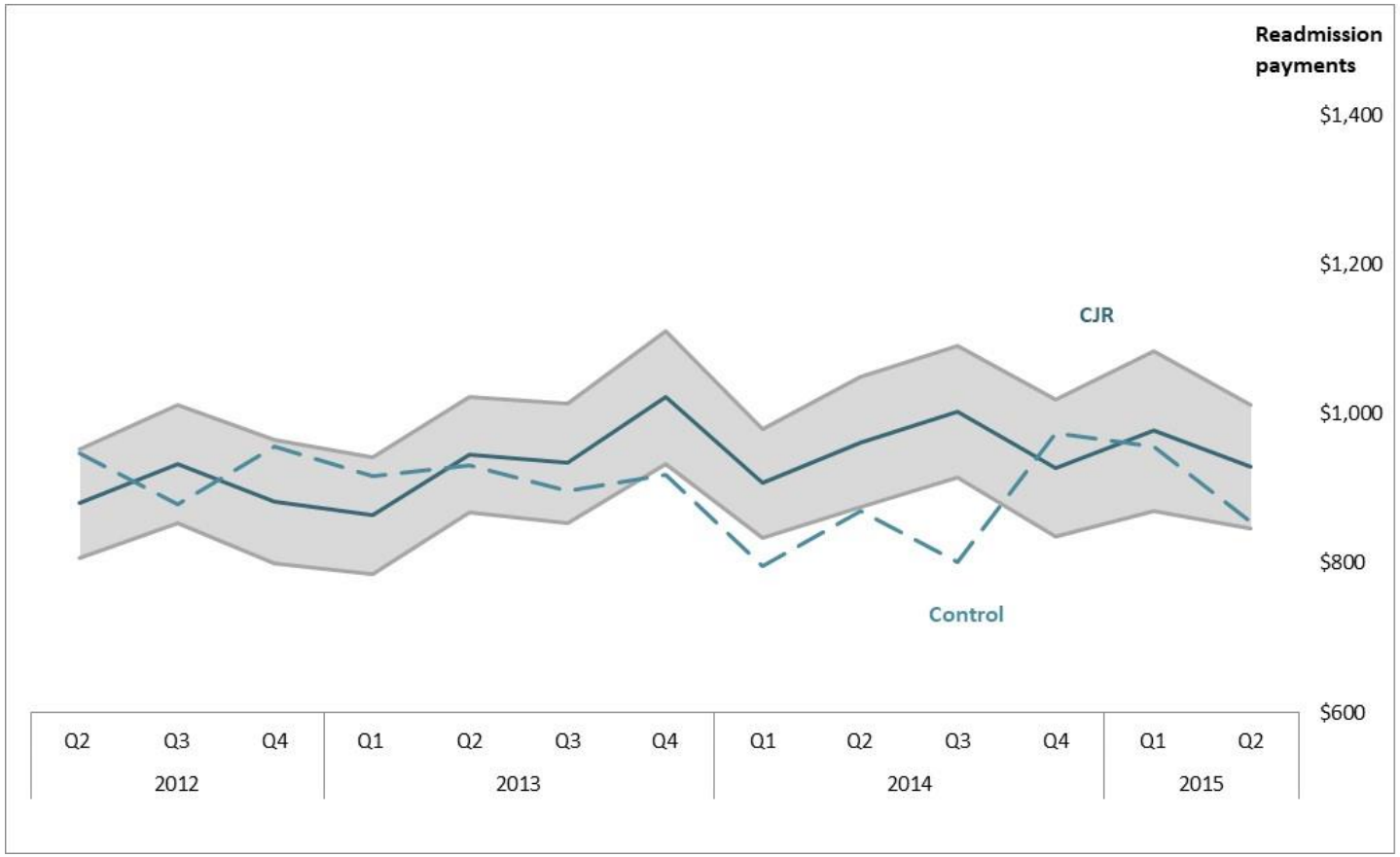
Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

LEJR = lower extremity joint replacement, SNF = skilled nursing facility.

Exhibit K-8: Risk-adjusted baseline trends for readmission payments, LEJR episodes at non-opt-in hospitals



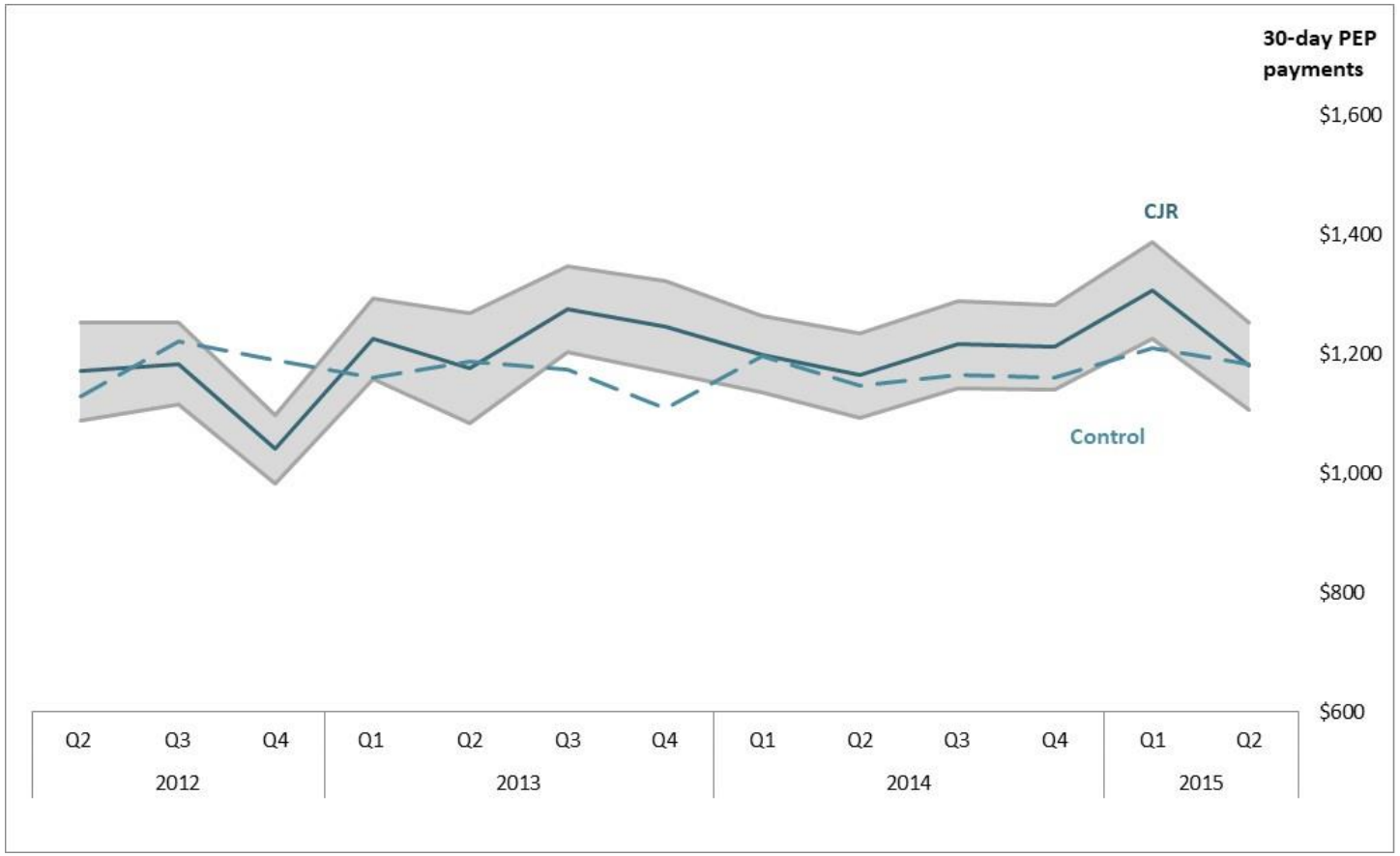
Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

LEJR = lower extremity joint replacement.

Exhibit K-9: Risk-adjusted baseline trends for 30-day post-episode payments, LEJR episodes at non-opt-in hospitals



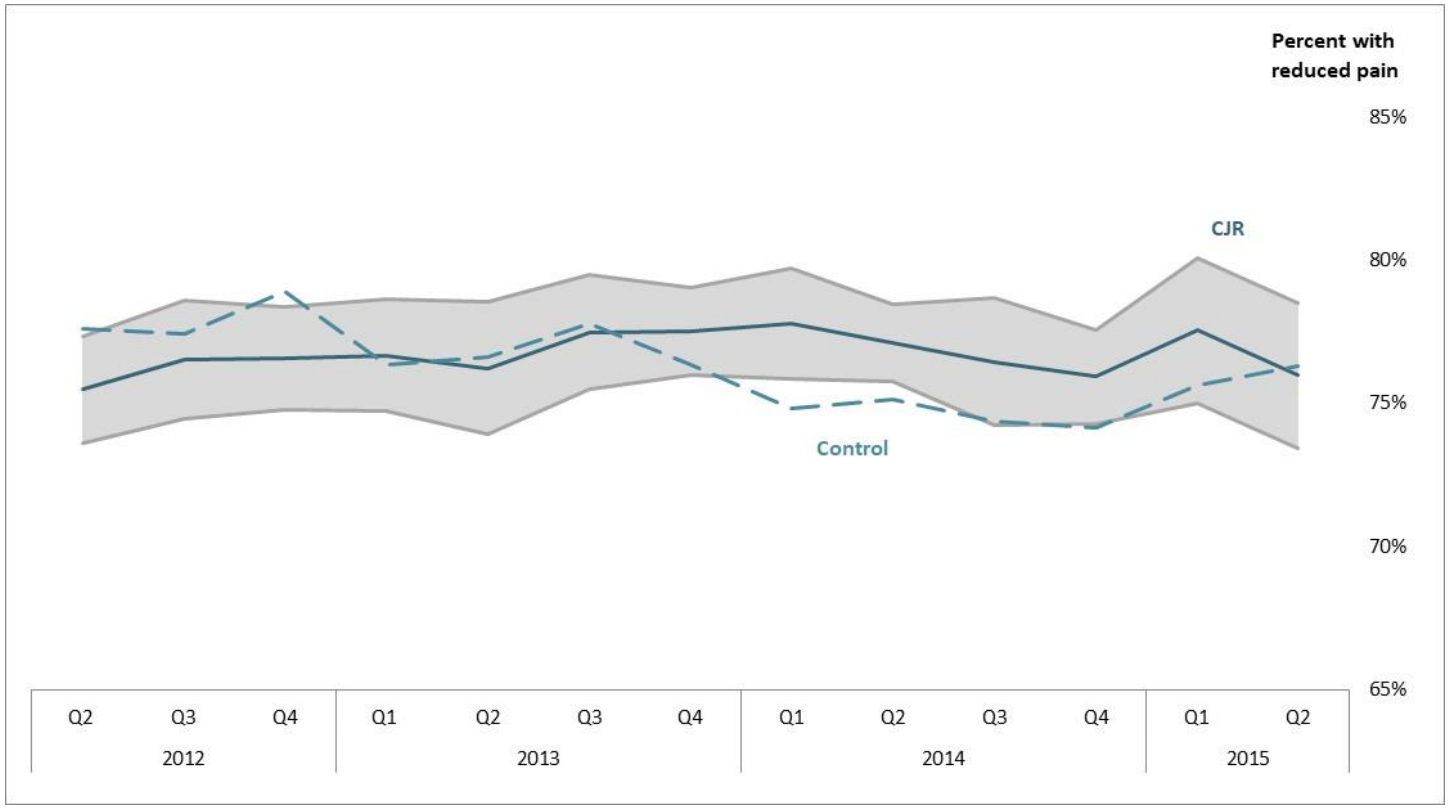
Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

LEJR = lower extremity joint replacement, PEP = post-episode period.

Exhibit K-10: Risk-adjusted baseline trends for percent of patients with reduced pain, LEJR episodes at non-opt-in hospitals discharged to HHA



Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated on or after January 2012 that ended by December 2019.

Notes: Baseline trends were estimated from a risk-adjusted trend model using the baseline, interim, and intervention periods.

The gray shading represents the 90% confidence intervals for the CJR estimate.

HHA = home health agency, LEJR = lower extremity joint replacement.

Appendix L: Factors Associated with Receiving Reconciliation Payments

Exhibit L-1: Factors related to hospital level of average reconciliation payment per episode, mandatory hospitals (n=1,128), PY1-4

Domain	Reference category	Measure	Difference in average reconciliation payment per episode from reference group	90% Confidence Interval	p-value
Intercept	NA	Intercept	-$2,904	[-$3,303 to -$2,505]	p<0.01
Performance year	PY1	PY2	$425	[$323 to $527]	p<0.01
		PY3	-$215	[-$389 to -$42]	p<0.05
		PY4	-$349	[-$584 to -$114]	p<0.05
Hospital historical average payments in relation to the PY target price	Above the target price	Started the PY with hospital historical average payments below the PY target price	$525	[$307 to $742]	p<0.01
PY performance quality category	Below acceptable	Acceptable	$629	[$360 to $898]	p<0.01
		Good	$1,001	[$801 to $1,201]	p<0.01
		Excellent	$1,198	[$931 to $1,465]	p<0.01
PY average quarterly volume	<15 episodes	15-49 episodes	$560	[$346 to $774]	p<0.01
		50 or more episodes	$852	[$569 to $1,135]	p<0.01
Ownership	For profit	Not for profit	$563	[$320 to $807]	p<0.01
		Government	$356	[-$28 to $740]	p=0.13
Census region	West	Northeast	$459	[$97 to $822]	p<0.05
		South	$381	[$57 to $706]	p<0.10
		Midwest	$222	[-$204 to $649]	p=0.39
Hospital characteristics	Above median	Below median hospital DSH patient percentage	$276	[$60 to $491]	p<0.05
	Above median	Below median bed count	$265	[$59 to $471]	p<0.05
	No affiliation	Any affiliation with a medical school	$253	[$42 to $464]	p<0.05
	Never participated	Ever participated in BPCI LEJR	-$39	[-$339 to $260]	p=0.83

Domain	Reference category	Measure	Difference in average reconciliation payment per episode from reference group	90% Confidence Interval	p-value
Patient characteristics	Above median	Below median average HCC score for PY episodes	$526	[$305 to $747]	p<0.01
	Above median	Below median percent of PY episodes age 80 years or older	$200	[$24 to $377]	p<0.10
	Above median	Below median percent of PY episodes dual eligible	$320	[$105 to $535]	p<0.05
	Above median	Below median percent of PY episodes with prior institutional stay	$240	[$55 to $425]	p<0.05
	Above median	Below median percent of PY episodes non-Hispanic Black or African-American	$125	[-$48 to $297]	p=0.23
	Above median	Below median percent of PY episodes with disability, no ESRD	-$124	[-$349 to $101]	p=0.36
	Below median	Above median percent of PY episodes MS-DRG 470 elective	-$174	[-$394 to $46]	p=0.19
	Above median	Below median percent of PY episodes that are female	$54	[-$91 to $200]	p=0.54

Source: CJR evaluation team analysis of December 2016 POS, FY 2016 CMS Annual IPPS, BPCI Salesforce participation list, CMS payment contractor CJR NPRA, quality performance, Medicare claims and enrollment, and target price data for mandatory CJR participant hospitals in PY1 (episodes initiated during or after April 2016 that ended by December 2016), PY2 (episodes ending in 2017), PY3 (episodes ending in 2018), and PY4 (episodes ending in 2019).

***Notes*:** PY1, PY2, and PY3 NPRA data are final, while the PY4 NPRA data are preliminary and will be finalized spring 2021.

Multivariate generalized linear regression model, which accounts for multiple observations (PY) per hospital and clustering of hospitals at the Metropolitan Statistical Area (MSA) level, was used to identify factors related to average reconciliation payment per episode that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

We restricted the sample to hospitals with 20 or more episodes in the performance year to improve the stability of results. Mandatory hospitals with positive amounts per episode in a performance year earned reconciliation payments under the CJR model. Hospitals with no or negative amounts per episode included hospitals with episode payments above their quality-adjusted target price and hospitals with episode payments below their quality-adjusted target price but with quality composite scores "below acceptable quality" making them ineligible for reconciliation payments. We calculated the potential repayment amount for PY1 because hospitals were not required to make a repayment in the first year of the CJR model. Stop gain and loss limits are applied to the overall reconciliation amount per episode.

Median values for categorizing variables included: bed count, 258; DSH patient percentage, 24.2%; dual eligible, 11.2%; average HCC score, 1.60; MS-DRG 470 elective, 81.4%; 80 years or older, 26.9%; prior institutional stay, 4.9%; disability, 15.1%; non-Hispanic Black or African-American, 3.9%; and female, 65.6%.

BPCI = Bundled Payments for Care Improvement, DSH = disproportionate share hospital, ESRD = end-stage renal disease, FY = fiscal year, HCC = hierarchical condition category, IPPS = Inpatient Prospective Payment System, LEJR = lower extremity joint replacement, MS-DRG = Medicare Severity-Diagnosis Related Group, NA = not applicable, NPRA = net payment reconciliation amount, POS = provider of services, PY = performance year.

Exhibit L-2: Factors related to hospital level of average reconciliation payment per episode, opt-in hospitals (n=278), PY1-4

Domain	Reference category	Measure	Difference in average reconciliation payment per episode from reference group	90% Confidence Interval	p-value
Intercept	NA	Intercept	-$1,902	[-$2,726 to -$1,078]	p<0.01
Performance year	PY1	PY2	$306	[$111 to $502]	p<0.01
		PY3	$522	[$272 to $773]	p<0.01
		PY4	$1,196	[$885 to $1,507]	p<0.01
Hospital historical average payments in relation to the PY target price	Above the target price	Started the PY with hospital historical average payments below the PY target price	$716	[$382 to $1,049]	p<0.01
PY performance quality category	Below acceptable	Acceptable	$39	[-$608 to $686]	p=0.92
		Good	$1,209	[$546 to $1,873]	p<0.01
		Excellent	$1,313	[$670 to $1,956]	p<0.01
PY average quarterly volume	<15 episodes	15-49 episodes	-$277	[-$870 to $315]	p=0.44
		50 or more episodes	-$54	[-$658 to $550]	p=0.88
Ownership	For profit	Not for profit	$503	[$110 to $896]	p<0.05
		Government	$678	[$158 to $1,197]	p<0.05
Census region	West	Northeast	$98	[-$567 to $762]	p=0.81
		South	$495	[-$31 to $1,022]	p=0.12
		Midwest	$220	[-$101 to $540]	p=0.26
Hospital characteristics	Above median	Below median hospital DSH patient percentage	-$68	[-$428 to $292]	p=0.76
	Above median	Below median bed count	$340	[-$58 to $738]	p=0.16
	No affiliation	Any affiliation with a medical school	$324	[-$16 to $663]	p=0.12
	Never participated	Ever participated in BPCI LEJR	$1,225	[$507 to $1,942]	p<0.01

Domain	Reference category	Measure	Difference in average reconciliation payment per episode from reference group	90% Confidence Interval	p-value
Patient characteristics	Above median	Below median average HCC score for PY episodes	$32	[-$432 to $496]	p=0.91
	Above median	Below median percent of PY episodes age 80 years or older	$182	[-$86 to $451]	p=0.26
	Above median	Below median percent of PY episodes dual eligible	$53	[-$207 to $312]	p=0.74
	Above median	Below median percent of PY episodes with prior institutional stay	$91	[-$122 to $305]	p=0.48
	Above median	Below median percent of PY episodes non-Hispanic Black or African-American	-$41	[-$335 to $253]	p=0.82
	Above median	Below median percent of PY episodes with disability, no ESRD	$156	[-$58 to $369]	p=0.23
	Below median	Above median percent of PY episodes MS-DRG 470 elective	-$268	[-$557 to $20]	p=0.13
	Above median	Below median percent of PY episodes that are female	$159	[-$58 to $377]	p=0.23

Source: CJR evaluation team analysis of December 2016 POS, FY 2016 CMS Annual IPPS, BPCI Salesforce participation list, CMS payment contractor CJR NPRA, quality performance, Medicare claims and enrollment, and target price data for opt-in CJR participant hospitals in PY1 (episodes initiated during or after April 2016 that ended by December 2016), PY2 (episodes ending in 2017), PY3 (episodes ending in 2018), and PY4 (episodes ending in 2019).

***Notes*:** PY1, PY2, and PY3 NPRA data are final, while the PY4 NPRA data are preliminary and will be finalized spring 2021.

Multivariate generalized linear regression model, which accounts for multiple observations (PY) per hospital and clustering of hospitals at the Metropolitan Statistical Area (MSA) level, was used to identify factors related to average reconciliation payment per episode that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

We restricted the sample to hospitals with 20 or more episodes in the performance year to improve the stability of results. Voluntary opt-in hospitals with positive amounts per episode in a performance year earned reconciliation payments under the CJR model. Hospitals with no or negative amounts per episode included hospitals with episode payments above their quality-adjusted target price and hospitals with episode payments below their quality-adjusted target price but with quality composite scores "below acceptable quality" making them ineligible for reconciliation payments. We calculated the potential repayment amount for PY1 because hospitals were not required to make a repayment in the first year of the CJR model. Stop gain and loss limits are applied to the overall reconciliation amount per episode.

Median values for categorizing variables included: bed count, 188; DSH patient percentage, 21.8%; dual eligible, 9.0%; average HCC score, 1.31; MS-DRG 470 elective, 89.0%; 80 years or older, 21.9%; prior institutional stay, 3.1%; disability, 13.9%; non-Hispanic Black or African-American, 2.2%; and female, 64.3%.

BPCI = Bundled Payments for Care Improvement, DSH = disproportionate share hospital, ESRD = end-stage renal disease, FY = fiscal year, HCC = hierarchical condition category, IPPS = Inpatient Prospective Payment System, LEJR = lower extremity joint replacement, MS-DRG = Medicare Severity-Diagnosis Related Group, NA = not applicable, NPRA = net payment reconciliation amount, POS = provider of services, PY = performance year.

Appendix M: Orthopedic Surgeon Survey Questions

Section 1. Before Surgery

*We are interested in changes in the way you assess patients as candidates for hip or knee replacement surgery. Please think **ONLY** about your Medicare fee-for-service patients when answering these questions. If you work at multiple hospitals, please think about the hospital where you work most.*

The CJR model holds participant hospitals financially accountable for the quality and cost of a CJR episode of care. Therefore, we would like to know about whether hospitals provide guidelines or directives about **modifiable health risk factors** that they want you to consider when determining whether to perform hip or knee replacement surgery.

1. Do hospitals provide guidelines or directives about **uncontrolled diabetes** that they want you to consider when determining whether to perform hip or knee replacement surgery?

   - ○ Yes
   - ○ No → **If No, go to #2**

1a. Have these guidelines or directives from hospitals changed in the past three years?

   - ○ Yes
   - ○ No

1b. How important are guidelines or directives from hospitals about uncontrolled diabetes in your decision making process when considering whether to perform hip or knee replacement surgery?

   - ○ Very important
   - ○ Somewhat important
   - ○ Not important

2. Do hospitals provide guidelines or directives about **obesity** that they want you to consider when determining whether to perform hip or knee replacement surgery?

   - ○ Yes
   - ○ No → **If No, go to #3**

2a. Have these guidelines or directives from hospitals changed in the past three years?

   - ○ Yes
   - ○ No

2b. How important are guidelines or directives from hospitals about obesity in your decision making process when considering whether to perform hip or knee replacement surgery?

- ○ Very important
- ○ Somewhat important
- ○ Not important

3. Do hospitals provide guidelines or directives about **patient smoking** that they want you to consider when determining whether to perform hip or knee replacement surgery?

- ○ Yes
- ○ No → **If No, go to #4**

3a. Have these guidelines or directives from hospitals changed in the past three years?

- ○ Yes
- ○ No

3b. How important are guidelines or directives from hospitals about patient smoking in your decision making process when considering whether to perform hip or knee replacement surgery?

- ○ Very important
- ○ Somewhat important
- ○ Not important

4. Do hospitals provide guidelines or directives about **depression or anxiety** that they want you to consider when determining whether to perform hip or knee replacement surgery?

- ○ Yes
- ○ No → **If No, go to #5**

4a. Have these guidelines or directives from hospitals changed in the past three years?

- ○ Yes
- ○ No

4b. How important are guidelines or directives from hospitals about depression or anxiety in your decision making process when considering whether to perform hip or knee replacement surgery?

- ○ Very important
- ○ Somewhat important
- ○ Not important

5. Is there another modifiable health risk factor, **other than** diabetes, obesity, smoking, and depression or anxiety, that hospitals want you to consider when determining whether to perform hip or knee replacement surgery? *Please specify the next most important factor **other than** diabetes, obesity, smoking and depression or anxiety.*

   - ○ Yes, please specify: _____
   - ○ No → **If No, go to #6**

5a. Have these guidelines or directives from hospitals changed in the past three years?

   - ○ Yes
   - ○ No

5b. How important are guidelines or directives from hospitals about this other modifiable health risk factor in your decision making process when considering whether to perform hip or knee replacement surgery?

   - ○ Very important
   - ○ Somewhat important
   - ○ Not important

6. When you determine that a patient has a modifiable health risk factor that prevents them from being a good candidate for surgery, what do you do? *Please mark all that apply*.

   - ○ Postpone surgery
   - ○ Refer the patient to a primary care provider to address the risk factor
   - ○ Refer the patient to a specialist to address the risk factor
   - ○ Refer the patient to a surgeon with expertise in treating patients with that type of health risk factor
   - ○ Plan to discharge the patient to institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities)
   - ○ Give the patient instructions about how to address the risk factor
   - ○ Other, specify: _____

Next, we would like to know about whether hospitals provide guidelines or directives about **environmental risk factors** that they want you to consider when determining whether to perform hip or knee replacement surgery.

7. Do hospitals provide guidelines or directives about **lack of caregiver support**, that they want you to consider when determining whether to perform hip or knee replacement surgery?

   - ○ Yes
   - ○ No → **If No, go to #8**

7a. Have these guidelines or directives from hospitals changed in the past three years?

- ○ Yes
- ○ No

7b. How important are guidelines or directives from hospitals about lack of caregiver support in your decision making process when considering whether to perform hip or knee replacement surgery?

- ○ Very important
- ○ Somewhat important
- ○ Not important

8. Do hospitals provide guidelines or directives about **lack of transportation,** that they want you to consider when determining whether to perform hip or knee replacement surgery?

- ○ Yes
- ○ No → **If No, go to #9**

8a. Have these guidelines or directives changed from hospitals in the past three years?

- ○ Yes
- ○ No

8b. How important are guidelines or directives from hospitals about lack of transportation in your decision making process when considering whether to perform hip or knee replacement surgery?

- ○ Very important
- ○ Somewhat important
- ○ Not important

9. Do hospitals provide guidelines or directives about the **safety of the home environment** to consider when determining whether to perform hip or knee replacement surgery?

- ○ Yes
- ○ No → **If No, go to #10**

9a. Have these guidelines or directives from hospitals changed in the past three years?

- ○ Yes
- ○ No

9b. How important are guidelines or directives from hospitals about the safety of the home environment in your decision making process when considering whether to perform hip or knee replacement surgery?

- ○ Very important
- ○ Somewhat important
- ○ Not important

10. Is there an environmental risk factor, **other than** lack of caregiver support, lack of transportation, or safety of the home environment, that hospitals want you to consider when determining whether to perform hip or knee replacement surgery? (Please specify the next most important factor other than lack of caregiver support, lack of transportation, or safety of the home environment)

    - ○ Yes, please specify: _____
    - ○ No → **If No, go to #11**

10a. Have these guidelines or directives from hospitals changed in the past three years?

    - ○ Yes
    - ○ No

10b. How important are guidelines or directives from hospitals about this other environmental risk factor in your decision making when considering whether to perform hip or knee replacement surgery?

    - ○ Very important
    - ○ Somewhat important
    - ○ Not important

11. When you determine that a patient has an environmental risk factor that prevents them from being a good candidate for surgery, what do you do? *Please mark all that apply*.

    - ○ Postpone surgery
    - ○ Refer the patient elsewhere to address the risk factor
    - ○ Plan to discharge the patient to institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities)
    - ○ Give the patient instructions about how to address the risk factor
    - ○ Other, specify: _____

Section 2. In-hospital care

*We are interested in any changes you've made in the past three years related to inpatient care for hip or knee replacement surgery patients, and how much influence hospitals had on these decisions. Please think **ONLY** about your Medicare fee-for-service patients when answering these questions.*

12. For those care processes that have changed in the past three years, please indicate the degree to which hospital guidelines or directives influenced any changes you made. If you made no changes, please indicate that instead. *Please select only one option in each row*.

Care processes	No change in process	No hospital influence on change	Hospital somewhat influenced change	Hospital greatly influenced change
Using only implants/prostheses approved by the hospital	O	O	O	O
Simplifying wound dressings	O	O	O	O
Changing pain management approaches	O	O	O	O
Earlier hospital discharge or decreasing length of hospital stay	O	O	O	O
Change to anesthesia protocol	O	O	O	O
Other: _____	O	O	O	O
Other: _____	O	O	O	O

Section 3. Post-surgical care

*We would like to understand the factors that influence your decisions regarding post-acute care for patients who have had hip or knee replacement surgery – especially the decision about institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities) versus discharge to home. Please think **ONLY** about your Medicare fee-for-service patients when answering these questions.*

13. How important do you consider the following factors when recommending institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities) versus discharge to home for patients following hip or knee replacement surgery? Please indicate how important each factor is below.

Factor	Very important	Somewhat important	Not important
Patient ambulation after surgery	O	O	O
Patient ability to transfer after surgery	O	O	O
Safety of the home environment	O	O	O
Whether patient lives alone	O	O	O
Caregiver support	O	O	O
Availability of transportation	O	O	O
Amount of pain management required	O	O	O
Comorbidities	O	O	O
Elective versus fracture	O	O	O
Hospital care coordinator recommendation	O	O	O
Other_____	O	O	O
Other_____	O	O	O

14. When recommending institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities) versus discharge to home for patients following hip or knee replacement surgery, do you consider different factors today than you did three years ago?

    - O Yes, please explain: _____
    - O No

15. Have hospital guidelines or directives about discharge to institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities) versus discharge to home changed in the past three years?

    - ○ Yes
    - ○ No

16. How much do hospital guidelines or directives influence your decisions about discharge to institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities)?

    - ○ Hospitals **greatly** influence my decisions about discharge to institutional post-acute care
    - ○ Hospitals **somewhat** influence my decisions about discharge to institutional post-acute care
    - ○ Hospitals **do not** influence my decisions about discharge to institutional post-acute care

17. In the past three years, has the proportion of hip and knee replacement surgery patients that you recommend for discharge to institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities) increased or decreased?

    - ○ I recommend **more** patients for institutional post-acute care than I did in 2016
    - ○ I recommend **fewer** patients for institutional post-acute care than I did in 2016
    - ○ No change in recommendations for institutional post-acute care

*We are interested in patients' status at their first post-operative appointment following **knee** replacement surgery. Please think **ONLY** about your Medicare fee-for-service patients when answering these questions.*

18. Do you perform knee replacement surgeries?

    - ○ Yes
    - ○ No➔ **If No, go to the Longer Term Outcomes section (starting just before question #22)**

19. In the past three years, have you seen a change in range of motion for the typical **knee** replacement surgery patient at their first post-operative appointment?

    - ○ Range of motion is typically **better** at the first visit than in 2016
    - ○ Range of motion is typically **worse** at the first visit than in 2016
    - ○ **No change** in range of motion at the first visit compared to 2016

20. In the past three years, has there been a change in the amount of pain described by the typical **knee** replacement surgery patient at their first post-operative appointment?

    - ○ Patients tend to report **less** pain at the first visit than in 2016
    - ○ Patients tend to report **more** pain at the first visit than in 2016
    - ○ **No change** in patient-reported pain at the first visit compared to 2016

21. In the past three years, have you seen a change in the rate of wound infections among typical **knee** replacement surgery patients at their first post-operative appointment?

    - ○ I see **fewer** wound infections at the first visit than I did in 2016
    - ○ I see **more** wound infections at the first visit than I did in 2016
    - ○ **No change** in wound infections at the first visit compared to 2016

Section 4. Longer Term Outcomes

*We are interested in whether longer term patient outcomes have changed in the past three years. This section asks separate questions about knee replacement, elective hip replacement, and hip fracture patients. Please think **ONLY** about your Medicare fee-for-service patients when answering these questions.*

22. In the past three years, have you seen a change in the number of **knee** replacement patients returning to you several months after surgery complaining of poor mobility or pain in their operative joint?

    - ○ **Fewer** patients have complaints than in 2016
    - ○ **More** patients have complaints than in 2016
    - ○ **No change** in patients with complaints than in 2016
    - ○ I do not perform knee replacement surgery

23. In the past three years, have you seen a change in the number of **elective hip** replacement patients returning to you several months after surgery complaining of poor mobility or pain in their operative joint?

    - ○ **Fewer** patients have complaints than in 2016
    - ○ **More** patients have complaints than in 2016
    - ○ **No change** in patients with complaints than in 2016
    - ○ I do not perform elective hip replacement surgery

24. In the past three years, have you seen a change in the number of patients who had hip replacement surgery as a result of a **fracture**, returning to you after several months with complaints of poor mobility or pain in their operative joint?

    - ○ **Fewer** patients have complaints than in 2016
    - ○ **More** patients have complaints than in 2016
    - ○ **No change** in patients with complaints compared to 2016
    - ○ I do not perform hip replacement surgery on hip fracture patients

Section 5. Hospital Performance Monitoring and Gainsharing

*We are interested in the information hospitals share with you about performance on quality and cost measures related to patients who had hip or knee replacement surgery, and if this has information sharing has changed in the past three years. Please think **ONLY** about your Medicare fee-for-service patients when answering these questions.*

25. Do hospitals share performance metrics with you – in scorecards, dashboards, or other reports – about outcomes for your patients who had hip or knee replacement surgery?

    - ○ Yes
    - ○ No → **If No, go to #29**

26. How frequently do hospitals share performance metrics with you about outcomes for your patients who had hip or knee replacement surgery? *Your best estimate is fine.*

    - ○ At least monthly
    - ○ Quarterly
    - ○ Twice a year
    - ○ Once a year

27. Which of the following performance metrics related to **your own** hip and knee replacement patients do hospitals share with you? *Please mark all that apply*.

    - ○ Surveys about your patients' satisfaction/experience with care
    - ○ Other patient-reported outcomes
    - ○ Post-discharge emergency department visits by your patients
    - ○ Readmissions for your patients
    - ○ Your patients' use of institutional post-acute care (skilled nursing facilities or inpatient rehabilitation facilities)
    - ○ Your patients' 90-day total episode cost of care
    - ○ Other (please specify): _____

28. To what degree does the information provided in these performance metrics lead you to modify your care practices for your hip and knee replacement patients?

    - ○ Performance metrics **greatly** influence my decisions to modify care
    - ○ Performance metrics **somewhat** influence my decisions to modify care
    - ○ Performance metrics **do not** influence my decisions to modify care

We are also interested in gainsharing arrangements related to your Medicare fee-for-service patients who have hip or knee replacement surgery.

29. Under the CJR model, CMS allows hospitals to share financial gains with surgeons. Do you have a **financial gainsharing arrangement** with any hospitals where you perform hip or knee replacement surgery for Medicare fee-for-service patients?

    - ○ Yes, I have a financial gainsharing arrangement with one or more hospitals
    - ○ No, but I am currently working on implementing a financial gainsharing arrangement with one or more hospitals
    - ○ No, but I would like to have a financial gainsharing arrangement with one or more hospitals
    - ○ No, and I am not interested in having a financial gainsharing arrangement with any hospitals
    - ○ Don't know

Section 6. Outpatient Knee Replacement Surgery

In 2018, Medicare removed knee replacement surgery from the inpatient-only list and began coverage for knee replacement surgery performed on an outpatient basis. We would like to know about outpatient knee replacement surgeries that you may perform.

30. Do you currently perform any knee replacement surgeries on Medicare fee-for-service patients on an outpatient basis?

    - ○ Yes
    - ○ No → **If No, go to #32**

31. How much do hospital guidelines or directives influence your decisions about which Medicare fee-for-service patients are appropriate for hospital **outpatient** knee replacement surgery? *Please mark one.*

    - ○ Hospital guidelines or directives **greatly** influence my decisions about outpatient knee replacement surgery
    - ○ Hospital guidelines or directives **somewhat** influence my decisions about outpatient knee replacement surgery
    - ○ Hospital guidelines or directives **do not** influence my decisions about outpatient knee replacement surgery
    - ○ Hospitals have no guidelines or directives about outpatient knee replacement surgery

Section 7. About You

32. How long have you been an orthopedic surgeon, excluding training?

- ○ Less than 3 years
- ○ 3 years up to 11 years
- ○ 11 years up to 20 years
- ○ More than 20 years

33. In the last 12 months, how many hip and knee replacement surgeries did you perform? Please answer this question thinking about ALL of your patients, not just your Medicare fee-for-service patients. *Your best estimate is fine.*

- ○ Less than 25
- ○ 25 to150
- ○ More than 150

34. At how many hospitals do you perform hip or knee replacement surgeries? Please answer this question only thinking about hospitals, not other surgery locations, such as ambulatory surgery centers.

- ○ One
- ○ Two
- ○ Three or more

35. Which best describes your employment status?

- ○ I am a hospital employee
- ○ I am in a hospital or health system-owned practice
- ○ I am in an academic department or practice
- ○ I am in a physician-owned practice
- ○ I am an independent contractor

36. Do you, or your physician group, participate in any of the following? *Please mark all that apply*.

- ○ Medicare Accountable Care Organization (ACO)
- ○ Medicare Bundled Payment for Care Improvement (BPCI or BPCI-Advanced)
- ○ Value-based payment models run by commercial payers
- ○ None of the above

Appendix N: Subpopulation Analysis Results

Claims-based sample sizes and risk-adjusted average outcomes

Exhibit N-1: Claims-based sample sizes and risk-adjusted average outcome values for Black or African American and White beneficiaries, mandatory CJR hospitals, LEJR episodes, PY1-4

	Intervention episodes				Baseline risk-adjusted average outcome value				Intervention risk-adjusted average outcome value			
	Black or African American		White		Black or African American		White		Black or African American		White	
Outcome	CJR	Control	CJR	Control	CJR	Control	CJR	Control	CJR	Control	CJR	Control
Unplanned readmissions	8,983	12,383	130,054	156,247	10.5%	10.1%	9.1%	8.8%	9.5%	9.6%	8.8%	8.8%
ED use	8,983	12,383	130,054	156,247	18.0%	17.4%	12.9%	12.4%	18.9%	18.6%	13.8%	13.3%
Mortality	9,097	12,220	132,980	153,066	1.6%	1.6%	2.8%	2.9%	1.3%	1.8%	2.6%	2.6%
Total episode payments	8,989	12,389	130,087	156,275	$30,723	$29,005	$28,947	$28,575	$27,248	$27,965	$26,235	$27,267

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

Note: ED = emergency department, LEJR = lower extremity joint replacement, PY = performance year.

Exhibit N-2: Claims-based sample sizes and risk-adjusted average outcome values for dually eligible and non-dually eligible beneficiaries, mandatory CJR hospitals, LEJR episodes, PY1-4

Outcome	Intervention episodes				Baseline risk-adjusted average outcome value				Intervention risk-adjusted average outcome value			
	Dually eligible		Non-dually eligible		Dually eligible		Non-dually eligible		Dually eligible		Non-dually eligible	
	CJR	Control	CJR	Control	CJR	Control	CJR	Control	CJR	Control	CJR	Control
Unplanned readmissions	18,001	16,523	135,766	162,733	13.5%	12.8%	8.5%	8.3%	12.7%	12.8%	8.1%	8.2%
ED use	18,001	16,523	135,766	162,733	19.2%	19.9%	12.2%	11.6%	20.1%	20.8%	13.2%	12.5%
Mortality	18,582	16,654	138,458	158,933	3.8%	4.3%	2.4%	2.5%	3.7%	3.9%	2.3%	2.3%
Total episode payments	18,017	16,537	135,796	162,754	$35,383	$33,989	$28,261	$27,839	$32,680	$33,303	$25,439	$26,459

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

Note: ED = emergency department, LEJR = lower extremity joint replacement, PY = performance year.

Exhibit N-3: Claims-based sample sizes and risk-adjusted average outcome values for Black or African American *and* dually eligible and White *and* non-dually eligible beneficiaries, mandatory CJR hospitals, LEJR episodes, PY1-4

Outcome	Intervention episodes				Baseline risk-adjusted average outcome value				Intervention risk-adjusted average outcome value			
	Black or African American *and* dually eligible		White *and* non-dually eligible		Black or African American *and* dually eligible		White *and* non-dually eligible		Black or African American *and* dually eligible		White *and* non-dually eligible	
	CJR	Control	CJR	Control	CJR	Control	CJR	Control	CJR	Control	CJR	Control
Unplanned readmissions	2,922	3,887	120,679	146,439	13.8%	12.3%	8.5%	8.3%	12.6%	11.2%	8.3%	8.3%
ED use	2,922	3,887	120,679	146,439	23.8%	23.7%	12.2%	11.7%	24.2%	25.7%	13.2%	12.6%
Mortality	2,964	3,893	123,187	143,110	2.0%	1.9%	2.5%	2.6%	1.6%	2.1%	2.4%	2.4%
Total episode payments	2,925	3,891	120,701	146,457	$33,140	$30,913	$28,279	$27,862	$30,402	$30,390	$25,533	$26,489

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

Note: ED = emergency department, LEJR = lower extremity joint replacement, PY = performance year.

Differential Impact Results

Exhibit N-4: Claims-based differential impact analysis results for Black or African American beneficiaries compared to White beneficiaries, mandatory CJR hospitals, LEJR episodes, PY1-4

Outcome	Baseline CJR difference	DiD impact estimates for subpopulations		Differential impact estimate [a]	Differential impact as % of baseline gap	p-value	90% CI
		Black or African American	White				
Unplanned readmissions	1.5pp	-0.5pp	-0.2pp	-0.3pp	-18.6%	p=0.62	[-1.2 to 0.7]
ED use	5.1pp	-0.3pp	0.1pp	-0.4pp	-7.3%	p=0.68	[-1.9 to 1.1]
Mortality	-1.2pp	-0.4pp	0.1pp	-0.5pp	39.7%	p<0.05	[-0.8 to -0.1]
Total episode payments	$1,776	-$2,435	-$1,404	-$1,031	-58.1%	p<0.05	[-$1,851 to -$212]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

Notes: The estimates in this exhibit are the result of a triple difference model. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The baseline CJR difference is calculated as the difference in risk-adjusted average outcome values between Black or African American patients at CJR hospitals and White patients at CJR hospitals.

Because CJR participant hospitals shifted a lower share of TKAs to the hospital outpatient setting, the control group includes outpatient TKA episodes that have been weighted to balance the episode volume in the CJR hospitals.

CI = confidence interval, DiD = difference-in-differences, ED = emergency department, LEJR = lower extremity joint replacement, pp = percentage point, PY = performance year, TKA = total knee arthroplasty.

[a] Positive values indicate that CJR is associated a more favorable impact for Black or African American patients than for White patients.

Exhibit N-5: Claims-based differential impact analysis results for dually eligible beneficiaries compared to non-dually eligible beneficiaries, mandatory CJR hospitals, LEJR episodes, PY1-4

Outcome	Baseline CJR difference	DiD impact estimates for subpopulations		Differential impact estimate [a]	Differential impact as % of baseline gap	p-value	90% CI
		Dually eligible	Non-dually eligible				
Unplanned readmissions	5.0pp	-0.7pp	-0.3pp	-0.5pp	-9.6%	p=0.28	[-1.2 to 0.2]
ED use	6.9pp	-0.0pp	0.1pp	-0.1pp	-2.0%	p=0.83	[-1.2 to 0.9]
Mortality	1.3pp	0.3pp	-0.0pp	0.3pp	24.7%	p=0.21	[-0.1 to 0.8]
Total episode payments	$7,122	-$2,017	-$1,441	-$576	-8.1%	p=0.14	[-$1,224 to $71]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a triple difference model. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The baseline CJR difference is calculated as the difference in risk-adjusted average outcome values between dually eligible patients at CJR hospitals and non-dually eligible patients at CJR hospitals.

Because CJR participant hospitals shifted a lower share of TKAs to the hospital outpatient setting, the control group includes outpatient TKA episodes that have been weighted to balance the episode volume in the CJR hospitals.

CI = confidence interval, DiD = difference-in-differences, ED = emergency department, LEJR = lower extremity joint replacement, pp = percentage point, PY = performance year, TKA = total knee arthroplasty.

[a] Positive values indicate that CJR is associated a more favorable impact for dually eligible patients than for non-dually eligible patients.

Exhibit N-6: Claims-based differential impact analysis results for Black or African American *and* dually eligible beneficiaries compared to White *and* non-dually eligible beneficiaries, mandatory CJR hospitals, LEJR episodes, PY1-4

Outcome	Baseline CJR difference	DiD impact estimates for subpopulations		Differential impact estimate	Differential impact as % of baseline gap	p-value	90% CI
		Black or African American *and* dually eligible	White *and* non-dually eligible				
Unplanned readmissions	5.2pp	0.0pp	-0.2pp	0.2pp	4.0%	p=0.82	[-1.3 to 1.7]
ED use	11.6pp	-1.5pp	0.1pp	-1.6pp	-14.1%	p=0.26	[-4.0 to 0.8]
Mortality	-0.6pp	-0.6pp	-0.0pp	-0.6pp	103.1%	p=0.12	[-1.3 to 0.0]
Total episode payments	$4,861	-$2,215	-$1,374	-$841	-17.3%	p=0.21	[-$1,958 to $276]

Source: CJR evaluation team analysis of Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015 (baseline) and episodes initiated during or after April 2016 that ended by December 2019 (intervention).

***Notes*:** The estimates in this exhibit are the result of a triple difference model. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

The baseline CJR difference is calculated as the difference in risk-adjusted average outcome values between Black or African American *and* dually eligible patients at CJR hospitals and White *and* non-dually eligible patients at CJR hospitals.

Because CJR participant hospitals shifted a lower share of TKAs to the hospital outpatient setting, the control group includes outpatient TKA episodes that have been weighted to balance the episode volume in the CJR hospitals.

CI = confidence interval, DiD = difference-in-differences, ED = emergency department, LEJR = lower extremity joint replacement, pp = percentage point, PY = performance year, TKA = total knee arthroplasty.

[a] Positive values indicate that CJR is associated a more favorable impact for Black or African American *and* dually eligible patients than for White *and* non-dually eligible patients.

Exhibit N-7: Survey-based differential impact analysis for Black or African American respondents compared to White respondents, functional status and pain measures, LEJR patients discharged from mandatory hospitals

Survey measure	Response range[a]	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Black or African American	White			
Ability to walk by yourself without resting	-4 to 4	0.01	0.06	-0.04	p=0.54	[-0.16 to 0.08]
Difficulty walking up or down 12 stairs	-3 to 3	0.04	-0.01	0.05	p=0.46	[-0.06 to 0.16]
Difficulty rising from sitting	-4 to 4	0.03	0.00	0.02	p=0.73	[-0.09 to 0.14]
Difficulty standing	-4 to 4	0.09	-0.01	0.10	p=0.12	[-0.00 to 0.20]
Use of a mobility aid	-2 to 2	0.03	0.01	0.02	p=0.71	[-0.06 to 0.10]
Difficulty getting on/off the toilet	-4 to 4	0.05	-0.01	0.06	p=0.29	[-0.04 to 0.16]
Frequency that pain interferes with normal activities	-4 to 4	0.06	-0.01	0.07	p=0.36	[-0.05 to 0.18]
Medication use for pain in the joint you had replaced	-3 to 3	-0.03	0.02	-0.05	p=0.32	[-0.13 to 0.03]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement.

[a] The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Estimated changes, and the difference between changes in the CJR and control group, are reported in "level" terms (that is, levels of the Likert scale for each measure).

Exhibit N-8: Survey-based differential impact analysis for Black or African American respondents compared to White respondents, satisfaction measures, LEJR patients discharged from mandatory hospitals

Survey measure	Response range[a]	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Black or African American	White			
Satisfaction with overall recovery since leaving the hospital	0 to 100	2.23	-0.01	2.24	p=0.29	[-1.23 to 5.71]
Composite measure of satisfaction with care management	0 to 100	2.87	-0.32	3.18	p<0.10	[0.41 to 5.95]
Health care providers listened to preferences	0 to 100	6.26	-0.65	6.91	p<0.01	[3.34 to 10.49]
Satisfaction with discharge destination	0 to 100	1.39	0.18	1.22	p=0.61	[-2.78 to 5.21]
Satisfaction with care coordination	0 to 100	1.07	-0.45	1.52	p=0.48	[-2.04 to 5.09]
Satisfaction with treatment instructions	0 to 100	2.46	-0.52	2.98	p=0.13	[-0.26 to 6.22]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement.

[a] Satisfaction outcomes are scaled from 0 to 100 points, where 0 = very dissatisfied, 25 = dissatisfied, 50 = neutral, 75 = satisfied, and 100 = very satisfied. The composite summarizes the level of satisfaction across the four measures of care management. Differences between CJR and control outcomes are reported in point terms.

Exhibit N-9: Survey-based differential impact analysis for Black or African American respondents compared to White respondents, care transitions measures, LEJR patients discharged from mandatory hospitals

Survey measure	Response range[a]	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Black or African American	White			
Discharged from the hospital at the right time	Yes	0.39pp	-0.76pp	1.15pp	p=0.55	[-2.01 to 4.30]
	No – Too Soon	0.02pp	-0.06pp	0.08pp	p=0.40	[-0.08 to 0.25]
	No – Too Late	-0.41pp	0.82pp	-1.23pp	p=0.54	[-4.54 to 2.08]
Received the right amount of post-discharge care	Yes	-0.27pp	-0.86pp	0.59pp	p=0.79	[-3.14 to 4.32]
	No – Too Little	0.34pp	1.19pp	-0.85pp	p=0.76	[-5.50 to 3.80]
	No – Too Much	-0.07pp	-0.33pp	0.26pp	p=0.64	[-0.67 to 1.19]
Had all the medical equipment needed at home	Yes	-1.78pp	-1.13pp	-0.64pp	p=0.77	[-4.32 to 3.03]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement, pp = percentage point.

[a] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

Exhibit N-10: Survey-based differential impact analysis for Black or African American respondents compared to White respondents, caregiver help measures, LEJR patients discharged from mandatory hospitals

Survey measure	Response range	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Black or African American	White			
Received any caregiver help [a]	Yes	1.58pp	-0.61pp	2.20pp	p<0.10	[0.35 to 4.05]
Composite measure of caregiver help [b]	0 to 100	0.08	-1.43	1.51	p=0.43	[-1.68 to 4.69]
Help needed putting on or taking off clothes [b]	0 to 100	-0.41	-1.89	1.48	p=0.41	[-1.51 to 4.48]
Help needed bathing [b]	0 to 100	0.79	-1.43	2.23	p=0.37	[-1.87 to 6.32]
Help needed using the toilet [b]	0 to 100	0.50	-1.06	1.56	p=0.30	[-0.94 to 4.06]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement, pp = percentage point.

[a] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

[b] Respondents were only asked about the amount of help needed with a given activity of daily living if they indicated that they received caregiver help. Measures of caregiver help required among respondents who received any help are scaled from 0 to 100 points, where 0 = complete help needed, 50 = some help needed, and 100 = no help needed. The composite summarizes the amount of help needed across all three activities of daily living. Differences between CJR and control outcomes are reported in point terms.

Exhibit N-11: Survey-based differential impact analysis for dually eligible respondents compared to non-dually eligible respondents, functional status measures, LEJR patients discharged from mandatory hospitals

Survey measure	Response range [a]	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Dually eligible	Non-dually eligible			
Ability to walk by yourself without resting	-4 to 4	0.04	0.04	-0.01	p=0.92	[-0.13 to 0.12]
Difficulty walking up or down 12 stairs	-3 to 3	0.09	-0.02	0.11	p<0.05	[0.02 to 0.20]
Difficulty rising from sitting	-4 to 4	0.06	-0.00	0.07	p=0.31	[-0.04 to 0.17]
Difficulty standing	-4 to 4	0.04	-0.00	0.04	p=0.54	[-0.07 to 0.15]
Use of a mobility aid	-2 to 2	-0.01	0.01	-0.01	p=0.77	[-0.09 to 0.07]
Difficulty getting on/off the toilet	-4 to 4	0.03	-0.01	0.04	p=0.57	[-0.07 to 0.15]
Frequency that pain interferes with normal activities	-4 to 4	0.06	-0.00	0.06	p=0.48	[-0.08 to 0.20]
Medication use for pain in the joint you had replaced	-3 to 3	0.01	0.02	-0.01	p=0.89	[-0.15 to 0.12]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement.

[a] The change in a given measure of functional status refers to the difference between a respondent's self-reported status at the time of the survey and the respondent's recalled status in the week prior to hospitalization. Estimated changes, and the difference between changes in the CJR and control group, are reported in "level" terms (that is, levels of the Likert scale for each measure).

Exhibit N-12: Survey-based differential impact analysis for dually eligible respondents compared to non-dually eligible respondents, satisfaction measures, LEJR patients discharged from mandatory hospitals

Survey measure	Response range[a]	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Dually eligible	Non-dually eligible			
Satisfaction with overall recovery since leaving the hospital	0 to 100	-0.59	-0.02	-0.57	p=0.79	[-4.04 to 2.91]
Composite measure of satisfaction with care management	0 to 100	-0.99	-0.16	-0.83	p=0.67	[-4.01 to 2.35]
Health care providers listened to preferences	0 to 100	-0.87	-0.38	-0.50	p=0.84	[-4.51 to 3.51]
Satisfaction with discharge destination	0 to 100	-1.35	0.32	-1.67	p=0.43	[-5.13 to 1.80]
Satisfaction with care coordination	0 to 100	-1.66	-0.21	-1.46	p=0.52	[-5.19 to 2.28]
Satisfaction with treatment instructions	0 to 100	-1.23	-0.31	-0.92	p=0.69	[-4.78 to 2.94]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement.

[a] Satisfaction outcomes are scaled from 0 to 100 points, where 0 = very dissatisfied, 25 = dissatisfied, 50 = neutral, 75 = satisfied, and 100 = very satisfied. The composite summarizes the level of satisfaction across the four measures of care management. Differences between CJR and control outcomes are reported in point terms.

Exhibit N-13: Survey-based differential impact analysis for dually eligible respondents compared to non-dually eligible respondents, care transitions measures, LEJR patients discharged from mandatory hospitals

Survey measure	Response range[a]	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Dually eligible	Non-dually eligible			
Discharged from the hospital at the right time	Yes	-0.07pp	-0.49pp	0.42pp	p=0.83	[-2.88 to 3.72]
	No – Too Soon	-0.00pp	-0.04pp	0.03pp	p=0.75	[-0.14 to 0.21]
	No – Too Late	0.07pp	0.52pp	-0.45pp	p=0.83	[-3.92 to 3.02]
Received the right amount of post-discharge care	Yes	0.21pp	-0.66pp	0.87pp	p=0.51	[-1.32 to 3.05]
	No – Too Little	-0.28pp	0.90pp	-1.18pp	p=0.51	[-4.14 to 1.78]
	No – Too Much	0.07pp	-0.24pp	0.31pp	p=0.50	[-0.46 to 1.09]
Had all the medical equipment needed at home	Yes	-0.33pp	-0.96pp	0.62pp	p=0.78	[-3.08 to 4.32]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement, pp = percentage point.

[a] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

Exhibit N-14: Survey-based differential impact analysis for dually eligible respondents compared to non-dually eligible respondents, caregiver help measures, LEJR patients discharged from mandatory hospitals

Caregiver help measure	Response range	Impact estimate by subpopulation		Differential impact of CJR model	p-value	90% CI
		Dually eligible	Non-dually eligible			
Received any caregiver help [a]	Yes	-1.00pp	-0.38pp	-0.62pp	p=0.77	[-4.21 to 2.96]
Composite measure of caregiver help [b]	0 to 100	-2.03	-1.28	-0.74	p=0.70	[-3.93 to 2.45]
Help needed putting on or taking off clothes [b]	0 to 100	-0.66	-1.68	1.03	p=0.62	[-2.43 to 4.48]
Help needed bathing [b]	0 to 100	-1.98	-1.20	-0.78	p=0.76	[-5.00 to 3.44]
Help needed using the toilet [b]	0 to 100	-2.63	-1.10	-1.54	p=0.44	[-4.83 to 1.76]

Source: CJR evaluation team analysis of patient survey data for episodes with discharge in March, April, September, or October 2019.

Notes: The estimates in this exhibit are the result of a cross-sectional regression model, weighted for sampling and nonresponse. Estimates that are significant at the 1%, 5%, or 10% significance level are indicated by red, orange, or yellow shaded cells, respectively.

CI = confidence interval, LEJR = lower extremity joint replacement, pp = percentage point.

[a] Indicates binary measure, reported as the percent of respondents reporting "Yes" to a given measure. Differences between CJR and control outcomes are reported in percentage point terms.

[b] Respondents were only asked about the amount of help needed with a given activity of daily living if they indicated that they received caregiver help. Measures of caregiver help required among respondents who received any help are scaled from 0 to 100 points, where 0 = complete help needed, 50 = some help needed, and 100 = no help needed. The composite summarizes the amount of help needed across all three activities of daily living. Differences between CJR and control outcomes are reported in point terms.