LEWIN GROUP®

CMS Comprehensive Care for Joint Replacement Model: Performance Year 3 Evaluation Report – Appendices

Third 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**

November 2020

CMS Comprehensive Care for Joint Replacement Model: Performance Year 3 Evaluation Report – Appendices

Third 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.

November 2020

List of Appendices

Appendix A: List of Acronyms & Glossary Terms

Exhibit A-1: List of acronyms

Acronym	Meaning
AAHKS	American Association of Hip and Knee Surgeons
AAOS	American Association of Orthopedic Surgeons
ACH	Acute Care Hospital
ADLs	Activities of Daily Living
AM-PAC	Activity Measure for Post-Acute Care
APM	Alternative Payment Model
ASA	American Society of Anesthesiologists
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
CY	Calendar Year
DiD	Difference-in-Differences
DME	Durable Medical Equipment
DSH	Disproportionate Share Hospital
ED	Emergency Department
EMR	Electronic Medical Record
ERAS	Enhanced Recovery After Surgery
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
HHI	Herfindahl-Hirschman Index
IP	Inpatient
IPO	Inpatient Only
IPPS	Inpatient Prospective Payment System
IQR	Interquartile Range
IRF	Inpatient Rehabilitation Facility
IRF-PAI	Inpatient Rehabilitation Facility-Patient Assessment Instrument
IT	Information Technology
LACE	Length of Stay, Acuity of Admission, Comorbidities, and Emergency Room Visits

Acronym	Meaning
LEJR	Lower Extremity Joint Replacement
LOS	Length of Stay
MA	Medicare Advantage
MCC	Major Complication or Comorbidity
MDS	Minimum Data Set
MSA	Metropolitan Statistical Area
MS-DRG	Medicare Severity-Diagnosis Related Group
MUA	Manipulation Under Anesthesia
NPRA	Net Payment Reconciliation Amount
OASIS	Outcome and Assessment Information Set
OP	Outpatient
PAC	Post-Acute Care
PDGM	Patient-Driven Groupings Model
PDPM	Patient-Driven Payment Model
PGP	Physician Group Practice
PRO	Patient-Reported Outcomes
PY	Performance Year
RAC	Recovery Audit Contractors
RAPT	Risk Assessment and Prediction Tool
SNF	Skilled Nursing Facility
THA	Total Hip Arthroplasty
TKA	Total Knee Arthroplasty

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.
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 are: 0.5% for "excellent" quality, 1% for "good" quality, and 2% for "acceptable" or "below acceptable" quality. These effective discount percentages for hospitals with repayment responsibility will increase in PY4/5 (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
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.
Post-acute care (PAC)	Rehabilitation and palliative care services received by the beneficiary from IRFs, SNFs, HHAs, or LTCHs following a hospitalization.
Post-episode care	Under the CJR model, care that occurs after the 90-day post-discharge period.
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.
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. By 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 PY 2-5.
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 (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.
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.
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: Payment, Utilization, Quality, and Activities of Daily Living Results

Exhibit B-1: Risk-adjusted claims-based difference-in-differences results for payment, utilization, and quality metrics, mandatory CJR hospitals, inpatient LEJR episodes, PY1-3 (overestimated)

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90th percentile LCI	90th percentile UCI
		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	104,482	109,650	$29,118	$26,506	$28,634	$27,562	-$1,540	-5.3%	p<0.01	-$2,112	-$968
	IRF payments	104,482	109,650	$2,215	$1,240	$2,136	$1,701	-$539	-24.4%	p<0.01	-$849	-$230
	SNF payments	104,482	109,650	$6,085	$4,281	$6,154	$5,285	-$935	-15.4%	p<0.01	-$1,278	-$592
	HH payments[a]	104,482	109,650	$2,415	$2,510	$2,298	$2,305	$88	3.6%	p=0.47	-$111	$288
	Readmission payments	104,482	109,650	$1,202	$1,043	$1,105	$1,085	-$140	-11.6%	p<0.10	-$258	-$21
	Part B payments	104,482	109,650	$4,979	$4,879	$4,814	$4,818	-$104	-2.1%	p<0.10	-$207	$0
	30-day PEP payments[a]	104,482	109,650	$1,478	$1,476	$1,488	$1,521	-$35	-2.4%	p=0.19	-$80	$9
	Anchor payments[a]	104,482	109,650	$12,186	$12,180	$12,189	$12,153	$30	0.2%	p=0.55	-$52	$111
Utilization	Anchor hospitalization LOS	104,150	109,056	3.5	2.9	3.4	2.8	0.0	-0.7%	p=0.62	-0.1	0.1
	First PAC IRF	104,482	109,650	13.7%	6.3%	12.8%	9.0%	-3.6	-26.3%	p<0.01	-5.7	-1.5
	First PAC SNF	104,482	109,650	41.2%	30.2%	42.2%	34.9%	-3.6	-8.8%	p<0.05	-6.0	-1.2
	First PAC HH	104,482	109,650	36.7%	49.4%	33.6%	38.2%	8.1	21.9%	p<0.05	2.6	13.6
	First PAC home without HH	104,482	109,650	8.4%	14.1%	11.4%	17.9%	-0.8	-9.8%	p=0.71	-4.5	2.9
	Any HH use[a]	104,482	109,650	71.8%	73.7%	69.2%	66.3%	4.7	6.6%	p=0.18	-1.0	10.5
	IRF days	7,309	9,953	11.7	11.7	11.5	11.7	-0.2	-1.3%	p=0.39	-0.5	0.1
	SNF days	33,356	34,589	26.8	22.1	26.8	24.6	-2.5	-9.4%	p<0.01	-3.4	-1.6
	HH visits	76,408	74,092	17.0	15.8	16.7	16.5	-1.0	-5.9%	p<0.05	-1.7	-0.3
	HH PT/OT visits	76,408	74,092	10.4	10.4	10.3	10.8	-0.6	-5.5%	p=0.14	-1.2	0.1
	Outpatient PT/OT visits[a]	71,672	73,534	13.0	14.0	13.1	14.0	0.1	0.7%	p=0.66	-0.2	0.4

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90th percentile LCI	90th percentile UCI
		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	104,447	109,629	10.4%	9.0%	10.1%	9.0%	-0.3	-3.1%	p<0.10	-0.6	0.0
	ED use	104,447	109,629	13.2%	14.1%	12.7%	13.6%	0.0	0.2%	p=0.94	-0.5	0.5
	Mortality rate	106,720	111,767	2.6%	2.5%	2.8%	2.6%	0.1	2.0%	p=0.58	-0.1	0.2
	Complications[b]	89,767	96,286	3.2%	2.7%	3.0%	2.7%	-0.2	-7.4%	p<0.10	-0.4	0.0
	MUA[b]	57,142	62,255	1.9%	1.9%	2.0%	1.9%	0.0	1.1%	p=0.88	-0.2	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 2018 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 99%, 95%, or 90% 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.

An inpatient LEJR episode begins with an inpatient anchor hospitalization that meets CJR episode eligibility requirements and ends 90 days after discharge. Inpatient LEJR episodes are CJR episodes under the CJR model.

Impact estimates based on inpatient LEJR episodes overestimate the impact of the CJR model.

DiD = difference-in-differences, ED = emergency department, HH = home health, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, LOS = length of stay, MUA = manipulation under anesthesia, OT = occupational therapy, PAC = post-acute care, PEP = post-episode period, PT = physical therapy, PY = performance year, SNF = skilled nursing facility.

[a] Results for the HH payments, 30-day PEP payments, anchor payments, any HH use, and outpatient PT/OT visits measures need to be interpreted with caution because data from the baseline period showed CJR and control group hospitals were not on parallel trends for these outcomes, which is required for an unbiased estimate.

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

Exhibit B-2: Risk-adjusted assessment-based difference-in-differences results for activities of daily living metrics, mandatory CJR hospitals, inpatient LEJR episodes, April 2016-September 2018

First PAC setting	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90th percentile LCI	90th percentile UCI
		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	5,325	7,517	10.7	11.2	10.2	10.9	-0.2	-1.6%	p=0.47	-0.6	0.2
SNF	Improved transfer, locomotion on unit, and walking in corridor	22,697	23,771	65.1%	66.1%	68.9%	70.1%	-0.1	-0.1%	p=0.96	-2.3	2.1
	Improved toilet use	22,723	23,713	44.0%	41.0%	46.9%	47.6%	-3.7	-8.4%	p<0.10	-7.2	-0.2
	Without self-reported pain[a]	21,845	22,668	58.2%	74.8%	53.0%	66.4%	3.2	5.6%	p<0.05	0.7	5.8
HHA	Improved ambulation/ locomotion	38,996	37,050	89.5%	90.4%	89.6%	90.6%	-0.1	-0.1%	p=0.81	-1.0	0.7
	Improved bed transferring	38,811	36,855	83.6%	84.8%	83.3%	85.1%	-0.5	-0.6%	p=0.57	-2.1	1.0
	Reduced pain	38,798	36,914	74.8%	83.0%	74.8%	82.7%	0.3	0.4%	p=0.81	-1.8	2.3

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 2018 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 99%, 95%, or 90% 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.

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.

Exhibit B-3: Risk-adjusted claims-based difference-in-differences results for payment, utilization, and quality metrics, mandatory CJR hospitals, all LEJR episodes, PY1-3 (underestimated)

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90th percentile LCI	90th percentile UCI
		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	108,038	115,812	$29,039	$26,314	$28,554	$27,207	-$1,378	-4.7%	p<0.01	-$1,927	-$828
	IRF payments	108,038	115,812	$2,202	$1,238	$2,128	$1,664	-$499	-22.7%	p<0.01	-$799	-$199
	SNF payments	108,038	115,812	$6,049	$4,185	$6,117	$5,118	-$865	-14.3%	p<0.01	-$1,202	-$528
	HH payments[a]	108,038	115,812	$2,415	$2,506	$2,295	$2,302	$85	3.5%	p=0.50	-$120	$289
	Readmission payments	108,038	115,812	$1,195	$1,034	$1,098	$1,069	-$132	-11.0%	p<0.10	-$250	-$14
	Part B payments	108,038	115,812	$4,975	$4,845	$4,797	$4,740	-$73	-1.5%	p=0.24	-$176	$29
	30-day PEP payments[a]	108,038	115,812	$1,470	$1,473	$1,480	$1,511	-$29	-2.0%	p=0.28	-$72	$15
	Anchor payments[a]	108,038	115,812	$12,179	$12,116	$12,178	$12,049	$66	0.5%	p=0.23	-$25	$156
Utilization	First PAC IRF	108,038	115,812	13.6%	6.3%	12.8%	8.9%	-3.5	-25.3%	p<0.01	-5.6	-1.4
	First PAC SNF	108,038	115,812	41.0%	29.5%	42.1%	33.6%	-3.1	-7.6%	p<0.05	-5.4	-.8
	First PAC HH	108,038	115,812	36.9%	49.8%	33.7%	38.9%	7.7	20.8%	p<0.05	2.2	13.2
	First PAC home without HH	108,038	115,812	8.4%	14.4%	11.5%	18.5%	-1.1	-13.4%	p=0.63	-5.0	2.7
	Any HH use[a]	108,038	115,812	71.8%	73.7%	69.1%	66.2%	4.7	6.6%	p=0.19	-1.2	10.7
	IRF days	7,477	10,181	11.7	11.7	11.5	11.7	-0.1	-1.2%	p=0.43	-0.4	0.2
	SNF days	33,451	34,701	26.8	22.1	26.8	24.6	-2.5	-9.4%	p<0.01	-3.4	-1.6
	HH visits	79,003	78,053	17.0	15.6	16.6	16.3	-1.0	-5.7%	p<0.05	-1.7	-0.2
	HH PT/OT visits	79,003	78,053	10.4	10.4	10.3	10.8	-0.6	-5.8%	p=0.12	-1.2	0.0
	Outpatient PT/OT visits[a]	74,851	78,827	13.0	14.0	13.2	14.1	0.1	0.7%	p=0.67	-0.2	0.4

Domain	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90th percentile LCI	90th percentile UCI
		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	108,003	115,791	10.4%	8.9%	10.1%	8.9%	-0.3	-2.9%	p<0.10	-0.6	0.0
	ED use	108,003	115,791	13.2%	14.1%	12.7%	13.6%	0.1	0.5%	p=0.82	-0.4	0.5
	Mortality rate	110,280	117,940	2.6%	2.5%	2.7%	2.5%	0.1	2.1%	p=0.54	-0.1	0.2
	Complications[b]	93,323	102,448	3.2%	2.6%	3.0%	2.7%	-0.2	-6.5%	p=0.10	-0.4	0.0
	MUA[b]	60,698	68,417	1.9%	1.8%	2.0%	1.9%	0.0	0.4%	p=0.95	-0.2	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 2018 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 99%, 95%, or 90% 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.

All LEJRs (IP+OP) include inpatient LEJR episodes and outpatient TKAs. An inpatient LEJR episode begins with an inpatient anchor hospitalization that meets CJR episode eligibility requirements and ends 90 days after discharge. Inpatient LEJR episodes are CJR episodes under the CJR model. An outpatient TKA begins with an outpatient anchor procedure that meets CJR episode eligibility requirements and ends 90 days after the date of service for the anchor procedure. Outpatient TKAs are not CJR episodes under the CJR model.

Impact estimates based on all LEJRs (IP+OP) underestimate the impact of the CJR model.

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

[a] Results for the HH payments, 30-day PEP payments, anchor payments, any HH use, and outpatient PT/OT visits measures need to be interpreted with caution because data from the baseline period showed CJR and control group hospitals were not on parallel trends for these outcomes, which is required for an unbiased estimate.

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

Exhibit B-4: Risk-adjusted assessment-based difference-in-differences results for activities of daily living metrics, mandatory CJR hospitals, all LEJR episodes, April 2016-September 2018

First PAC setting	Measure	CJR	Control group	CJR		Control group		DiD	DiD % of baseline	p-value	90th percentile LCI	90th percentile UCI
		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	5,414	7,616	10.5	11.0	10.1	10.8	-0.2	-1.7%	p=0.46	-0.6	0.2
SNF	Improved transfer, locomotion on unit, and walking in corridor	22,729	23,791	65.6%	67.0%	69.5%	70.7%	0.2	0.3%	p=0.89	-2.0	2.3
	Improved toilet use	22,755	23,733	43.4%	40.4%	46.4%	47.0%	-3.7	-8.6%	p<0.10	-7.2	-0.2
	Without self-reported pain[a]	21,878	22,688	58.2%	74.8%	53.0%	66.4%	3.2	5.6%	p<0.05	0.7	5.8
HHA	Improved ambulation/ locomotion	40,199	39,121	89.6%	90.5%	89.7%	90.8%	-0.2	-0.2%	p=0.67	-1.0	0.6
	Improved bed transferring	40,010	38,913	83.9%	85.2%	83.7%	85.5%	-0.5	-0.6%	p=0.57	-2.0	1.0
	Reduced pain	39,999	38,983	74.3%	82.7%	74.3%	82.5%	0.2	0.3%	p=0.87	-1.9	2.3

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 2018 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 99%, 95%, or 90% 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.

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.

Appendix C: 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 prior to 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 its web site, which is updated quarterly.

[1] www.medicare.gov/NursingHomeCompare/

Appendix D: Definitions of MSA, Hospital, and Patient Characteristics

Exhibit D-1: MSA characteristic variable definitions

Variable	Definition	Source
Annual number of LEJR discharges	Average number of LEJR discharges annually in a given MSA	2012-2014 Medicare Claims
HHI for LEJR at ACH	Sum of the squared LEJR market shares of all ACH providers (CJR and control group), multiplied by 10,000. The HHI values can range from 0 (large number of firms in the market) to 10,000 (a single firm controls the market). Values between 1,500 and 2,500 indicate moderately concentrated markets and values greater than 2,500 are considered highly concentrated.	2012-2014 Medicare Claims
Number of hospitals performing LEJR	Number of hospitals performing LEJR in a given MSA	2012-2014 Medicare Claims
Population size	Census population estimates for a given MSA	2014 American Community Survey (5-Year Estimates)

Note: ACH = acute care hospital, HHI = Herfindahl-Hirschman Index, LEJR = lower extremity joint replacement, MSA = metropolitan statistical area.

Exhibit D-2: Hospital characteristic variable definitions

Variable	Definition	Source
Annual number of LEJR discharges	Total number of LEJR discharges (MS-DRG 469 or 470) initiated at the hospital	January 2012 - December 2014 Medicare Claims
Annual number of LEJR episodes	Total number of LEJR episodes initiated at the hospital meeting CJR eligibility criteria	January 2012 - December 2014 Medicare Claims
Annual number of Medicare discharges (all MS-DRGs)	Total number of Medicare discharges (LEJR and non-LEJR) at the hospital	January 2012 - December 2014 Medicare Claims
Bed count	Number of beds	FY 2016 CMS Annual IPPS Final Rule data (based on FY 2012-2013 cost report data)
Census region	Location of hospital among four Census Regions	December 2016 CMS POS file
DSH patient percentage	The sum of the percentage of Medicare inpatient days attributable to patients eligible for both Medicare Part A and Supplemental Security Income (SSI), and the percentage of total inpatient days attributable to patients eligible for Medicaid but not Medicare Part A	FY 2016 CMS Annual IPPS Final Rule data (based on FY 2012-2013 cost report and Social Security Administration data)
First PAC in an institutional setting for LEJR discharges	Percent of LEJR discharges at the hospital that were first discharged to a SNF or an IRF within 5 days of discharge	January 2012 - December 2014 Medicare Claims
First PAC IRF for LEJR episodes	Percent of LEJR episodes at the hospital first discharged to a IRF within 5 days of discharge	January 2012 - December 2014 Medicare Claims

Variable	Definition	Source
First PAC SNF for LEJR episodes	Percent of LEJR episodes at the hospital first discharged to a SNF within 5 days of discharge	January 2012 - December 2014 Medicare Claims
First PAC home with HH for LEJR episodes	Percent of LEJR episodes at the hospital first discharged home with HH within 14 days of discharge	January 2012 - December 2014 Medicare Claims
First PAC home without HH for LEJR episodes	Percent of LEJR episodes at the hospital not discharged to a SNF or IRF within 5 days of discharge and not discharged home with HH within 14 days of discharge	January 2012 - December 2014 Medicare Claims
HCC score for LEJR discharges	Average HCC score for LEJR discharges initiated at the hospital. Patients who are healthier than the average Medicare beneficiary have HCC scores of less than 1.0, while patients who are unhealthier than the average Medicare beneficiary have HCC scores of greater than 1.0.	January 2012 - December 2014 Medicare Claims and Enrollment
Historical episode payments relative to PY1 quality-adjusted target price	Whether the hospital started the CJR model with historical episode payments below the PY1 quality-adjusted target price	PY1 CMS payment contractor hospital quality-adjusted target price data
LEJR share of hospital discharges	Percent of the hospital's total discharges that were LEJR discharges	January 2012 - December 2014 Medicare Claims
Medical residents per 1,000 beds	Number of medical residents assigned per 1,000 beds	FY 2016 CMS Annual IPPS Final Rule data (based on FY2012-2013 cost report)
Medicare days percent	Medicare days as a percent of total inpatient days	FY 2016 CMS Annual IPPS Final Rule data (based on FY2012-2013 cost report)
MS-DRG 470 elective discharges	Percent of LEJR discharges initiated at the hospital that were MS-DRG 470 (major joint replacement or reattachment of lower extremity without major complications or comorbidities) without fracture	January 2012 - December 2014 Medicare Claims
Ownership	Ownership type (i.e., for-profit, not-for-profit, government)	December 2016 CMS POS file
Part of a chain	Whether the hospital is part of a chain of providers	December 2014 CMS PECOS
Percent of MSA's LEJR performed at hospital	Percent of LEJR in a given MSA that were performed at the hospital	January 2012 - December 2014 Medicare Claims
Prior Bundled Payments for Care Improvement experience	Whether the hospital ever participated in the Bundled Payments for Care Improvement initiative (LEJR or non-LEJR)	2018 CMS Bundled Payments for Care Improvement Salesforce database
Quality composite score	The CJR model quality composite score ranges from 0 (worst) to 20 (best) and is comprised of: THA/TKA complications rate (weighted 50%); HCAHPS linear mean roll-up measure (weighted 40%); and submission of CJR model PRO data (weighted 10%)	PY1 and PY2 CJR quality performance data

Variable	Definition	Source
Earned reconciliation payment	The hospital earned a reconciliation payment in the performance year because its actual payments were below its quality-adjust target payments and it had "acceptable," "good," or "excellent" quality performance.	PY1, PY2, and PY3 CJR NPRA data
Safety-net hospital	Whether the hospital is in the upper quartiles of DSH patient percentage or UCC per claim, based on national distributions of these variables[1]	FY 2016 CMS Annual IPPS Final Rule data (based on FY2012-2013 cost report for DSH percent and FY2012-2014 for UCC per claim)
Teaching hospital	Whether the hospital has any affiliation with a medical school	December 2016 CMS POS file
Total payment for LEJR episodes	Hospital average standardized total allowed payment, inpatient stay and 90-day post-discharge period, for LEJR episodes	January 2012 - December 2014 Medicare Standardized Payments
Total payment for LEJR discharges	Hospital average standardized total allowed payment, inpatient stay and 90-day post-discharge period, for LEJR discharges	January 2012 - December 2014 Medicare Standardized Payments
Urban	Hospital located in metropolitan or micropolitan statistical area	December 2016 CMS POS file

Note: CMS = Centers for Medicare & Medicaid Services, DSH = disproportionate share hospital, FY = fiscal year, HCAHPS = Hospital Consumer Assessment of Healthcare Providers and Systems, HH = home health, IPPS = inpatient prospective payment system, IRF = inpatient rehabilitation facility, LEJR = lower extremity joint replacement, MSA = metropolitan statistical area, MS-DRG = Medicare Severity-Diagnosis Related Group, NPRA = net payment reconciliation amount, PAC = post-acute care, PECOS = Provider Enrollment, Chain, and Ownership System, POS = Provider of Services, PRO = patient-reported outcomes, PY = performance year, SNF = skilled nursing facility, SSI = Supplemental Security Income, THA = total hip arthroplasty, TKA = total knee arthroplasty, UCC = uncompensated care.

[1] Safety-net definition taken from: Norton EC, Kim J, Das A, Chen LM. Moneyball in Medicare. *NBER Working Paper* No. 22371. Cambridge, MA: NBER.

Exhibit D-3: 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 2018 (intervention) Medicare Enrollment Database
Congestive heart failure	Percent of patients with congestive heart failure (HCC flag #85)	July 2011 – December 2014 (baseline) and April 2015 – December 2018 (intervention) Medicare Claims
Dementia	Percent of patients with dementia (with and without complications; HCC flags #51 and #52)	July 2011 – December 2014 (baseline) and April 2015 – December 2018 (intervention) Medicare Claims
Diabetes	Percent of patients with diabetes	July 2010 – December 2014 (baseline) and April 2014 – December 2018 (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 2018 (intervention) Medicare Enrollment Database
Eligible for Medicaid	Percent eligible for Medicaid based on Medicare enrollment file	January 2012 - December 2014 (baseline) and April 2016 – December 2018 (intervention) Medicare Enrollment Database
Fracture status	Percent of patients with 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 2018 (intervention) Medicare Claims
Sex	Percent of female patients	January 2012 - December 2014 (baseline) and April 2016 – December 2018 (intervention) Medicare Enrollment Database
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 2018 (intervention) Medicare Claims
Hypertension	Percent of patients with hypertension	July 2011 – December 2014 (baseline) and April 2015 – December 2018 (intervention) Medicare Claims
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 2018 (intervention) Medicare Claims

Variable	Definition	Source
Obesity	Percent of patients obese or with a BMI of greater than 30	July 2010 – December 2014 (baseline) and April 2014 – December 2018 (intervention) Medicare Claims
Prior acute care hospital 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 2018 (intervention) Medicare Claims
Prior care use	Percent of patients with any care use (inpatient, skilled nursing facility, inpatient rehabilitation facility, home health, or long-term care hospital) during the six months prior to anchor hospitalization	July 2011 - December 2014 (baseline) and October 2015 – December 2018 (intervention) Medicare Claims
Prior IRF use	Percent of patients with one or more inpatient rehabilitation facility stays during the six months prior to the anchor hospitalization	July 2011 - December 2014 (baseline) and October 2015 – December 2018 (intervention) Medicare Claims
Prior SNF use	Percent of patients with one or more skilled nursing facility stays during the six months prior to the anchor hospitalization	July 2011 - December 2014 (baseline) and October 2015 – December 2018 (intervention) Medicare Claims
Prior HH use	Percent of patients with one or more instances of home health use during the six months prior to the anchor hospitalization	July 2011 - December 2014 (baseline) and October 2015 – December 2018 (intervention) Medicare Claims
Race/ethnicity	Percent of patients by race/ethnicity: White, Black, Hispanic, Other race, Unknown	January 2012 - December 2014 (baseline) and April 2016 – December 2018 (intervention) Medicare Enrollment Database

Note: BMI = body mass index, CMS = Centers for Medicare & Medicaid Services, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, ICD = International Classification of Diseases, IRF = inpatient rehabilitation facility, LTCH = long-term care hospital, MS-DRG = Medicare Severity-Diagnosis Related Group, SNF = skilled nursing facility.

Appendix E: Methodology

I. Data Sources

A. Secondary data sources

Secondary data sources were used to:

1) Characterize Comprehensive Care for Joint Replacement (CJR) and control group markets (Area Health Resource File (AHRF), Fiscal Year (FY) Acute Inpatient Prospective Payment System (IPPS) Final Rule data files, FY Inpatient Rehabilitation Facility (IRF) Prospective Payment System Final Rule data files, Medicare fee-for-service (FFS) claims, and US Census Bureau's American Community Survey data);
2) Identify and characterize CJR participant hospitals and control group hospitals (Provider of Services file, Acute IPPS Final Rule data files, Medicare FFS claims, CJR programmatic data, Bundled Payments for Care Improvement Salesforce Database, and Medicare Provider Enrollment, Chain, and Ownership System);
3) Sample CJR participant hospitals for participation in site visits and telephone interviews (CJR programmatic data, Medicare FFS claims, Agency for Healthcare Research and Quality Compendium of U.S. Health Systems, and American Hospital Association (AHA) Hospital Survey);
4) 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);
5) 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
6) 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, Medicare Provider Enrollment, Chain, and Ownership System, and Bundled Payments for Care Improvement Salesforce Database).

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

Exhibit E-1: Secondary data sources

Data source	Date range	Dataset contents	Use
AHA Hospital Survey	2014	Annual survey of acute care hospitals that collects information on hospital organizational structure, system affiliation, facility/service lines, inpatient/outpatient utilization, finances/expenses, physician arrangements, staffing, and corporate/purchasing affiliations.	Used to characterize CJR and control group hospitals. Information on PAC ownership was used to inform site visit and telephone interview sampling.
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 characterize CJR and control group markets.
AHRQ Compendium of U.S. Health Systems	2016	Identifies health systems for acute care hospitals in the United States.	Used to characterize CJR and control group hospitals. Information on health system ownership was used to inform site visit and telephone interview sampling.
Bundled Payments for Care Improvement Salesforce 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. Used to create a measure of Bundled Payments for Care Improvement dose for the volume analysis.
CJR programmatic data	Intervention	List of CJR participant hospitals, as well as their PY1, PY2 and PY3 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.

Data source	Date range	Dataset contents	Use
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 identify and characterize acute care IPPS hospitals located in CJR and control group markets. Used to identify section 401 hospitals located in control group hospitals to exclude from the mandatory analysis.
FY IRF PPS Final Rule data files	FY 2016 (Data is from FY 2014)	CMS IRF PPS data are used to set payment rates. Data files identify IRF facilities (by Medicare provider identification number), their geographic location, and annual number of IRF discharges.	Used to identify PPS IRF facilities in CJR and control group markets and produce market level IRF variables (IRF present in MSA; number of IRF discharges per 10,000 population).
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).
Medicare FFS beneficiary enrollment data	Q4 2007 to Q4 2018	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). 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	Q4 2007 to Q4 2018	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 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). Claims data were also used to generate the number of LEJR discharges for the volume analysis. Claims data were used to identify OP TKA procedures in CJR and control markets for descriptive analyses and site visits and telephone interview sampling. Claims data were used to generate LEJR volume, LEJR share, average HCC score, and post-discharge setting information for telephone interview sampling.

Data source	Date range	Dataset contents	Use
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.
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.

Data source	Date range	Dataset contents	Use
Medicare PECOS	December 2014 (end of baseline period)	Information on Medicare providers (hospitals and PAC providers) including ownership and chain relationships. For this evaluation, a hospital is considered to have "chain ownership" when it bills under the same TIN as another hospital or it has at least one TIN in the PECOS ownership table that is different than the hospital's TIN.	Used to create an indicator of hospital chain ownership for characterizing CJR and control group hospitals.
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 and located in CJR and control group markets.
US Census Bureau's American Community Survey	2014 5-year estimates	Annual survey from the US Census Bureau that provides sociodemographic (population size, age, sex, race/ ethnicity) and socioeconomic (median household income) population estimates at the MSA level.	Used to characterize CJR and control group markets.

Note: ACO = Accountable Care Organization, AHA = American Hospital Association, AHRF = Area Health Resource File, CEC = comprehensive ESRD care model, CME = Common Medicare enrollment, 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, LTCH = long-term care hospital, 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, NP = nurse practitioner, NPRA = net payment reconciliation amount, OASIS = Outcome and Assessment Information Set, PA = physician assistant, PAC = post-acute care, PECOS = Provider, Enrollment, Chain, and Ownership System, POS = provider of services, PPS = prospective payment system, PY = performance year, Q=quarter, SNF = skilled nursing facility, TAP = monthly Medicare claims file, TIN = tax identification number, UCP = uncompensated care payment.

B. Primary data sources

We collected and analyzed primary data from site visits and telephone interviews to inform questions that are not readily answered by secondary data. In this appendix we describe the qualitative methods employed during the third performance year.[1,2] We conducted site visits to twenty-one hospitals in three Metropolitan Statistical Areas (MSAs) and conducted two rounds of telephone interviews with 79 providers.

[1] https://innovation.cms.gov/Files/reports/CJR-firstannrptapp.pdf
[2] https://innovation.cms.gov/Files/reports/cjr-secondannrpt.pdf

1. Site visits

We conducted site visits to hospitals and associated providers in three mandatory CJR MSAs. Our aim was to capture the MSA level effects of the CJR model on care coordination strategies, relationships with associated providers, and patient and caregiver experience from various perspectives including CJR and non-CJR hospitals, orthopedic surgical practices, post-acute care providers, and outpatient physical therapy (PT) providers. Non-participant hospitals were included to provide insight into spillover effects and changes in volume due to the CJR model.

a. Interviewees

We spent four to ten days in each MSA conducting interviews with multiple hospitals. In each MSA, we conducted in-person interviews with representatives from two to eight CJR participant hospitals, as well as one to three hospitals that were not part of CJR or the control group, and were located within the CJR hospital referral region (HRR).[3] We also conducted in-person interviews with orthopedic surgeons and post-acute care (PAC) providers that received LEJR patients from the hospitals we visited (Exhibit E-2 lists targeted interviewees).

Exhibit E-2: Target interviewees for site visits

Organization	Interview session	Sample job titles
CJR hospital	Executive and financial leadership	CEO, COO, CFO
	Orthopedic service line and care redesign leadership	Orthopedic surgery or surgical service line leader, head of surgery department, head of operating room, VP for QI, nurse in charge of QI or CJR initiatives
	Patient care, discharge planning and PAC partnerships	CJR program coordinator, care coordinator/nurse navigator, discharge planner, VP of QI, RN, PT/OT, hospitalist (if involved in LEJR patient care)
	Data management	Analyst working with cost and quality information, individual in charge of PRO submission, individual that works with IT or data analysis vendors
Orthopedic surgeons and surgical groups	Surgeons	Orthopedic surgeons, physician assistants
Post-acute care providers	Clinical and financial leadership	Administrator, executive director, director of nursing, chief nursing officer
	Physical therapy	Director of rehabilitation, physical therapists
Non-CJR hospital	Executive and financial leadership	CEO, COO, CFO
	Orthopedic service line and care redesign leadership	Orthopedic surgery or surgical service line leader, head of surgery department, head of operating room, VP for QI, nurse in charge of QI or CJR initiatives

Note: CEO = chief executive officer, CFO = chief financial officer, COO = chief operating officer, HRR = hospital referral region, IT = information technology, LEJR = lower extremity joint replacement, PAC = post-acute care, PRO = patient reported outcomes, QI = quality improvement, RN = registered nurse, VP= vice president.

[3] HRRs were developed by the Dartmouth Atlas of Health Care to define regional health care markets based on Medicare patient referral patterns for tertiary care. An HRR contains at least one hospital that performs major cardiovascular procedures and neurosurgery.

b. Protocols

To capture the variability in responses to the CJR model, the list of interview topics was broad and diverse. We developed semi-structured interview guides tailored to the expertise of each type of interviewee. Flexibility was encouraged such that if an orthopedic surgeon was also responsible for the "data management" interview topics, for example, questions intended for the data analyst would be asked of the surgeon. Given the diversity of hospitals, the number of interviews and interviewees varied by hospital.

c. Hospital selection criteria

Performance year (PY) 3 site visits focused on the 34 mandatory MSAs. We created a map identifying three groups of hospitals: 1) Mandatory or opt-in hospitals, 2) low-volume or rural hospitals that did not opt to continue participation in the model, and 3) non-participant hospitals in the HRR that performed LEJR. We selected three MSAs after evaluating the following:

- Included at least one CJR participant hospital and other non-participating hospitals in the area that performed LEJR,
- Geographic variation in the distribution of site visit candidates,
- Variation in the degree of competition within the MSA (e.g., number of hospitals and LEJR volume),
- Site visit candidates' past Bundled Payments for Care Improvement LEJR participation,
- Variation in PAC supply and use,
- Variation in outpatient total knee arthroplasty (TKA) rate in the market,
- We excluded MSAs that were visited in prior years.

d. Hospital recruitment

To recruit sampled hospitals for participation, we emailed information to the CJR point-of-contact (POC) for 18 hospitals, inviting them to participate in a brief introductory call with our team. We also reached out to 10 non-CJR hospitals thorough contacts received from CJR participant hospitals in the same system, or through hospital websites or information desks. Of these hospitals, 15 (83%) CJR hospitals and 7 (70%) non-CJR hospitals participated in an introductory call during which we described the purpose of the visit, the content that would be covered, and provided hospital representative(s) with an opportunity to ask questions. One hospital ultimately was unable to make the scheduled appointment. Exhibit E-3 shows that hospital response rates improved in year 3.

Exhibit E-3: Years 1 through 3 CJR hospital site visit response rates

	Contacted in year 1 (n=30)	Contacted in year 2 (n=38)	Contacted in year 3 (n=17)[a]	Total (n=85)
Agreed to participate	9 (30%)	11 (29%)	15 (88%)	35 (41%)
Declined to participate	13 (43%)	13 (34%)	0 (0%)	26 (31%)
Did not respond	7 (23%)	9 (24%)	2 (12%)	18 (21%)
System participating (removed)	1 (3%)	5 (13%)	0 (0%)	6 (7%)

[a] In year 3, 15 hospitals agreed to participate in site visits and one hospital later withdrew.

Exhibit E-4 summarizes year 3 site visit interviews conducted across the three MSAs.

Exhibit E-4: Site visit interviews by MSA, year 3

MSA	CJR hospitals visited	CJR hospital interviews	Non-CJR hospitals visited	Non-CJR hospital interviews	Orthopedic surgeon and surgical practice interviews	Post-acute care provider interviews
1	4	15	1	2	2	5
2	8	32	3	6	8	9
3	2	8	3	6	4	4
Total	14	55	7	14	14	18

Note: MSA = metropolitan statistical area.

Exhibit E-5 presents characteristics of CJR participant hospitals that participated in year 3 site visits compared to all CJR participant hospitals.

Exhibit E-5: Characteristics of CJR hospitals that participated in year 3 site visits versus all mandatory CJR hospitals

	Year 3 site visit (n=14)	All mandatory CJR hospitals[a,b] (n=403)
Patient HCC score, PY1 mean	1.41	1.67
Health system membership, % yes	79%	87%
Annual total Medicare discharges, baseline mean	4,091	4,350
Proportion of Medicare discharges for LEJR, baseline mean	15%	6%
Percent OP TKA in 2018	15%	15%
Number of beds, mean	268	286
Teaching facility, % yes	50%	44%
Own a PAC provider, % yes	65%	67%

Source: CJR evaluation team analysis of December 2016 POS, FY 2016 CMS Annual IPPS, 2014 AHA Hospital Survey, 2016 AHRQ Compendium of U.S. Health Systems, and Medicare claims and enrollment data for LEJR discharges and episodes in the baseline period (2012-2014) and PY1 (episodes starting on or after April 2016 and ending by December 2016).

Notes: AHA = American Hospital Association, AHRQ = Agency for Healthcare Research and Quality, HCC = hierarchical condition category, IPPS = Inpatient Prospective Payment System, LEJR = lower extremity joint replacement, OP = outpatient, PAC = post-acute care, POS = provider of services, PY = performance year, TKA = total knee arthroplasty.

[a] Two hospitals have missing data on first PAC setting, LEJR share of baseline, annual total discharges and HCC score, 81 hospitals have missing data for PAC ownership and 12 hospitals have missing data for percent outpatient TKA.

[b] "All mandatory CJR" hospitals are defined as mandatory CJR hospitals with any LEJR episodes in PY3 at the time of sampling.

2. Telephone interviews

In PY3 we conducted two rounds of telephone interviews with representatives from hospitals selected from the 34 mandatory CJR MSAs. The telephone interviews were used to collect information from CJR participant hospitals on care coordination efforts and the perceived impact of these efforts, and hospital experiences with removal of elective TKA from the Medicare inpatient only list.

a. Interviewees

For Round 5, the team aimed to interview hospital staff most knowledgeable about care coordination efforts. Most often, interviews were conducted with care coordination staff (e.g., care coordinators, care navigators, social workers, nurses), though leadership or administrative staff (e.g., Chief of Nursing, Director of Case Management) participated as well. For Round 6, the team requested that individuals most familiar with hospital response to the rule change removing elective TKA from the Medicare inpatient only list and knowledgeable about TKA surgical decision making participate in the interviews. Most often, interviews included members of executive leadership (e.g., Chief Executive Officer, Chief Financial Officer), care management staff (e.g., Orthopedic Service Line Director, Chief Nursing Officer), care coordination staff (e.g., Care Navigator, Care Coordinator, Manager of Care Coordination), or payment program and utilization management representatives (e.g., Director of Value-Based Programs, Payment Innovation Project Manager, Director of Utilization Management).

b. Protocols

We developed semi-structured interview guides tailored to the topic of each round of telephone interviews. The fifth round of telephone interviews included a pre-interview survey. The pre-interview survey was completed by 31 of 34 hospitals (91.2%). The survey responses identified the most important care coordination activities for each hospital, and were used to tailor the interview protocol to address each hospital's approach to care coordination under the CJR model. For hospitals that did not complete the pre-interview survey, interviewers asked hospital representatives to expand on the care coordination activities considered most important to successful patient outcomes.

c. Hospital selection criteria

The fifth round of telephone interviews sampled CJR participant hospitals that were mandatory throughout the intervention period and opt-in hospitals in mandatory MSAs in the 34 mandatory CJR MSAs. The sixth round of telephone interviews sampled mandatory hospitals only. To reduce participant burden, hospitals were excluded from the round 5 or round 6 sample if they were also selected for year 3 site visits.

For the fifth round of telephone interviews, there were 383 CJR participant hospitals located in mandatory MSAs in the universe. Hospitals were excluded from the eligible sample if they were part of a large national health system that declined to participate or had fewer than 20 annual episodes in 2017. Of the remaining 258 eligible hospitals, 100 were randomly sampled.

For the sixth round of telephone interviews, there were 403 CJR participant hospitals located in mandatory MSAs in the universe.[4] Hospitals were excluded from the eligible sample because they either did not perform any outpatient musculoskeletal procedures in calendar year (CY) 2017,[5] performed fewer than 20 TKA procedures in CY 2017 or CY 2018,[6] or were a part of a large national health system that declined to participate. From the final group of 252 eligible hospitals, we selected all hospitals that performed 50% or more of their TKAs in the outpatient setting in 2018[7] (n=31) and randomly selected 100 hospitals from hospitals performing less than 50% of their TKAs in the outpatient setting.

d. Hospital recruitment

We worked with the CJR POC at each hospital to obtain contact information for individuals who were knowledgeable about the topic of interest. We first contacted potential interviewees over email and included key information and related materials (i.e., frequently asked questions document and informed consent information). We encountered some challenges in recruiting participants, including difficulty obtaining current contact information for sampled hospitals, a low response rate to our initial outreach requiring subsequent outreach efforts, and hospital representatives having limited awareness of or time to participate in evaluation activities.

In the fifth round of telephone interviews, we successfully interviewed 34 (34%) of the 100 sampled hospitals. Three hospitals (3%) declined to participate and 46 hospitals (46%) did not respond to our request. In the sixth round of telephone interviews, we successfully interviewed 45 (34%) of the 131 sampled hospitals. Three hospitals (2%) declined to participate and 83 (63%) did not respond to our request. Exhibit E-6 presents participation rates for telephone interviews conducted in year 3.

Exhibit E-6: Telephone interview participation rates, rounds 5 and 6

	R5 TI	R6 TI
Declined Participation	3 (3%)	3 (2%)
Responded but could not schedule	1 (1%)	13 (10%)
Responded after the interview window	1 (1%)	NA
Did Not Respond	61 (61%)	70 (53%)
Interviewed	34 (34%)	45 (34%)

Note: NA = not appliable, R5 = round 5, R6 = round 6, TI = telephone interview.

[4] The round 6 universe was greater than the round 5 universe due to the addition of twenty hospitals to the CJR model when the Bundled Payments for Care Improvement initiative ended on 9/30/2018.

[5] We excluded hospitals that did not perform any outpatient procedures in CY 2017 because these hospitals were likely not performing outpatient TKA in CY 2018.

[6] We excluded hospitals that performed fewer than 20 TKA in CY 2017 or CY 2018 because these hospitals likely do not have the familiarity to speak to the interview topic.

[7] We calculated the percent of a hospital's TKA performed in the hospital outpatient setting using Medicare claims data from January – September 2018.

Exhibit E-7 presents characteristics of CJR participant hospitals that participated in telephone interviews compared to all CJR participant hospitals.

Exhibit E-7: Characteristics of CJR hospitals that participated in telephone interviews versus all CJR hospitals

		R5 TI (n=34)[a]	R5 TI mandatory universe (n=383)[b]	R6 TI (n=45)[c]	R6 TI mandatory universe (n=403)[d]
Number of MSAs represented		17	34	16	34
Patient HCC score, Intervention mean		1.55	1.67	1.48	1.67
Not for profit, % yes		59%	60%	76%	61%
Health system membership, % yes		85%	87%	89%	87%
Annual total Medicare discharges, baseline mean		4,157	4,242	5,293	4,350
Proportion of Medicare discharges for LEJR, baseline mean		9%	6%	9%	6%
Number of beds, mean		241	281	299	286
Teaching facility, % yes		26%	44%	49%	44%
Own a PAC provider, % yes		65%	54%	75%	67%
Percent OP TKA in 2018		20%	15%	29%	15%
Baseline %	First discharged to IRF	15%	15%	15%	14%
	First discharged to SNF	44%	46%	39%	46%
	First discharged home with HH	28%	27%	36%	27%
	First discharged home without HH	13%	13%	10%	12%
Intervention %	First discharged to IRF	7%	8%	6%	8%
	First discharged to SNF	31%	37%	28%	37%
	First discharged home with HH	39%	37%	47%	38%
	First discharged home without HH	23%	17%	19%	17%

Source: CJR evaluation team analysis of December 2016 POS, FY 2016 CMS Annual IPPS, 2014 AHA Hospital Survey, 2016 AHRQ Compendium of U.S. Health Systems, and Medicare claims and enrollment data for LEJR discharges and episodes in the baseline period (2012-2014) and PY1 (episodes starting on or after April 2016 and ending by December 2016).

Notes: AHA = American Hospital Association, AHRQ = Agency for Healthcare Research and Quality, FY = fiscal year, HCC = hierarchical condition category, HH = home health, IPPS = Inpatient Prospective Payment System, IRF = inpatient rehabilitation factiliy, LEJR = lower extremity joint replacement, MSA = metropolitan statistical area, OP = outpatient, PAC = post-acute care, POS = provider of services, PY = performance year, R5 = round 5, R6 = round 6, SNF = skilled nursing facility, TI = telephone interview, TKA = total knee arthroplasty.

[a] Three Round 5 hospitals had missing data for PAC ownership.

[b] One Round 5 hospital had missing data on first PAC during intervention and HCC score; eleven hospitals had missing data on the percent OP TKA; 73 hospitals had missing PAC ownership data.

[c] Five Round 6 hospitals had missing data for PAC ownership.

[d] Two Round 6 hospitals had missing data on first PAC setting, LEJR share of baseline, annual total discharges, and mean HCC; twelve hospitals had missing data on the percent OP TKA; 81 hospitals had missing PAC ownership data.

II. Study Population

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

A. Defining the CJR and control group populations

The Centers for Medicare & Medicaid Services (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 three performance years of the model from April 1st 2016 to December 31, 2018. Our analysis primarily focused on episodes from hospitals that were mandated to participate in PY3 (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). We also present an intention-to-treat (ITT) analysis for episodes and hospitals included in the original design of the model (Section II.C.1).

Exhibit E-8 shows the names and core-based statistical area (CBSA) identification numbers of the CJR and control group MSAs included in the mandatory and ITT analyses. The MSAs included in the mandatory analysis are starred. Section II.C provides additional detail about how the control group MSAs were identified and the weights generated for mandatory and ITT analyses.

Exhibit E-8: CJR and control group MSAs included in the ITT and mandatory 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*

CJR		Control	
CBSA ID	MSA name, state	CBSA ID	MSA name, state
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
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*

CJR		Control	
CBSA ID	**MSA name, state**	**CBSA ID**	**MSA name, state**
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
		40340	Rochester, MN
		40380	Rochester, NY
		40900	Sacramento--Roseville--Arden-Arcade, CA
		41500	Salinas, CA
		41620	Salt Lake City, UT*

CJR		Control	
CBSA ID	MSA name, state	CBSA ID	MSA name, state
		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.
CBSA = core-based statistical area, ITT = intention-to-treat, MSA = metropolitan statistical area.

B. Hospital characteristics

We analyzed the characteristics of mandatory hospitals and we compared their characteristics to all other IPPS hospitals in the United States. Further, we compared characteristics of opt-in and non-opt-in hospitals in voluntary MSAs. We identified CJR participant hospitals using a participant list provided by CMS (October 2019). We identified all other acute care hospitals in the United States that were paid under the Medicare IPPS using the CMS Provider of Services file. Hospitals that did not perform any LEJR during the baseline period (2012-2014) were excluded.

Data were compiled on hospital, patient, and market characteristics from a variety of secondary sources, including: Provider of Services file (ownership, Census region, teaching status), FY IPPS Final Rule data files (bed count, disproportionate share hospital (DSH) patient percentage), Medicare FFS claims (annual LEJR count, average Hierarchical Condition Category (HCC) score for LEJR patients, episode payment, first PAC discharge setting, LEJR as a percent of total discharges, annual total Medicare discharges, hospital LEJR market share, Herfindahl-Hirschman Index (HHI)), Bundled Payment for Care Improvement initiative participant list, CJR model net

payment reconciliation amount (NPRA) and quality data, Area Health Resource File (skilled nursing facility (SNF) beds per 10,000 population), and American Community Survey (MSA population count).

We calculated summary statistics (mean, range, median, and (interquartile range) IQR) to describe mandatory hospitals. Comparisons were made across groups (mandatory vs. all other IPPS hospitals; opt in vs. non-opt in) using chi-square tests for categorical variables and t-tests for continuous variables.

C. Creation of the Analytic Weights

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 E-9 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 E-9: CMS' stratified random sample of CJR MSAs and analytic weights for ITT analysis

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	CJR weight	# Control group MSAs	Proportion of MSAs in the control group	Control group weight
Less than median population	1	Lowest quartile	25	8	32.0%	1.0	17	68.0%	8/17
	2	2nd lowest quartile	18	6	33.3%	1.0	12	66.7%	6/12
	3	3rd lowest quartile	19	8	42.1%	1.0	11	57.9%	8/11
	4	Highest quartile	22	11	50.0%	1.0	11	50.0%	11/11
More than median population	5	Lowest quartile	15	5	33.3%	1.0	10	66.7%	5/10
	6	2nd lowest quartile	28	10	35.7%	1.0	18	64.3%	10/18
	7	3rd lowest quartile	22	9	40.9%	1.0	13	59.1%	9/13
	8	Highest quartile	22	10	45.5%	1.0	12	54.5%	10/12
		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: ITT = intention-to-treat, MSA = metropolitan statistical area.

1. ITT weights

We used an ITT analysis to evaluate the impact of the CJR model on the *ever CJR participating hospitals*. The ITT includes all CJR participant hospitals located in the original 67 MSAs that were randomly selected into the CJR model from the onset of the model. To account for the differential probability of selection and minimize differences between the CJR and control groups, we created weights that were used in the descriptive and risk-adjusted ITT analyses. The control group was weighted to represent the CJR group; CJR MSAs all had a weight of 1, while the control group weights were calculated as the number of CJR MSAs in the stratum divided by the number of control group MSAs in the stratum (Exhibit E-10).[8]

Specifically, for each MSA sampling stratum

- *Weight for every CJR participating hospitals = 1*
- *Weight for control group hospitals =*

$$= \frac{(probability\ in\ treatment)}{(probability\ in\ control)} = \frac{(probability\ in\ original\ 67\ CJR\ MSAs)}{(probability\ in\ 104\ non\ CJR\ MSAs)}$$

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

2. Mandatory weights

We used an average treatment effects of the treated (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 excluded low-volume and rural hospitals from the 104 non-CJR MSAs because these hospitals were excluded from mandatory participation.

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

[8] In the first annual report, the results were weighted to represent the entire sample of CJR-eligible MSAs (control group MSAs in addition to CJR MSAs). See https://innovation.cms.gov/Files/reports/CJR-firstannrptapp.pdf

the 34 mandatory CJR MSAs were not selected randomly and so we could not rely on simple proportions.[9] We therefore calculated exact probabilities using combinatorics. We used the exact 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 10th of a percent).

Exhibit E-10 shows the analytic weights calculated for control group MSAs included in the mandatory analysis.

[9] 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).

Exhibit E-10: Analytic weights for control group MSAs included in the mandatory analysis

MSA sampling stratum	MSA	Weight
4	All MSAs	0.83
8	All MSAs	1.00
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	0.77
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.

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; or
- 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.[10]
- the episodes were attributed to the Bundled Payments for Care Improvement initiative.[11]

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).

---

[10] 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

[11] 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.

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. Therefore, we also provide impact estimates based on episodes for all LEJRs, including OP TKAs, to account for the impact of the CJR model on the inpatient or outpatient decision. For the OP 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"). 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.

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 E-11 and E-12 list each claims-based and assessment-based measure respectively.

Exhibit E-11: 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	Acute inpatient care (anchor hospitalization) length of stay (in days)
	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 HH 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[e]
	Emergency department visit, 90-day PDP
	All-cause mortality, inpatient stay and 90-day PDP[f]
	Manipulation under anesthesia (MUA), 90-day PDP[g]
	Incidence of any complications, 90-day PDP[e,g]

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

Notes: HHA = home health agency, IRF = inpatient rehabilitation facility, MUA = manipulation under anesthesia, 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] Updated specification documents were released by CMS in March 2019 for the unplanned readmission measure and complications measure, and these measures were revised accordingly. Available at: https://www.qualitynet.org/inpatient/measures/readmission/methodology and https://www.qualitynet.org/inpatient/measures/complication/methodology

[f] 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.

[g] MUA and complications are measured among elective episodes only; further, MUA is among patients with knee replacements.

Exhibit E-12: Functional status and pain

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 Outcome and Assessment Information Set (OASIS) data.

Note: HHA = home health agency, IRF = inpatient rehabilitation facility, MDS = minimum data 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 E-13 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 E-13: 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, LTCH = long-term care hospital, 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 sample

1. PY3 mandatory CJR hospitals compared with control group hospitals

We compared baseline characteristics of the 378 PY3 mandatory CJR participant hospitals to the 377 control group hospitals with any LEJR episodes during the baseline period. Baseline hospital characteristics were balanced across CJR and control group hospitals on nearly all characteristics (Exhibits E-14a and 14b). However, compared to control group hospitals, CJR participant hospitals were more likely to be safety-net hospitals (34.2% vs. 18.3%, $p<0.01$) and have higher average disproportionate share hospital percentages (25.5% vs. 21.2%, $p<0.05$). We did not observe any differences between CJR and control group hospitals in the averages of the baseline characteristics examined (volume, episode payment, first PAC discharge setting, bed count, Medicare days percent, Medical residents per 1,000 beds, and disproportionate share percent) (Exhibit E-14b). For a comparison between CJR and control group hospitals included in the intent-to-treat analysis, see the CJR model Second Annual Report (available at https://innovation.cms.gov/Files/reports/cjr-secondannrpt.pdf).

Exhibit E-14a & E-14b: Characteristics of mandatory CJR hospitals compared with control group hospitals, among hospitals with any LEJR during baseline (2012 – 2014)

Baseline characteristic		CJR hospitals (N=378) %	Control group hospitals (N=377) %	p-value
Ownership	Non-profit	67.4	65.4	p=0.88
	For-profit	20.0	19.3	
	Government	12.7	15.3	
Census region	Northeast	25.9	13.9	p=0.45
	South	55.4	54.6	
	Midwest	6.8	31.2	
	West	11.8	0.2	
Part of chain	Yes	82.4	82.8	p=0.96
Teaching hospital	Yes	50.1	40.0	p=0.13
Prior Bundled Payments for Care Improvement experience	Ever participated in the Bundled Payments for Care Improvement initiative (LEJR or non-LEJR)	22.4	18.0	p=0.51
Safety-net	Safety-net hospital	34.2	18.3	p<0.01

Baseline characteristic	CJR hospitals (N=378) Mean	Control group hospitals (N=377) Mean	p-value
Annual number of LEJR episodes[a]	305	337	p=0.49
Standardized total episode allowed payment, inpatient stay plus 90 day post-discharge period[a]	$29,257	$27,978	p=0.12
First PAC IRF[a]	13.5%	12.8%	p=0.82
First PAC SNF[a]	41.3%	38.5%	p=0.59
First PAC home with HH[a]	36.7%	36.4%	p=0.94
First PAC home without HH[a]	8.5%	12.3%	p=0.13
Bed count	388.8	341.6	p=0.35
Medicare days percent	37.5%	39.8%	p=0.31
Medical residents per 1,000 beds	90.5	74.0	p=0.56
Disproportionate share percent	25.5%	21.2%	p<0.05

Source: CJR evaluation team analysis of December 2016 POS, December 2014 PECOS, FY 2016 CMS Annual IPPS, 2014 AHA Hospital Survey, and Medicare claims and enrollment data for episodes initiated in 2012 through 2014 that ended between April 2012 and March 2015.

Notes: PY3 mandatory CJR hospitals are defined as all hospitals located in CJR-participating MSAs and required to participate in the CJR model as of December 2018. The control group is constructed by applying MSA-level weights to the 104 original (ITT) control group MSAs. The weights are based on the probability that the MSA is selected into the 34 mandatory CJR MSAs through the two-step selection process (see Section II.C.2).

This exhibit includes CJR and control groups hospitals that had at least one LEJR episode during the baseline period and were paid under the inpatient prospective payment system. The CJR and control groups are weighted by number of episodes and the control group is further weighted by the MSA sampling strata (probability of selection) to be representative of the CJR group.

Tests of significance for categorical variables use a design-based F-test, while tests of means use a t-test. The standard errors are adjusted for clustering of hospitals within MSAs. Differences that are significant at the 99%, 95%, or 90% significance level are indicated by red,orange, or yellow shaded cells, respectively.

AHA = American Hospital Association, FY = fiscal year, HH = home health, IPPS = Inpatient Prospective Payment System, IRF = inpatient rehabilitation facility, ITT = intention-to-treat, LEJR = lower extremity joint replacement, MSA = metropolitan statistical area, PAC = post-acute care, PECOS = Provider Enrollment and Chain/Ownership System, POS = provider of services, PY = performance year, SNF = skilled nursing facility.

[a] The baseline outcomes are not risk adjusted. Reporting standardized allowed payments.

D. Analytic methodology

While the CJR and control group populations are overall quite similar in terms of market, hospital, and patient characteristics, 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 difference-in-differences (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.[12]

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, second, and third performance years in the Final Rule: episodes starting on or after April 1, 2016 and ending by December 31, 2018.[13]

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, market, 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*

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.

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

[13] 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.

- 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 providers 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 three performance years during the intervention period, Equation 2 was modified to include three time period indicators t_1 (equals 1 during PY1 intervention period and zero otherwise) and t_2 (equals 1 during PY2 intervention period and zero otherwise), and t_3 (equals 1 during PY3 intervention period and zero otherwise).

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

- Coefficient b_5 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_6 and b_7 represent the DiD estimators for PY2 and PY3 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.[14] For all DiD models, statistical significance was assessed at the 10% level.

In addition, we ran DiD models for all LEJRs, which included inpatient and outpatient episodes, to account for the greater share of outpatient TKAs in the control group. These results are included in Appendix B.

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

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

can be tested if sufficient baseline data on the CJR and control groups are available, the second assumption is untestable.

We 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. 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 next section). We used a joint test of equality to conclude whether there is evidence to reject the parallel trend assumption. A statistically significant joint test result suggests that the CJR and control group trends during the baseline period were not the same.

The 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, market, and patient characteristics in the baseline period

and the test is:

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

For most outcomes, joint tests of the equality of the coefficients on the CJR-year interaction terms were not statistically significant, indicating a lack of evidence to reject the parallel trends assumption. For analysis of mandatory CJR hospitals, there is evidence to reject the parallel trends assumption for four measures: home health payments ($p<0.05$), 30-day post-episode payments ($p<0.05$), number of outpatient PT/OT visits ($p<0.10$), and pre-surgical PT/OT visits for elective procedures ($p<.10$).

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

a. Claims-based risk adjustments

In the DiD models that we estimated, we controlled for potential differences in beneficiary demographics, clinical characteristics observed before hospitalization, and provider characteristics

(represented by $X_{i,k,t}$ 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).[15] 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.[16] To control for prior health conditions, we used HCC indicators for the 12 months preceding the anchor hospitalization,[17] 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 addition, we included state dummies in all regression models to control for geographic differences in health care spending.

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 E-15). 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.

[15] Models were also estimated separately for fracture episodes and elective episodes in addition to risk adjusting for fracture in models that combined fracture and elective episodes.

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

[17] 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 E-15: Predictive risk factors used to risk-adjust claims-based outcomes

Domain	Variables
Characteristics of the procedure	▪ Anchor MS-DRG ▪ Hip fracture status[a] ▪ Procedure type (hip or knee)
Patient demographics and enrollment	▪ Age (under 65, 65-79, 80+) ▪ Sex ▪ Race ▪ 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 use variables used in risk adjustment varied by model[b] • Binary indicators for any acute care inpatient, emergency department visits, IRF, SNF, HHA, hospice, other Part A inpatient, LTCH, and custodial nursing facility service utilization in the six months preceding the start of the episode • Binary indicators for any acute care inpatient, IRF, SNF, HHA, hospice, and other Part A inpatient service use in the one month preceding the start of the episode • Number of days of acute care inpatient, IRF, SNF, HHA, hospice, other Part A inpatient, and LTCH service 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

Source: Risk adjustement 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, and Bundled Payments for Care Improvement initiative participant list.

Notes: ACO = Accountable Care Organization, 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] Models were also estimated separately for hip fracture episodes and elective episodes in addition to risk adjusting for fracture in models that combined fracture and elective episodes.

[b] 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, HH payment, Part B payment, unplanned readmissions, ED use, MUA, number of SNF days, Anchor LOS, first PAC, and any pre-surgical PT/OT. 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 HH visits, number of outpatient PT/OT visits, and number of HH PT/OT visits.

b. Assessment-based risk adjustment

We applied existing risk-adjustment models for the National Quality Forum (NQF) endorsed and CMS quality measures for the IRF (average change in mobility score),[18] SNF (improved transfer, locomotion on unit, and walking in corridor),[19,20] and HHA settings (improved ambulation/locomotion, improved bed transferring, and improvement in the frequency of pain when moving around).[21,22,23] We made some modifications to the risk-adjustment models for these measures to better align with the needs of the evaluation. 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 E-16). We considered alternative aggregation levels to control for prior service use (Exhibit E-16) and selected a specific subset of prior service use variables for each outcome that improved the model's goodness of fit. 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.

---

[18] 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

[19] 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

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

[21] 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

[22] 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

[23] 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

For the SNF measures, we included additional Minimum Data Set (MDS)-based risk-factors to the NQF-endorsed risk-adjustment models based 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.

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. The final model included covariates that had p-values less than 0.05 and significantly improved the model's goodness of fit.

Exhibit E-16: Predictive risk factors used to risk-adjust assessment-based outcomes

Domain	Variables
Characteristics of the procedure	▪ Anchor MS-DRG ▪ Hip fracture status[a] ▪ 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 use variables used in risk adjustment varied by model[b] • Binary indicators for any acute care inpatient, IRF, SNF, HHA, hospice, other Part A inpatient, LTCH, and custodial nursing facility service utilization in the six months preceding the start of the episode • Binary indicators for any acute care inpatient, IRF, SNF, HHA, hospice, and other Part A inpatient service use in the one month preceding the start of the episode • Number of days of acute care inpatient, IRF, SNF, HHA, hospice, other Part A inpatient, and LTCH service 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
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-based variable used in risk adjustment varied by model[b] ▪ 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, and Bundled Payments for Care Improvement initiative participant list.

Notes: ACO = Accountable Care Organization, HCC = hierarchical condition category, 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] Models were also estimated separately for hip fracture episodes and elective episodes in addition to risk adjusting for fracture in models that combined fracture and elective episodes.

[b] 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 E-17). 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), and any pre-surgical PT/OT. A Poisson model was used to estimate inpatient length of stay. 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 home health (HH) 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 E-17: 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, home health ▪ 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	▪ Unplanned readmission ▪ Emergency department visit ▪ MUA, among elective knee replacement episodes ▪ Complications, among elective episodes ▪ All-cause mortality ▪ Any pre-surgical PT/OT, among elective episodes ▪ 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
Poisson	▪ Inpatient length of stay

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.

4. Sensitivity analyses

As discussed in our second CJR evaluation report, a number of sensitivity analyses were performed on the findings for the claims-based outcomes in the main analysis:[24] 1) we assessed the relative impact of the stratum-level weights by excluding the weights from the DiD estimate and standard errors; 2) we excluded episodes generated under MSSP, Pioneer ACO, or Next Generation ACO to identify whether these exclusions would change the DiD estimate; and 3) we tested the sensitivity of the DiD estimate to including stratum fixed-effects. We found that the alternative specifications used in the sensitivity analyses did not materially affect any of the findings in the main analysis and thus provided evidence that the main analysis and the conclusions presented in the report were robust.

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. We explored the change in total episode payment with these episodes included, and we found that our findings were robust. Unadjusted baseline and intervention mean outcomes including these episodes were comparable to mean outcomes that excluded these episodes.

For this annual report, we 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 percent 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 period prior to joining BPCI and the time period after exiting BPCI);[25] 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.

[24] https://innovation.cms.gov/Files/reports/cjr-secondannrpt.pdf

[25] 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.

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 for all LEJRs (IP+OP) increased from -$1,378 to -$1,154, a $224 difference. The DiD for total payments for inpatient LEJR episodes increased from -$1,540 to -$1,306, a $234 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 for all LEJRs (IP+OP) increased from -$1,378 to -$1,172, a $206 difference. The DiD for total payments for inpatient LEJR episodes increased from -$1,540 to -$1,289, a $251 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.

Our final sensitivity analysis was including the number of days between the first and last PAC assessments as a factor in the risk adjustment models for the assessment-based outcomes. Fewer days between the first and last PAC assessments would provide less time for patients to show functional improvement. Since the CJR model may impact the number of days between the first and last PAC assessment we do not include it as a causal factor in our main model, but only as a sensitivity (e.g., CJR participant hospitals may encourage SNFs to discharge CJR patients earlier, reducing the number of SNF days and the number of days between the first and last MDS assessment). Findings from this sensitivity analysis are presented in Section II.B.5.c of the annual report.

IV. OP TKA descriptive methods

CMS removed TKA from the inpatient only list, effective January 2018, and Medicare now covers TKAs performed in the hospital outpatient (OP) setting.[26] We monitored the occurrence of OP TKA in CJR and control group hospitals and MSAs from January through October 3, 2018.[27]

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 at the hospital and MSA levels. For each CJR and control group hospital, we calculated the percent of TKAs performed on outpatients 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 each MSA, we calculated the percent of TKAs performed on outpatients by dividing the number of outpatient TKAs meeting episode eligibility by the sum of all TKAs meeting episode eligibility at CJR or control group hospitals in the MSA.

[26] 2018 OPPS final rule (https://s3.amazonaws.com/public-inspection.federalregister.gov/2017-23932.pdf).

[27] 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, 2018. 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 is based on estimates from a DiD model of per-episode standardized paid amounts. The DiD estimates are 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. This method produces a per-episode estimate of the change in non-standardized paid amounts. The total change in non-standardized paid amounts is produced by multiplying the per-episode estimate by the total number of episodes.

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 was calculated by dividing reconciliation payments by the total number of CJR episodes.

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 all CJR patients and a random sample of control patients who had LEJR surgery during the third performance year. Each wave covered two months of LEJR episodes (episodes that began March and April 2018 and August and September 2018). Exhibit E-18 describes the patient survey waves.

Exhibit E-18: Patient survey sample by survey wave

Wave	Discharge date	CJR LEJR episodes	CJR patients sampled	Control LEJR episodes	Control patients sampled
3	March & April 2018	6,389	6,389	7,386	6,696
4	August & September 2018	6,089	6,089	7,323	6,441

Source: CJR evaluation team analysis of survey data for patients with discharge from LEJR surgery in March, April, August, or September 2018.

Note: LEJR = lower extremity joint replacement.

In both waves, we selected all available CJR patients. We selected an approximately equal number of control patients through a proportional random sample based on each control hospital's LEJR volume during the sampling period.[28]

1. Survey administration

We mailed surveys to patients between 60 and 120 days after their LEJR discharge (97 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. Sensitivity analyses did not find any evidence that average time between discharge and survey response differed between the CJR and control groups, 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

Across the sample pooled across waves 3 and 4, the response rate was 67.6% for CJR patients and 68.6% for control patients – a small and insignificant difference (Exhibit E-19). The CJR analytic sample consisted of 8,433 completed survey responses, or 67.6% of all CJR episodes during the sampling period. The responses included patients from 327 of the 351 mandatory CJR participant

[28] We initially drew an equal number of control episodes. After the sample was pulled, CMS identified 7 CJR hospitals that were retroactively identified as "rural" and excluded from mandatory participation; these CJR patients were dropped from the sample.

hospitals with episodes during the four months covered by the two survey waves. The analogous control analytic sample consisted of 9,014 completed survey responses, or 61.3% of all control episodes, and included patients from 364 of 377 control hospitals.

Exhibit E-19: Overall sample size and response rate, waves 3 and 4 combined

Group	Patients surveyed (starting sample)		Survey responses received (analytic sample)		Response rate	
	CJR	Control	CJR	Control	CJR	Control
All LEJRs	12,478	13,137	8,433	9,014	67.6%	68.6%

Source: CJR evaluation team analysis of survey data for patients with discharge from LEJR surgery in March, April, August, or September 2018.

Note: LEJR = lower extremity joint replacement.

Response rates across waves 3 and 4 varied based on patients' first discharge setting and decreased with intensity of the PAC setting (Exhibit E-20). Among patients who went straight home with no PAC, the response rate was 74.2% for CJR patients and 73.8% for control patients. Response rates for patients who went home but received home health care were nearly identical to those who went home without home health care. Among patients discharged from the hospital to a SNF, 54.7% of CJR patients and 59.7% of control patients responded to the survey, a significant difference ($p<0.05$). Among beneficiaries discharged from the hospital to an IRF, 54.1% of CJR patients and 59.5% of control patients responded to the survey, which was also a significant difference ($p<0.10$).

Exhibit E-20: Sample size and response rate by discharge setting, waves 3 and 4 combined

Group	Patients surveyed (starting sample)[a]		Survey responses received (analytic sample)		Response rate	
	CJR	Control	CJR	Control	CJR	Control
Inpatient rehabilitation facility	736	1,042	398	620	54.1%	59.5%*
Skilled nursing facility	3,440	3,760	1,881	2,243	54.7%	59.7%**
Home health care	6,193	5,760	4,584	4,256	74.0%	73.9%
Home without home health	2,001	2,511	1,485	1,852	74.2%	73.8%

Source: CJR evaluation team analysis of survey data for patients with discharge from LEJR surgery in March, April, August, or September 2018.

Notes: Significance of difference in response rate determined by t-test: *p<0.10 ** p < 0.05.

[a] The starting sample and analytic sample do not sum to the same values as Exhibit E-19 because data on post-acute care discharge were missing for about 1% of observations.

Among all LEJR episodes that occurred during our sampling period, the largest proportion resulted in the patient discharged home with HH care, while the smallest proportion resulted in discharge to an IRF. This pattern holds in our analytic sample, which means our ability to identify statistically significant differences will be greater among respondents discharged home with HH care than among respondents discharged to an IRF.

B. Analytic methodology

This section describes the general analytic approach for all patient survey analyses included in the third annual report. Unless otherwise noted, the methods described below were applied identically across all 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 E-21. Appendix G includes the patient survey questionnaire.

Exhibit E-21: 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	Healthcare 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: [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 at the end of the episode, 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 then calculated the mean change for CJR and control groups and 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 I needed, about right, or not enough). The third measure, did you have all the medical equipment you needed when you went 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, summarizes responses to three questions. Satisfaction with care management summarizes 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" summarizes three survey questions that each have 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 0 again indicates 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 wave 1 (which comprised approximately 85% of the total wave 1 responses) indicated that the survey items we grouped into composites are internally consistent and, for each composite, reflect a single construct that we can summarize with one number.

3. Weighting

We employed entropy balancing to address potential differences in key patient characteristics across the CJR and control groups, 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 groups on key attributes (see domains 1-4 in Exhibit E-22), and minimize differences in observable patient characteristics between CJR or control respondents relative to the full CJR population. For the stratified analysis by PAC setting, we weighted the sample of CJR and control respondents up to the CJR population within each PAC setting.

C. Results estimation

For each of the patient survey measures, we estimated the difference between CJR and control patients.

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

$$Y_{i,k,t} = b_1 + b_2 \cdot CJR_i + X_{i,k,t}{}' \cdot B + u_{i,k,t} \quad \textbf{(4)}$$

Where:

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

For subgroup analyses within each of the four discharge settings, we applied the above Equation 4 separately for patients discharged to each setting.

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 panels in Exhibit E-22). We selected these 15 patient-level characteristics to serve as covariates for all survey analyses, based on experience with the Bundled Payments for Care Improvement initiative patient survey, conceptual considerations (i.e., factors predicted to be important based on theory), and congruence with claims and assessment-based analyses.

Exhibit E-22: Risk adjustment to control for differences in patient demographics and clinical risk factors

Domain	Variables
Characteristics of the procedure	▪ Hip fracture ▪ Knee 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)[c] ▪ Hospital prior BPCI experience (non-LEJR) ▪ Herfindahl-Hirschman Index for LEJR at acute care hospitals in MSA ▪ Medicare Advantage penetration in MSA (%)

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.

[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.

[c] CJR participant hospitals that previously participated in the risk-bearing phase of BPCI for LEJR were included in the analysis. However, to be included in the control group, hospitals could not have participated in the risk-bearing phase of BPCI for LEJR.

In addition to those 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 BPCI LEJR experience, prior BPCI non-LEJR experience), and MSA-level variables (i.e., LEJR Herfindahl-Hirschman Index and Medicare Advantage Penetration) (fifth panel in Exhibit E-22).

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

Our analysis compared 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 3 and 4 sample as a whole, which is a conservative threshold for identifying potentially problematic differences between two groups (Exhibit E-23).[29]

Exhibit E-23: Summary statistics in pre-hospital functional status between CJR and control respondents – waves 3 and 4 pooled

Measure	All-LEJR		
	CJR mean	Control mean	Standard difference
Walking without rest	2.69	2.59	0.09
Going up or down stairs	2.21	2.15	0.08
Rising from sitting	2.70	2.71	0.04
Standing	2.94	2.92	0.04
Use of a mobility device	2.24	2.20	0.05
Getting on or off the toilet	2.98	2.98	0.03
Pain limiting regular activities	1.94	1.92	0.04
Medication intensity	2.72	2.65	0.06

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

***Notes*:** LEJR = lower extremity joint replacement.

Means and standardized differences are unweighted.

Standardized differences in pre-hospital functional status between CJR and control respondents were around the 0.10 threshold or below for each of the discharge settings except for IRFs (Exhibit E-24). Standardized differences between CJR and control respondents in this setting exceeded 0.10 for all measures. Differences in the amount of change in functional status between CJR and control respondents discharged to the IRF should be interpreted with caution.

[29] 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.

Exhibit E-24: Summary statistics in pre-hospital functional status between CJR and control respondents - waves 3 and 4, by discharge setting

Measure	Inpatient rehabilitation facility			Skilled nursing facility			Home with home health care			Home without home health care		
	CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.
Walking without rest	2.74	2.49	0.25	2.40	2.33	0.09	2.75	2.67	0.10	2.80	2.74	0.06
Going up or down stairs	2.42	2.22	0.19	2.15	2.08	0.10	2.22	2.17	0.08	2.19	2.16	0.08
Rising from sitting	3.20	3.05	0.21	2.75	2.77	0.10	2.64	2.66	0.03	2.65	2.63	0.11
Standing	3.41	3.13	0.28	2.96	2.93	0.08	2.90	2.90	0.06	2.91	2.87	0.07
Use of a mobility device	2.13	2.03	0.12	1.99	1.99	0.05	2.31	2.28	0.05	2.34	2.30	0.05
Getting on or off the toilet	3.44	3.29	0.20	3.01	3.01	0.07	2.94	2.95	0.06	2.96	2.89	0.10
Pain limiting regular activities	2.94	2.59	0.21	2.28	2.24	0.09	1.79	1.75	0.07	1.72	1.73	0.06
Medication intensity	3.06	2.89	0.18	2.83	2.76	0.07	2.67	2.59	0.09	2.63	2.59	0.06

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

***Notes*:** Std diff = standardized difference.

Means and standardized differences are unweighted.

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

Differences in functional status and care 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 E-22). For each of these measures, we calculated the standardized difference in the unweighted mean between CJR and control respondents. Standardized differences were all below 0.10 for the pooled wave 3 and 4 sample as a whole (Exhibit E-25).

Exhibit E-25: Summary statistics in claims-based patient characteristics between CJR and control respondents – waves 3 and 4 pooled

Measure	All-LEJR		
	CJR mean	Control mean	Std. diff.
Hip fracture	0.09	0.09	0.01
Knee procedure	0.54	0.53	0.03
MS-DRG 469	0.03	0.04	0.04
Age	73.67	73.50	0.02
Female	0.64	0.65	0.03
Eligible for Medicaid	0.07	0.07	0.01
Disability, no ESRD	0.13	0.13	0.01
Assignment to ACO	0.42	0.44	0.04
HCC score	1.35	1.34	0.02
Prior SNF stay[a]	0.03	0.03	0.03

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

***Notes*:** 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, Std diff = standardized difference.

Means and standardized differences are unweighted.

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

Standardized differences in claims-based patient and episode characteristics between CJR and control respondents generally remained below the 0.10 threshold for each of the discharge settings except IRFs (Exhibit E-26). Standardized differences between CJR and control respondents in the IRF setting exceeded 0.10 for 5 measures out of 10. These differences indicated greater patient complexity in the CJR group relative to the control group. Our stratified analysis adjusts for observable patient characteristics within each setting, but to the extent that CJR patients are more complex in ways we cannot observe, our estimates will be biased downward (appearing to indicate worse outcomes for CJR patients), when differences were at least somewhat due to underlying shifts in the populations using each PAC setting.

Exhibit E-26: Summary statistics in claims-based patient characteristics between CJR and control respondents - waves 3 & 4, by discharge setting

Measure	Inpatient rehabilitation facility			Skilled nursing facility			Home with home health			Home without home health		
	CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.	CJR mean	Control mean	Std. diff.
Hip fracture	0.48	0.38	0.21	0.25	0.19	0.13	0.02	0.02	0.01	0.02	0.02	0.01
Knee procedure	0.29	0.37	0.18	0.47	0.51	0.09	0.57	0.55	0.05	0.57	0.53	0.08
MS-DRG 469	0.12	0.10	0.07	0.06	0.07	0.03	0.02	0.02	0.02	0.02	0.03	0.06
Age	78.14	76.74	0.16	77.40	76.62	0.10	72.47	72.24	0.03	71.59	71.57	0.00
Female	0.68	0.71	0.06	0.74	0.74	0.01	0.62	0.63	0.03	0.59	0.58	0.02
Eligible for Medicaid	0.10	0.08	0.07	0.12	0.10	0.05	0.05	0.05	0.00	0.05	0.04	0.03
Disability, no ESRD	0.14	0.16	0.07	0.15	0.14	0.03	0.12	0.12	0.01	0.12	0.13	0.00
Assignment to ACO	0.38	0.43	0.10	0.41	0.45	0.09	0.44	0.45	0.01	0.39	0.43	0.08
HCC score	2.13	1.93	0.14	1.83	1.75	0.06	1.17	1.16	0.01	1.12	1.04	0.10
Prior SNF stay[a]	0.03	0.04	0.07	0.09	0.09	0.00	0.01	0.01	0.00	0.01	0.01	0.02

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

***Notes*:** 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, Std diff = standardized difference.

Means and standardized differences are unweighted.

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

G. Limitations

The patient survey does not include every LEJR episode that was initiated by a CJR or control hospital during the sampling period, and is limited to four months of intervention period data. Those survey measures that focus on past events (e.g., recalled functional status a week prior to surgery, PAC received weeks or months prior to the survey) are subject to recall bias, which may lead to mismeasurement of outcomes. This type of measurement error wouldn't 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).

After the CJR model began, fewer patients were first discharged to an IRF or SNF and more patients were discharged to an HHA in the CJR group, relative to the control group. Analysis of claims and patient assessment data indicate a greater increase in patient complexity after the start of the CJR model among CJR patients first discharged to an IRF, and to a lesser extent to a SNF and an HHA, relative to control respondents. The stratified patient survey analysis adjusts for observable patient characteristics within each setting, but decisions regarding PAC setting are likely based, in part, on patient characteristics we cannot observe in the data (such as health conditions not included in the HCC score). To the extent that CJR patients are more complex in ways we cannot observe, the estimates by PAC setting will be biased downward (i.e., appearing to indicate worse outcomes for CJR patients, when differences were at least somewhat due to underlying shifts in the populations using each PAC setting). Stratifying patient survey results by PAC setting establishes a lower-bound estimate of the impact of the CJR model within each setting. That is, we can be confident that the true impact of the CJR model is not more negative than our estimates.

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 (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.[30]

B. Time periods

The analysis was at the market-quarter level and covered October 2007 to December 2018. 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 9, 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 (January – December 2018)** begins the date that new changes to the CJR model were implemented (January 1, 2018) and ends with the end of PY3 (December 31, 2018).

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) procedures in calendar year 2018, identified using CPT code 27447 in Part B institutional claims.[31]

[30] 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.

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

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[32] 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, physician group practices, 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 three years of the model, there were 39 CJR PY3 mandatory hospitals located across nine markets that were formerly Bundled Payments for Care Improvement LEJR participants, and therefore their baseline market shares are included in both the CJR dose and the Bundled Payments for Care Improvement dose potentially overstating bundled payment penetration in these markets. If we 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] $$V_{i,t} = b_0 + b_{1i} + b_{2t} + b_{3i} + (b_4 \cdot Z_{it}) + (b_5 \cdot CJRdose_i \cdot CJRInterimCJR_1) + (b_6 \cdot CJRdose_i \cdot CJRPY1 - PY2_2) + (b_7 \cdot CJRdose_i \cdot CJRPY3_3) + (b_8 \cdot BPCIdose_i \cdot BPCIPost_1) + (b_9 \cdot BPCIdose_i \cdot BPCIPost_2) + (b_{10} \cdot BPCIdose_i \cdot BPCIPost_3) + \varepsilon_{it}$$

Where:

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

[32] 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_{3i} 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 accountable care organizations (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).
- $CJRInterimCJR_1$, $CJRPY1 – PY2_2$, **and** $CJRPY3_3$ equal 1 during the interim period and each CJR intervention period, respectively (July 2015 – March 2016, April 2016 – December 2017, and January – December 2018)
- $BPCIdose_i$ 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).[33]

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

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

the market and the inverse probability of selection into the CJR model.[34] Finally, we tested whether the CJR and control group discharge rates were significantly different at the CJR PY1/ PY2 and CJR PY3 time periods.

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 Bundled Payments for Care Improvement 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 Bundled Payments for Care Improvement provider whether the provider had yet to participate, dropped out, or continued to participate through the end of the Bundled Payments for Care Improvement 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-participating MSAs, and can lead to overlap between the CJR and Bundled Payments for Care Improvement doses, either of which would bias estimates toward zero. However, constructing the measures in this way was necessary so that the CJR and Bundled Payments for Care Improvement doses were not endogenous to performance under the CJR and Bundled Payments for Care Improvement initiatives.

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

VIII. Factors associated with earning reconciliation payments under the CJR models

We identified market, 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

Hospitals were included if they were located in the 34 metropolitan statistical areas (MSA) required to continue participation in the CJR model in PY3. 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 241 hospitals with net payment reconciliation amounts (NPRA) in PY1, 289 hospitals in PY2, and 277 hospitals in PY3.

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.[35]

C. Statistical model

The analysis was conducted at the hospital-performance year level. The relationship between the average reconciliation payment per episode and market, 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.[36] 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 $p<0.10$.

$$[1] \quad NPRA_{h,t} = b_0 + b_1 \cdot Market_h + b_2 \cdot Hospital_h + b_3 \cdot Patient_{ht} + b_4 \cdot Model_{ht} + \varepsilon_{it}$$

Where:

- Market covariates measured at baseline: Number of SNF beds per 10,000 population; HHI for LEJR in ACH
- Hospital covariates measured at baseline: Census region, bed count, ownership, DSH patient percentage, any affiliation with a medical school, and ever participated in BPCI LEJR.

[35] 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.

[36] The Lewin Group. CMS Comprehensive Care for Joint Replacement Model: Performance Year 2 Evaluation Report. CMS.gov. https://innovation.cms.gov/Files/reports/cjr-secondannrpt.pdf. Published June 2019. Accessed October 17, 2019.

- 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, 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 percent difference 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 E-27).

Exhibit E-27: Median values used to create binary versions of the continuous covariates for the average reconciliation payment per episode regression analysis

Covariate	Median value
Bed count	224
DSH patient percentage	26.9%
Percent difference between hospital historical average payments and PY target price	3.0%
Percent of episodes DRG 470 elective	79.6%
Percent of episodes female	66.1%
Percent of episodes 80 years or older	26.6%
Percent of episodes non-Hispanic Black	3.9%
Average HCC score	1.59
Percent of episodes dual eligible	14.3%
Percent of episodes disabled (no ESRD)	16.1%
Percent of episodes with a prior institutional PAC stay	5.3%

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), and PY3 (episodes ending between January and December 2018).

***Note*:** DRG = diagnosis related group, DSH = disproportionate share hospital, ESRD = end stage renal disease, FY = fiscal year, HCC = hierarchical condition category, IPPS = Inpatient Prospective Payment System, PAC = post-acute care, PY = performance year.

D. Limitations

The analysis examined PY1, PY2, and PY3 NPRA. PY1 and PY2 NPRA are considered final as of the writing of this report, while the PY3 results are preliminary and subject to change when they are finalized in spring 2020. 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 patient characteristics

For our univariate analysis, we estimated an OLS DiD regression[37] on various beneficiary characteristics. This analysis does 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.

B. Analyses of composite measure of patient complexity

Because multiple beneficiary characteristics are related to patient complexity, we developed a composite measure of patient complexity using predicted episode spending. This composite measure was created in four steps.

First, we estimated an OLS regression of total episode payments on the beneficiary characteristics included in our risk adjustment model (Equation 5).

$$Y_{i,k,t} = b_0 + X_{i,k,t}{}' \cdot B + u_{i,k,t} \quad \textbf{(5)}$$

Where $Y_{i,k,t}$ is the total episode spending of beneficiary *i* at hospital *k* which occurred in quarter *t.* We used a variety of beneficiary characteristics in $X_{i,t}{}'$ including: HCC score, squared HCC score, indicators for each HCC flag, age, sex, race/ethnicity, disability status at Medicare enrollment (not ESRD), Medicaid eligibility status, MS-DRG, procedure type (total hip arthroplasty (THA) or TKA), diabetes, obesity, hypertension, tobacco use, and prior utilization measures.[38] We restricted this regression to only include baseline episodes so that the estimated coefficients are not influenced by the CJR model. The results of this regression represent the relationship between patient complexity and total payments before the CJR model was implemented.

Second, using the estimated coefficients from the baseline total payments regression, we calculated a predicted payment value for every baseline and intervention episode of mandatory CJR and control hospitals (Equation 6).

$$Predicted\ Y_{i,k,t} = \hat{b}_0 + X_{i,t}{}' \cdot \hat{B} \quad \textbf{(6)}$$

where *Predicted* $Y_{i,t}$ can be interpreted as a composite measure of beneficiary *i*'s complexity as measured by the relationship defined in Equation 5 during the baseline period.

Third, for each hospital *k* we calculated an average predicted payment value across all baseline quarters, *Predicted* $Y_{k,0}$, and an average across all intervention quarters, *Predicted* $Y_{k,1}$.

Fourth, for each hospital *k* we constructed a *predicted payment ratio* using these hospital averages by Equation 7.

$$Predicted\ ratio_k = \frac{Predicted\ Y_{k,1}}{Predicted\ Y_{k,0}} \quad \textbf{(7)}$$

[37] See Section III.D.1 for additional details about our DiD design.
[38] See Section III.D.2.a for additional details about these variables.

A predicted ratio for hospital *k* that is greater than one indicates that this hospital saw an increase in their average patient complexity. A predicted ratio below indicates a decrease in average patient complexity.

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.

1. Analytic sample

Low-volume hospitals are more likely to have extreme values of the predicted ratio due to small sample sizes in baseline or intervention. To ensure that this analysis is not influenced by extreme values, we excluded hospitals that have relatively low volume. For a given episode type, we ranked all hospitals based on the minimum of their baseline and intervention volume and then excluded the lower 25th percentile. We performed a sensitivity analysis including all hospitals (see below).

2. Sensitivity analyses

A number of sensitivity analyses were performed on the predicted payment findings. First, we calculated an alternative predicted payment ratio that uses the percent of episodes above the median instead of the mean predicted payment ratio of the baseline and intervention period. Second, because we were concerned that hospitals with unusual baseline episode spending patterns (e.g., unusual utilization patterns of post-acute care given case mix) could be influencing the payment regression coefficients, we used hospital fixed effects in addition to the beneficiary characteristics in Equation 5. Lastly, instead of excluding hospitals with volume in the lower 25th percentile, we included all hospitals and weight each ratio by the hospital's minimum of their baseline and intervention volume. The alternative specifications used in the sensitivity analyses did not materially affect any of the findings in the patient mix analysis and thus provided evidence that this analysis and the conclusions presented in this report are robust.

X. Qualitative and Mixed Methods Analysis

A. Data collection

Provider telephone interviews were staffed with one interviewer and one note taker. 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 7.5.18; Scientific Software Development GmbH, Berlin, Germany) for coding and analysis.

Site visit interviews were staffed with a minimum of one interviewer and one note taker. Notes were taken during site visits and, if the interviewees agreed, the interview was recorded. Site visit recordings were transcribed and reviewed for accuracy. Transcripts and notes from the interviews were used to create summaries of the site visits, and provider names and other identifiers were removed.

B. Thematic analysis and case study approach

We developed analytic codebooks based on the telephone interview protocols. The codebooks contained categories to use in the ATLAS.ti software to characterize notes from telephone interviews and identify key themes across hospitals. 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.

We developed summary templates based on the site visit protocols. The summaries contained topic categories to characterize notes from site visits and identify key themes across hospitals and MSAs. Note takers completed site visit summaries and discussed findings with interviewers to ensure consistency. We used information from the summaries as well as claims data to draft case studies for each MSA.

The case study approach provides an in-depth description of CJR model experiences for each site visit participant hospital, as well as a detailed summary of MSA-level factors that may have influenced a response to the model (e.g., MSA surgeon supply). Case studies used other evaluation data sources to further describe hospital and MSA-level characteristics. Claims-based data were used to assess how payments and utilization may have shifted under the CJR model for each hospital. This approach provides a more comprehensive understanding of the effects of the CJR model on hospitals, orthopedic surgery groups, and PAC providers. Case study findings were included in the annual report to provide contextual examples when appropriate, and the three PY3 MSA-level case studies are located in the case study supplement.

1. Limitations

The analysis of the site visit and telephone interview data provide descriptions of themes and patterns in response to the protocols, which may not represent all CJR participants. The Round 6

telephone interviews aimed to capture the perspectives of hospitals that shifted a large portion of their TKA to the outpatient setting in response to the rule change removing elective TKA from the Medicare inpatient only list.[39] We therefore included an intentional sample of all hospitals with greater than or equal to 50% outpatient TKA in 2018 (n=31), and then a random sample of an additional 100 hospitals. For site visits, medium sized MSAs were purposefully sampled, which allowed a large enough sample of hospitals to identify MSA-level themes. Findings are limited to individual interviewees and may not represent all CJR participant hospitals.

C. Clinical Review Panel

Information obtained from eight 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 E-28. 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 the 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 Centers for Medicare and Medicaid Innovations (CMMI) 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

[39] https://www.federalregister.gov/documents/2017/12/14/R1-2017-23932/medicare-program-hospital-outpatient-prospective-payment-and-ambulatory-surgical-center-payment

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 E-28: 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

***Note*:** TKA = total knee arthroplasty.

Appendix F: Outcome Definitions

Exhibit F-1: Claims-based outcome definitions

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[1]
Medicare payments	Total Medicare standardized allowed amounts per episode[2]	The sum of Medicare payment and beneficiary out-of-pocket amounts for related items and services covered by Medicare Part A and Part B[3] 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, 2018; 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, 2018; 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, 2018; 5) have non-missing Medicare standardized allowed payment information for the episode.

[1] 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.

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

[3] 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[1]
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, 2018; 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, 2018; 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, 2018; 5) have non-missing Medicare standardized allowed payment information for the episode.
Utilization	Anchor hospitalization length of hospitalization (LOS)	The number of days between the admission date and the discharge date for the LEJR inpatient anchor hospitalization (MS-DRG 469 or 470). Anchor LOS is winsorized by MS-DRG and quarter at the 1st and 99th percentiles. This is only calculated for inpatient episodes, not for outpatient TKA episodes.	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 October 3, 2018;[4] 5) have non-missing Medicare standardized allowed payment information for the episode.

[4] Beneficiaries discharged from their anchor hospitalization on or before October 3, 2018 have a 90-day post-discharge period ending on or before December 31, 2018.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[1]
Utilization (cont'd)	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, 2018; 5) have non-missing Medicare standardized allowed payment information for the episode.
	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, 2018; 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, 2018; 5) have non-missing Medicare standardized allowed payment information for the episode.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[1]
Utilization (cont'd)	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, 2018; 5) have non-missing Medicare standardized allowed payment information for the episode.
	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, 2018; 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, 2018; 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, 2018; 5) have at least one SNF day 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[1]
Utilization (cont'd)	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, 2018; 5) have at least one HHA visit during this period; 6) have non-missing Medicare standardized allowed payment information for the episode.
	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, 2018; 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, 2018; 5) have at least one outpatient PT/OT 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[1]
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).[5] 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, 2018; 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.
	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, 2018; 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.

[5] 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[1]
Quality (cont'd)	All-cause mortality rate	Death from any cause during the anchor hospitalization or 90-day post-discharge period. For beneficiaries with multiple LEJR hospitalizations during the baseline and intervention periods, one hospitalization was randomly selected across the baseline and intervention periods for inclusion in this measure.	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, 2018; 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.
	Manipulation under anesthesia (MUA)	The proportion of knee replacement episodes with any MUA procedures (CPT 27570) during the 90-day post-discharge period.	90-day post-discharge period	Beneficiaries who: 1) have an elective procedure (non-fracture); 2) have a knee replacement; 3) have a complete FFS enrollment history six months prior to the anchor hospitalization; 4) have consistent, reliable sex and age data (age <115); 5) maintain Parts A and B enrollment throughout the measurement period; 6) have a measurement period that ends on or before December 31, 2018; 7) have non-missing Medicare standardized allowed payment information for the episode.

Measure category	Outcome name	Definition	Measurement period(s)	Eligible sample[1]
Quality (cont'd)	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).[6] 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, 2018; 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.

Note: 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, LOS = length of hospitalization, LTCH = long-term care hospital, MS-DRG = Medicare Severity-Diagnosis Related Group, MUA = manipulation under anesthesia, 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.

[6] 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[7]
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.	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.

[7] 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.

First PAC setting	Outcome name	Definition	Measurement period(s)	Eligible sample[7]
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.

Appendix G: 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 "healthcare 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 healthcare 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 H: Patient Survey Results

Exhibit H-1: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, overall results

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	7969	8529	0.77	0.75	0.02 (0.7%)	p=0.44
	Difficulty walking up or down 12 stairs	-3 to 3	7455	8036	0.75	0.75	-0.01 (-0.4%)	p=0.72
	Difficulty rising from sitting	-4 to 4	8107	8711	1.21	1.24	-0.03 (-1.1%)	p<0.10
	Difficulty standing	-4 to 4	8123	8719	1.16	1.17	-0.01 (-0.3%)	p=0.60
	Use of a mobility aid	-2 to 2	7844	8401	0.15	0.16	-0.01 (-0.4%)	p=0.49
	Difficulty getting on/off the toilet	-4 to 4	8050	8615	1.37	1.39	-0.01 (-0.5%)	p=0.30
	Frequency that pain interferes with normal activities	-4 to 4	8141	8733	1.96	1.98	-0.02 (-1.0%)	p=0.21
	Medication use for pain in the joint you had replaced	-3 to 3	7911	8519	0.61	0.61	-0.00 (-0.1%)	p=0.81
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	8183	8796	79.9	80.6	-0.7	p=0.21
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	7747	8334	82.7	82.7	0.0	p=0.97
	Healthcare providers listened to preferences	0 to 100	8081	8650	77.5	78.4	-0.9	p=0.12
	Satisfaction with discharge destination	0 to 100	8110	8726	82.3	81.5	0.8	p=0.23
	Satisfaction with care coordination	0 to 100	8180	8758	82.1	81.8	0.3	p=0.61
	Satisfaction with treatment instructions	0 to 100	8103	8692	84.1	84.1	-0.0	p=0.95
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	8084	8700	86.7	87.8	-1.1	p<0.10
	Received the right amount of post-discharge care	0 to 100	8188	8762	84.7	85.9	-1.2	p<0.05
	Had all the medical equipment needed at home	0 to 100	7691	8262	92.1	92.4	-0.3	p=0.58

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	7834	8395	94.9	95.1	-0.2	p=0.78
	Composite measure of caregiver help[d]	0 to 100	7264	7798	67.5	69.4	-1.9	p<0.01
	Help needed putting on or taking off clothes[d]	0 to 100	7800	8350	59.8	62.0	-2.2	p<0.01
	Help needed bathing[d]	0 to 100	7748	8315	64.1	66.7	-2.6	p<0.01
	Help needed using the toilet[d]	0 to 100	7346	7883	79.6	81.3	-1.7	p<0.05

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 99%, 95%, or 90% significance level are indicated by red, orange, or yellow shaded cells, respectively.

[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 H-2: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, patients first discharged to an IRF

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	362	543	0.22	0.31	-0.09 (-3.4%)	p=0.29
	Difficulty walking up or down 12 stairs	-3 to 3	332	512	0.51	0.50	0.01 (0.4%)	p=0.94
	Difficulty rising from sitting	-4 to 4	366	569	0.97	1.05	-0.09 (-2.6%)	p=0.29
	Difficulty standing	-4 to 4	372	570	0.86	1.03	-0.18 (-5.0%)	p<0.05
	Use of a mobility aid	-2 to 2	370	540	-0.15	-0.07	-0.08 (-3.7%)	p<0.10
	Difficulty getting on/off the toilet	-4 to 4	368	562	1.11	1.26	-0.14 (-4.0%)	p<0.01
	Frequency that pain interferes with normal activities	-4 to 4	366	574	1.77	1.80	-0.04 (-1.2%)	p=0.67
	Medication use for pain in the joint you had replaced	-3 to 3	367	560	0.46	0.50	-0.04 (-1.2%)	p=0.54
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	374	605	76.4	78.2	-1.8	p=0.34
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	352	564	78.2	80.8	-2.6	p=0.25
	Healthcare providers listened to preferences	0 to 100	372	586	74.4	76.6	-2.1	p=0.40
	Satisfaction with discharge destination	0 to 100	375	601	78.8	78.4	0.4	p=0.89
	Satisfaction with care coordination	0 to 100	372	598	77.8	78.6	-0.8	p=0.78
	Satisfaction with treatment instructions	0 to 100	371	594	79.3	82.0	-2.7	p=0.26
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	362	584	84.4	87.4	-3.0	p<0.10
	Received the right amount of post-discharge care	0 to 100	378	597	83.7	86.4	-2.7	p<0.10
	Had all the medical equipment needed at home	0 to 100	343	545	90.7	94.0	-3.3	p<0.10

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	353	559	89.1	92.3	-3.2	p=0.20
	Composite measure of caregiver help[d]	0 to 100	320	506	70.6	75.5	-4.9	p<0.10
	Help needed putting on or taking off clothes[d]	0 to 100	349	555	63.0	68.3	-5.3	p<0.10
	Help needed bathing[d]	0 to 100	350	550	64.7	70.1	-5.4	p<0.05
	Help needed using the toilet[d]	0 to 100	329	514	82.3	85.6	-3.3	p<0.10

Source: CJR evaluation team analysis of patient survey data for patients first discharged to an IRF 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 99%, 95%, or 90% significance level are indicated by red, orange, or yellow shaded cells, respectively.

IRF = inpatient rehabilitation facility.

[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 H-3: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, patients first discharged to a SNF

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	1698	2081	0.56	0.47	0.09 (3.7%)	p<0.10
	Difficulty walking up or down 12 stairs	-3 to 3	1592	1948	0.66	0.63	0.03 (1.5%)	p=0.44
	Difficulty rising from sitting	-4 to 4	1752	2138	1.10	1.11	-0.01 (-0.3%)	p=0.80
	Difficulty standing	-4 to 4	1757	2135	1.04	1.02	0.01 (0.5%)	p=0.74
	Use of a mobility aid	-2 to 2	1681	2048	0.03	0.03	0.01 (0.4%)	p=0.67
	Difficulty getting on/off the toilet	-4 to 4	1740	2105	1.30	1.31	-0.02 (-0.5%)	p=0.61
	Frequency that pain interferes with normal activities	-4 to 4	1753	2132	1.89	1.91	-0.02 (-0.8%)	p=0.59
	Medication use for pain in the joint you had replaced	-3 to 3	1688	2067	0.57	0.57	-0.00 (-0.1%)	p=0.96
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	1811	2175	77.0	78.5	-1.5	p=0.26
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	1684	2043	76.2	77.0	-0.8	p=0.58
	Healthcare providers listened to preferences	0 to 100	1777	2130	72.8	74.4	-1.6	p=0.37
	Satisfaction with discharge destination	0 to 100	1818	2177	70.9	71.0	-0.1	p=0.94
	Satisfaction with care coordination	0 to 100	1813	2158	76.9	77.9	-1.0	p=0.52
	Satisfaction with treatment instructions	0 to 100	1764	2139	80.3	81.1	-0.8	p=0.57
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	1716	2083	83.6	87.0	-3.5	p<0.01
	Received the right amount of post-discharge care	0 to 100	1788	2151	81.4	84.0	-2.6	p<0.05
	Had all the medical equipment needed at home	0 to 100	1599	1956	91.0	91.6	-0.6	p=0.61

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	1635	1981	87.4	87.7	-0.2	p=0.87
	Composite measure of caregiver help[d]	0 to 100	1486	1814	72.8	75.4	-2.6	p<0.10
	Help needed putting on or taking off clothes[d]	0 to 100	1622	1977	66.7	69.5	-2.9	p<0.10
	Help needed bathing[d]	0 to 100	1613	1959	67.4	71.0	-3.6	p<0.05
	Help needed using the toilet[d]	0 to 100	1516	1839	82.3	83.4	-1.1	p=0.26

Source: CJR evaluation team analysis of patient survey data for patients first discharged to a SNF 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 99%, 95%, or 90% significance level are indicated by red, orange, or yellow shaded cells, respectively.

SNF = skilled nursing facility.

[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 H-4: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, patients discharged directly home with HH care

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	4401	4082	0.89	0.88	0.01 (0.5%)	p=0.64
	Difficulty walking up or down 12 stairs	-3 to 3	4107	3855	0.84	0.86	-0.02 (-1.0%)	p=0.38
	Difficulty rising from sitting	-4 to 4	4454	4151	1.28	1.32	-0.04 (-1.5%)	p<0.05
	Difficulty standing	-4 to 4	4459	4160	1.26	1.28	-0.02 (-0.9%)	p=0.21
	Use of a mobility aid	-2 to 2	4314	4011	0.25	0.28	-0.03 (-1.3%)	p<0.05
	Difficulty getting on/off the toilet	-4 to 4	4418	4107	1.45	1.46	-0.01 (-0.5%)	p=0.44
	Frequency that pain interferes with normal activities	-4 to 4	4480	4172	1.99	2.00	-0.01 (-0.7%)	p=0.54
	Medication use for pain in the joint you had replaced	-3 to 3	4349	4072	0.62	0.64	-0.03 (-1.0%)	p=0.27
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	4468	4166	81.2	82.4	-1.1	p<0.05
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	4239	3954	84.3	85.1	-0.8	p=0.23
	Healthcare providers listened to preferences	0 to 100	4412	4105	78.5	80.1	-1.6	p<0.05
	Satisfaction with discharge destination	0 to 100	4407	4118	85.9	86.1	-0.2	p=0.72
	Satisfaction with care coordination	0 to 100	4473	4162	83.9	84.8	-1.0	p<0.10
	Satisfaction with treatment instructions	0 to 100	4442	4125	85.3	86.2	-0.9	p=0.13
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	4471	4178	88.1	88.5	-0.4	p=0.61
	Received the right amount of post-discharge care	0 to 100	4506	4196	86.8	87.2	-0.4	p=0.28
	Had all the medical equipment needed at home	0 to 100	4280	3992	91.6	92.2	-0.6	p=0.44

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	4359	4057	97.1	97.4	-0.3	p=0.51
	Composite measure of caregiver help[d]	0 to 100	4059	3799	66.1	67.4	-1.3	p=0.13
	Help needed putting on or taking off clothes[d]	0 to 100	4344	4030	57.7	59.2	-1.4	p=0.11
	Help needed bathing[d]	0 to 100	4314	4028	63.3	65.2	-1.9	p<0.10
	Help needed using the toilet[d]	0 to 100	4091	3835	78.5	80.1	-1.6	p=0.13

Source: CJR evaluation team analysis of patient survey data for patients discharged directly home with HH care 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 99%, 95%, or 90% significance level are indicated by red, orange, or yellow shaded cells, respectively.

HH = home health.

[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 H-5: Risk-adjusted survey-based results for change in functional status, satisfaction with overall recovery, satisfaction with care management, care transitions, and caregiver help, patients discharged directly home without HH care

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	1427	1782	0.90	0.85	0.05 (1.7%)	p=0.25
	Difficulty walking up or down 12 stairs	-3 to 3	1345	1682	0.82	0.82	0.00 (0.2%)	p=0.88
	Difficulty rising from sitting	-4 to 4	1454	1812	1.26	1.25	0.01 (0.5%)	p=0.71
	Difficulty standing	-4 to 4	1454	1813	1.21	1.19	0.02 (0.7%)	p=0.56
	Use of a mobility aid	-2 to 2	1397	1763	0.25	0.25	0.00 (0.1%)	p=0.91
	Difficulty getting on/off the toilet	-4 to 4	1443	1801	1.41	1.39	0.02 (0.7%)	p=0.50
	Frequency that pain interferes with normal activities	-4 to 4	1460	1815	2.00	2.02	-0.02 (-1.1%)	p=0.65
	Medication use for pain in the joint you had replaced	-3 to 3	1427	1781	0.65	0.60	0.05 (1.7%)	p=0.22
Satisfaction with overall recovery[b]	Satisfaction with overall recovery since leaving the hospital	0 to 100	1451	1809	81.1	79.3	1.8	p=0.21
Satisfaction with care management[b]	Composite measure of satisfaction with care management	0 to 100	1396	1733	84.4	82.6	1.9	p<0.10
	Healthcare providers listened to preferences	0 to 100	1440	1789	78.7	77.7	1.0	p=0.43
	Satisfaction with discharge destination	0 to 100	1433	1789	85.8	83.5	2.3	p<0.10
	Satisfaction with care coordination	0 to 100	1444	1799	83.6	82.0	1.6	p=0.28
	Satisfaction with treatment instructions	0 to 100	1447	1793	84.3	82.6	1.7	p=0.22
Experience with care transitions[c]	Discharged from the hospital at the right time	0 to 100	1453	1816	87.0	86.8	0.2	p=0.90
	Received the right amount of post-discharge care	0 to 100	1434	1779	84.3	84.8	-0.5	p=0.73
	Had all the medical equipment needed at home	0 to 100	1389	1732	93.6	92.4	1.2	p=0.14

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	1407	1759	96.9	97.0	-0.1	p=0.90
	Composite measure of caregiver help[d]	0 to 100	1324	1644	67.2	67.1	0.1	p=0.96
	Help needed putting on or taking off clothes[d]	0 to 100	1406	1750	58.6	57.9	0.8	p=0.59
	Help needed bathing[d]	0 to 100	1392	1740	66.1	64.9	1.3	p=0.34
	Help needed using the toilet[d]	0 to 100	1335	1660	79.9	77.8	2.1	p=0.11

Source: CJR evaluation team analysis of patient survey data for patients discharged directly home without HH care 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 99%, 95%, or 90% significance level are indicated by red, orange, or yellow shaded cells, respectively.

HH = home health.

[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 recalled status among all CJR respondents.

[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 I: Change in Patient Characteristics

Exhibit I-1: Change in patient complexity measures, LEJR patients whose first PAC setting was an inpatient rehabilitation facility, PY1-3

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90th percentile LCI	90th percentile UCI
		Baseline average (N=18,812)	Intervention average (N=6,662)	Baseline average (N=20,258)	Intervention average (N=9,157)					
IRF	Age 80+	42.6%	48.8%	39.8%	41.1%	4.9	11.6%	p<0.01	1.8	8.1
	Female	71.1%	68.8%	71.1%	70.6%	-1.8	-2.6%	p<0.10	-3.6	0.0
	Black	7.9%	7.1%	8.0%	8.1%	-0.9	-11.1%	p=0.32	-2.3	0.6
	Eligible for Medicaid	16.2%	15.9%	12.0%	10.7%	0.9	5.8%	p=0.41	-0.9	2.8
	Disability, no ESRD	16.2%	16.2%	16.3%	15.3%	1.0	6.3%	p=0.23	-0.4	2.4
	HCC Score	1.84	2.18	1.79	1.98	0.15	8.0%	p<0.01	0.1	0.2
	Obesity	16.4%	23.3%	18.5%	27.8%	-2.5	-15.0%	p=0.11	-5.0	0.1
	Diabetes	34.0%	33.1%	32.4%	31.2%	0.2	0.6%	p=0.86	-1.7	2.1
	Hypertension	79.2%	79.1%	80.3%	81.3%	-1.1	-1.4%	p=0.20	-2.5	0.3
	Dementia	8.9%	11.7%	9.2%	9.7%	2.4	26.6%	p<0.01	1.0	3.7
	CHF	20.3%	22.0%	19.9%	21.5%	0.1	0.5%	p=0.94	-1.9	2.1
	Prior ACH stay	15.3%	17.5%	16.8%	17.1%	1.8	12.0%	p<0.05	0.4	3.3
	Prior IRF stay	5.4%	6.5%	6.2%	6.8%	0.5	8.6%	p=0.31	-0.3	1.2
	Prior SNF stay	3.7%	4.6%	3.8%	4.2%	0.4	12.2%	p=0.33	-0.3	1.2
	Prior HH use	17.0%	19.4%	17.3%	19.3%	0.4	2.2%	p=0.76	-1.6	2.3
	Any prior care	35.9%	40.2%	37.7%	40.6%	1.4	4.0%	p=0.21	-0.4	3.3
	MS-DRG 469	8.1%	11.5%	8.2%	9.6%	2.1	25.8%	p<0.05	0.7	3.4
	Hip fracture	32.4%	49.4%	31.5%	37.3%	11.3	34.9%	p<0.01	6.4	16.2
	Mobility index[a]	8.1	7.4	8.2	7.9	-0.4	-4.9%	p<0.05	-0.6	-0.1
	Self-care index[a]	20.5	19.2	20.9	19.9	-0.3	-1.5%	p=0.48	-0.9	0.4
	Cognitive index[a]	25.8	24.3	25.3	24.0	-0.1	-0.4%	p=0.81	-0.9	0.7

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 2018 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 2018 (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 99%, 95%, or 90% 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, DiD = difference-in-differences, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, IRF = inpatient rehabilitation facility, IRF-PAI = Inpatient Rehabilitation Facility Patient Assessment Instrument, LCI = lower confidence interval, LEJR = lower extremity joint replacement, MS-DRG = Medicare Severity-Diagnosis Related Group, PAC = post-acute care, PY = performance year, SNF = skilled nursing facility, UCI = upper confidence interval.

[a] 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 16,054 CJR baseline episodes, 5,373 CJR intervention episodes, 17,615 control group baseline episodes, and 7,590 control group intervention episodes.

Exhibit I-2: Change in patient complexity measures, LEJR patients whose first PAC setting was a skilled nursing facility, PY1-3

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90th percentile LCI	90th percentile UCI
		Baseline average (N=57,721)	Intervention average (N=31,281)	Baseline average (N=65,910)	Intervention average (N=32,302)					
SNF	Age 80+	40.4%	43.7%	38.4%	39.6%	2.1	5.3%	p<0.10	0.1	4.2
	Female	72.8%	73.3%	73.0%	73.8%	-0.3	-0.4%	p=0.64	-1.4	0.8
	Black	6.7%	6.8%	7.3%	8.4%	-1.0	-15.6%	p=0.12	-2.2	0.1
	Eligible for Medicaid	18.4%	18.6%	13.5%	14.5%	-0.9	-4.8%	p=0.45	-2.8	1.0
	Disability, no ESRD	14.9%	15.6%	14.6%	16.0%	-0.6	-4.3%	p=0.35	-1.8	0.5
	HCC Score	1.71	1.98	1.59	1.80	0.06	3.2%	p=0.22	0.0	0.1
	Obesity	16.3%	27.4%	16.8%	29.2%	-1.3	-7.9%	p=0.43	-4.0	1.4
	Diabetes	33.3%	34.7%	29.3%	29.8%	0.9	2.7%	p=0.21	-0.3	2.1
	Hypertension	78.9%	81.0%	77.8%	79.2%	0.7	0.9%	p=0.25	-0.3	1.7
	Dementia	12.9%	15.0%	12.3%	12.9%	1.5	11.4%	p=0.13	-0.1	3.1
	CHF	18.8%	20.2%	17.2%	18.4%	0.1	0.6%	p=0.90	-1.3	1.6
	Prior ACH stay	15.6%	17.3%	14.7%	15.7%	0.7	4.6%	p=0.11	0.0	1.4
	Prior IRF stay	1.1%	1.3%	1.1%	1.2%	0.0	1.9%	p=0.92	-0.3	0.4
	Prior SNF stay	9.5%	10.6%	9.0%	9.9%	0.1	1.5%	p=0.60	-0.3	0.6
	Prior HH use	16.4%	17.7%	15.3%	16.8%	-0.3	-1.6%	p=0.79	-1.9	1.4
	Any prior care	36.0%	39.2%	35.1%	38.4%	-0.1	-0.3%	p=0.91	-1.9	1.7
	MS-DRG 469	7.1%	9.1%	6.6%	7.8%	0.8	11.0%	p=0.29	-0.4	2.0
	Hip fracture	24.8%	30.9%	21.1%	25.3%	1.9	7.5%	p=0.39	-1.7	5.4
	Bathing poor[a]	87.5%	86.6%	81.1%	79.3%	0.9	1.1%	p=0.67	-2.6	4.5
	Cognition not intact[a]	20.6%	21.1%	19.6%	20.0%	0.1	0.5%	p=0.94	-2.1	2.3
	Severe cognitive impairment[a]	6.4%	7.3%	5.9%	6.7%	0.1	1.2%	p=0.88	-0.8	0.9
	Early-loss ADL score[a]	5.0	5.0	4.7	4.6	0.1	2.0%	p<0.10	0.0	0.2
	Mid-loss ADL score (i.e., motion score) [a]	8.3	8.3	7.8	7.6	0.2	2.4%	p<0.10	0.0	0.4
	Late-loss ADL score[a]	13.4	13.4	12.8	12.6	0.2	1.5%	p=0.13	0.0	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 2018 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 2018 (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 99%, 95%, or 90% 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, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, IRF = inpatient rehabilitation facility, LCI = lower confidence interval, 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, UCI = upper confidence interval.

[a] 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 57,721 CJR baseline episodes, 31,281 CJR intervention episodes, 65,910 control group baseline episodes, and 32,302 control group intervention episodes.

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

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90th percentile LCI	90th percentile UCI
		Baseline average (N=51,280)	Intervention average (N=51,914)	Baseline average (N=61,683)	Intervention average (N=47,381)					
HHA	Age 80+	14.5%	15.5%	12.7%	13.2%	0.5	3.7%	p=0.28	-0.3	1.3
	Female	59.2%	61.8%	59.3%	60.8%	1.1	1.9%	p<0.10	0.2	2.1
	Black	5.8%	5.8%	7.2%	7.3%	-0.1	-2.1%	p=0.82	-1.0	0.7
	Eligible for Medicaid	10.0%	9.1%	9.3%	8.4%	-0.1	-1.5%	p=0.88	-1.7	1.4
	Disability, no ESRD	16.5%	15.0%	17.3%	16.1%	-0.2	-1.5%	p=0.80	-1.8	1.3
	HCC Score	1.11	1.22	1.08	1.17	0.01	1.0%	p=0.68	0.0	0.1
	Obesity	14.8%	30.8%	16.2%	31.4%	0.7	5.0%	p=0.78	-3.5	5.0
	Diabetes	25.1%	26.8%	25.0%	24.4%	2.4	9.4%	p<0.01	1.4	3.3
	Hypertension	71.3%	73.1%	72.5%	72.8%	1.6	2.2%	p<0.05	0.5	2.6
	Dementia	2.2%	2.0%	2.3%	2.1%	-0.1	-2.6%	p=0.74	-0.3	0.2
	CHF	9.5%	10.1%	9.6%	9.6%	0.5	5.7%	p=0.23	-0.2	1.3
	Prior ACH stay	10.0%	9.6%	10.1%	9.9%	-0.3	-2.8%	p=0.45	-0.9	0.3
	Prior IRF stay	0.4%	0.4%	0.4%	0.5%	0.0	-11.0%	p=0.49	-0.1	0.1
	Prior SNF stay	1.1%	1.0%	1.0%	1.0%	-0.1	-12.7%	p=0.14	-0.3	0.0
	Prior HH use	10.3%	9.6%	9.7%	9.9%	-0.8	-8.1%	p=0.12	-1.7	0.1
	Any prior care	23.4%	23.4%	23.3%	24.0%	-0.8	-3.3%	p=0.46	-2.5	0.9
	MS-DRG 469	1.7%	1.5%	1.7%	1.6%	0.0	-2.6%	p=0.82	-0.4	0.3
	Hip fracture	2.4%	2.4%	2.3%	2.4%	-0.1	-2.5%	p=0.85	-0.6	0.5
	Toilet transferring[a]	15.2%	27.1%	14.5%	28.8%	-2.4	-15.9%	p=0.38	-7.0	2.1
	Transferring[a]	30.8%	68.6%	30.1%	69.5%	-1.5	-4.8%	p=0.63	-6.6	3.6
	Ambulation / locomotion[a]	50.0%	80.3%	52.0%	80.4%	1.9	3.7%	p=0.44	-2.1	5.8
	Lower body dressing[a]	89.1%	94.6%	88.8%	94.4%	0.0	0.0%	p=0.96	-1.5	1.4
	Upper body dressing[a]	26.2%	46.5%	25.0%	42.8%	2.5	9.4%	p=0.23	-0.9	5.8
	Bathing[a]	75.7%	88.5%	78.1%	87.7%	3.4	4.4%	p<0.10	0.4	6.3

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90th percentile LCI	90th percentile UCI
		Baseline average (N=51,280)	Intervention average (N=51,914)	Baseline average (N=61,683)	Intervention average (N=47,381)					
HHA cont'd	Toileting hygiene[a]	31.5%	62.3%	30.1%	60.1%	0.7	2.2%	p=0.80	-4.0	5.4
	Grooming[a]	21.2%	40.6%	21.0%	39.4%	0.9	4.3%	p=0.71	-3.2	5.0
	Cognitive functioning[a]	1.3%	1.6%	1.4%	1.8%	0.0	-2.4%	p=0.91	-0.5	0.4
	Confusion[a]	1.2%	1.3%	1.3%	1.3%	0.0	3.2%	p=0.88	-0.4	0.5
	Memory deficit[a]	2.1%	2.1%	2.4%	2.4%	-0.1	-2.4%	p=0.84	-0.5	0.4
	Impaired decision-making[a]	4.3%	4.7%	4.7%	5.6%	-0.5	-10.6%	p=0.50	-1.6	0.7
	Overall status[a]	8.9%	14.4%	6.9%	10.7%	1.8	19.9%	p=0.24	-0.7	4.2

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 2018 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 2018 (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 99%, 95%, or 90% 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, DiD = difference-in-differences, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, HHA = home health agency, IRF = inpatient rehabilitation facility, LCI = lower confidence interval, 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, UCI = upper confidence interval.

[a] 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 I-4: Change in patient complexity measures, LEJR patients who were discharged home without home health care, PY1-3

First PAC setting	Measure	CJR		Control group		Net differences	Net differences % of baseline	p-value	90th percentile LCI	90th percentile UCI
		Baseline average (N=11,842)	Intervention average (N=14,625)	Baseline average (N=20,035)	Intervention average (N=20,810)					
None	Age 80+	14.5%	12.4%	12.2%	11.5%	-1.3	-9.2%	p=0.16	-2.9	0.2
	Female	54.1%	56.2%	56.3%	56.8%	1.6	3.0%	p=0.16	-0.3	3.5
	Black	4.7%	4.8%	6.1%	6.2%	0.0	-0.3%	p=0.98	-1.1	1.0
	Eligible for Medicaid	9.8%	8.2%	8.1%	6.9%	-0.4	-4.0%	p=0.64	-1.7	1.0
	Disability, no ESRD	16.0%	14.0%	16.1%	15.0%	-0.8	-5.2%	p=0.39	-2.4	0.7
	HCC Score	1.13	1.13	1.07	1.10	-0.02	-1.9%	p=0.54	-0.1	0.0
	Obesity	13.2%	29.6%	14.1%	27.3%	3.3	24.7%	p<0.10	0.3	6.2
	Diabetes	23.9%	22.9%	23.4%	23.0%	-0.6	-2.4%	p=0.50	-2.0	0.8
	Hypertension	69.5%	69.9%	71.0%	71.2%	0.3	0.4%	p=0.78	-1.2	1.8
	Dementia	4.0%	2.6%	3.5%	2.4%	-0.3	-7.0%	p=0.52	-1.0	0.4
	CHF	10.1%	9.2%	9.4%	9.2%	-0.7	-7.2%	p=0.28	-1.8	0.4
	Prior ACH stay	11.6%	10.1%	11.1%	10.4%	-0.7	-6.3%	p=0.30	-1.9	0.4
	Prior IRF stay	0.6%	0.4%	0.6%	0.5%	0.0	-5.2%	p=0.82	-0.2	0.2
	Prior SNF stay	2.8%	1.4%	1.9%	1.2%	-0.7	-25.4%	p<0.05	-1.3	-0.2
	Prior HH use	5.0%	3.6%	4.0%	3.1%	-0.5	-9.2%	p=0.43	-1.4	0.5
	Any prior care	23.4%	21.9%	22.5%	22.1%	-1.1	-4.9%	p=0.28	-2.9	0.6
	MS-DRG 469	3.0%	2.2%	2.6%	2.4%	-0.5	-17.7%	p=0.28	-1.3	0.3
	Hip fracture	6.4%	3.3%	4.8%	2.6%	-0.9	-13.8%	p=0.24	-2.1	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 2018 (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 99%, 95%, or 90% 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, DiD = difference-in-differences, ESRD = end-stage renal disease, HCC = hierarchical condition category, HH = home health, IRF = inpatient rehabilitation facility, LCI = lower confidence interval, LEJR = lower extremity joint replacement, MS-DRG = Medicare Severity-Diagnosis Related Group, PAC = post-acute care, PY = performance year, SNF = skilled nursing facility, UCI = upper confidence interval.

Exhibit I-5: Changes in patient characteristics between baseline and intervention by MS-DRG and fracture status, PY1-3

Patient characteristics		Net differences in average values			
		MS-DRG 470, elective	MS-DRG 469, elective	MS-DRG 470, fracture	MS-DRG 469, fracture
Age	20-64 (pp)	-0.3	-0.9	0.0	0.1
	65-79 (pp)	0.6	-3.4	-1.3	1.3
	80+ (pp)	-0.3	4.3	1.2	-1.5
Sex	Female (pp)	0.2	0.9	-1.4	-2.0
Race/ethnicity	White (pp)	0.9	1.4	0.0	-2.7
	Black (pp)	-0.8	-1.9	0.2	1.7
	Hispanic (pp)	-0.3	1.0	-0.3	0.5
	Other (pp)	0.1	-0.3	0.1	0.6
	Unknown (pp)	0.1	-0.2	0.0	0.0
Medicaid	Eligible for Medicaid (pp)	-1.2	-3.9	0.1	3.2
Disability	Disability, no ESRD (pp)	-0.5	-1.0	-0.4	1.7
Health status	HCC score	0.0	0.1	0.0	0.1
	Obesity (pp)	0.6	0.3	-0.1	1.1
	Diabetes (pp)	0.6	-1.2	0.9	-0.2
	Hypertension	0.4	-1.1	-0.2	2.2
	Dementia	0.0	0.1	1.7	2.9
	CHF	-0.5	2.3	-0.1	0.6
Utilization in the six months prior to the anchor hospitalization	ACH stay (pp)	-0.4	0.5	1.3	0.7
	IRF stay (pp)	-0.2	-0.7	0.4	-1.7
	SNF stay (pp)	-0.3	0.9	-0.5	2.4
	HH use (pp)	-0.7	0.9	-0.6	-1.1
	Any prior care (pp)	-1.2	0.8	0.7	-1.5

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 2018 (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 99%, 95%, or 90% 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, CMMI = Center for Medicare & Medicaid Innovation, DiD = difference-in-differences, 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.

Appendix J: Activities of Daily Living Sensitivity Analyses

Exhibit J-1: Change in days between beginning and ending patient assessments by discharge setting, always mandatory hospitals, PY1-3

First PAC setting	CJR		Control group		CJR		Control group		Net differences	p-value
	Baseline episodes (N)	Intervention episodes (N)	Baseline episodes (N)	Intervention episodes (N)	Baseline average	Intervention average	Baseline average	Intervention average		
IRF	16,054	17,615	5,373	7,589	11.0	11.7	10.8	11.1	0.3	p=0.13
SNF	45,311	49,728	22,908	23,934	24.5	21.7	23.4	22.4	-1.8	p<0.05
HHA	42,487	48,900	39,040	37,082	24.4	20.6	23.6	22.9	-3.1	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) and episodes initiated during or after April 2016 that ended by September 2018 (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. Estimates that are significant at the 99%, 95%, or 90% significance level are indicated by red, orange, and yellow shaded cells, respectively.

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.

Exhibit J-2: Risk-adjusted assessment-based difference-in-differences results for activities of daily living metrics by discharge setting, reported ADL results and sensitivity estimate, always mandatory hospitals, PY1-3

First PAC setting	Measure	Main / sensitivity analysis	CJR		Control group		DiD	DiD % of baseline	p-value
			Baseline risk-adjusted average	Intervention risk-adjusted average	Baseline risk-adjusted average	Intervention risk-adjusted average			
IRF	Average change in mobility score	Reported results	10.7	11.2	10.2	10.9	-0.2	-1.6%	p=0.47
		Controlling for days between assessments	10.7	11.2	10.2	10.9	-0.1	-1.4%	p=0.54
SNF	Improved transfer, locomotion on unit, and walking in corridor	Reported results	65.1%	66.1%	68.9%	70.1%	-0.1	-0.1%	p=0.96
		Controlling for days between assessments	64.5%	68.1%	68.2%	70.2%	1.6	2.6%	p=0.17
	Improved toilet use	Reported results	44.0%	41.0%	46.9%	47.6%	-3.7	-8.4%	p<0.10
		Controlling for days between assessments	43.5%	42.6%	46.3%	47.8%	-2.4	-5.4%	p=0.23
HHA	Improved ambulation/ locomotion	Reported results	89.5%	90.4%	89.6%	90.6%	-0.1	-0.1%	p=0.81
		Controlling for days between assessments	89.5%	90.5%	89.5%	90.5%	0.0	0.0%	p=0.94
	Improved bed transferring	Reported results	83.6%	84.8%	83.3%	85.1%	-0.5	-0.6%	p=0.57
		Controlling for days between assessments	83.6%	85.0%	83.3%	85.1%	-0.4	-0.5%	p=0.69
	Reduced pain	Reported results	74.8%	83.0%	74.8%	82.7%	0.3	0.4%	p=0.81
		Controlling for days between assessments	74.8%	83.2%	74.6%	82.6%	0.4	0.6%	p=0.72

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 2018 (intervention).

***Notes*:** The estimates in this exhibit are the result of a DiD model. DiD estimates that are significant at the 99%, 95%, or 90% significance level are indicated by red, orange, and yellow shaded cells, respectively.

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

Because the CJR model may impact both the length of post-acute care and ADL outcomes, the number of days between assessments is not included as a causal risk factor in the risk adjustment models for the main analysis.

One measure, self-reported moderate to extreme pain for patients first discharged to a SNF is not included in this exhibit because it is not risk-adjusted.

ADL = activities of daily living, 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.

Exhibit J-3: Change in transfer, locomotion on unit, and walking in corridor scores from admission to discharge, LEJR episodes at always mandatory hospitals first discharged to a skilled nursing facility, PY1-3

Change from admission to discharge	CJR		Control group		Net differences	Net differences % of baseline	p-value
	Baseline average (N=44,889)	Intervention average (N=22,719)	Baseline average (N=49,158)	Intervention average (N=23,798)			
Improvement	66.2%	65.6%	68.3%	68.3%	-0.5	-0.8%	p=0.66
No change	30.6%	30.3%	27.6%	26.5%	0.8	2.6%	p=0.54
No change and score of 0-6 at admission (best)	7.3%	7.6%	10.0%	11.4%	-1.2	-16.9%	p=0.19
No change and score of 7-12 at admission (worst)	23.2%	22.7%	17.6%	15.1%	2.0	8.7%	p=0.10
Decline	3.3%	4.1%	4.2%	5.2%	-0.2	-7.5%	p=0.49

Source: CJR evaluation team analysis of Medicare claims and enrollment data and MDS 3.0 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 2018 (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 99%, 95%, or 90% significance level are indicated by red, orange, and yellow shaded cells, respectively.

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

DiD = difference-in-differences, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, PY = performance year.

Exhibit J-4: Change in toilet use scores from admission to discharge, LEJR episodes at always mandatory hospitals first discharged to a skilled nursing facility, PY1-3

Change from admission to discharge	CJR		Control group		Net differences	Net differences % of baseline	p-value
	Baseline average (N=45,001)	Intervention average (N=22,833)	Baseline average (N=49,536)	Intervention average (N=23,864)			
Improvement	44.0%	38.2%	47.8%	46.8%	-4.8	-11.0%	p<0.05
No change	54.7%	60.4%	50.5%	51.1%	5.2	9.4%	p<0.05
No change and score of 0-2 at admission (best)	11.9%	12.7%	17.1%	18.1%	-0.2	-1.9%	p=0.82
No change and score of 3-4 at admission (worst)	42.7%	47.6%	33.4%	32.9%	5.4	12.6%	p<0.01
Decline	1.3%	1.4%	1.7%	2.1%	-0.3	-25.6%	p<0.10

Source: Lewin analysis of Medicare claims and enrollment data and MDS 3.0 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 2018 (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 99%, 95%, or 90% significance level are indicated by red, orange, and yellow shaded cells, respectively.

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

DiD = difference-in-differences, LEJR = lower extremity joint replacement, MDS = Minimum Data Set, PY = performance year.

Exhibit J-5: Change in ambulation/locomotion scores from admission to discharge, LEJR episodes at always mandatory hospitals first discharged to a home health agency, PY1-3

Change from admission to discharge	CJR		Control group		Net differences	Net differences % of baseline	p-value
	Baseline average (N=42,421)	Intervention average (N=39,010)	Baseline average (N=48,848)	Intervention average (N=37,069)			
Improvement	86.1%	93.6%	86.7%	93.9%	0.3	0.4%	p=0.70
No change	13.6%	6.2%	13.0%	5.9%	-0.3	-2.4%	p=0.70
No change and score of 0-2 at admission (best)	11.9%	4.5%	11.5%	4.1%	0.0	-0.2%	p=0.98
No change and score of 3-6 at admission (worst)	1.6%	1.7%	1.4%	1.9%	-0.3	-18.5%	p=0.23
Decline	0.4%	0.2%	0.4%	0.2%	0.0	-3.8%	p=0.78

Source: CJR evaluation team analysis of Medicare claims and enrollment data 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 2018 (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 99%, 95%, or 90% significance level are indicated by red, orange, and yellow shaded cells, respectively.

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

DiD = difference-in-differences, LEJR = lower extremity joint replacement, OASIS = Outcome and Assessment Information Set, PY = performance year.

Exhibit J-6: Change in bed transfer scores from admission to discharge, LEJR episodes at always mandatory hospitals first discharged to a home health agency, PY1-3

Change from admission to discharge	CJR		Control group		Net differences	Net differences % of baseline	p-value
	Baseline average (N=42,421)	Intervention average (N=39,010)	Baseline average (N=48,848)	Intervention average (N=37,069)			
Improvement	76.4%	89.6%	76.3%	90.1%	-0.7	-0.9%	p=0.60
No change	23.1%	10.2%	23.3%	9.7%	0.7	3.2%	p=0.54
No change and score of 0-2 at admission (best)	22.9%	10.0%	23.2%	9.5%	0.8	3.5%	p=0.50
No change and score of 3-5 at admission (worst)	0.2%	0.2%	0.1%	0.2%	-0.1	-33.0%	p=0.41
Decline	0.5%	0.2%	0.4%	0.2%	-0.1	-17.3%	p=0.35

Source: CJR evaluation team analysis of Medicare claims and enrollment data 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 2018 (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 99%, 95%, or 90% significance level are indicated by red, orange, and yellow shaded cells, respectively.

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

DiD = difference-in-differences, LEJR = lower extremity joint replacement, OASIS = Outcome and Assessment Information Set, PY = performance year.

Exhibit J-7: Change in pain interfering with activity scores from admission to discharge, LEJR episodes at always mandatory hospitals first discharged to a home health agency, PY1-3

Change from admission to discharge	CJR		Control group		Net differences	Net differences % of baseline	p-value
	Baseline average (N=42,421)	Intervention average (N=39,010)	Baseline average (N=48,848)	Intervention average (N=37,069)			
Improvement	71.1%	83.9%	72.1%	85.6%	-0.7	-1.0%	p=0.67
No change	26.3%	15.2%	25.8%	13.6%	1.2	4.4%	p=0.45
No change and score of 0-2 at admission (best)	2.8%	1.5%	2.8%	1.1%	0.4	14.7%	p=0.33
No change and score of 3-4 at admission (worst)	23.4%	13.6%	23.0%	12.5%	0.7	3.2%	p=0.60
Decline	2.6%	0.9%	2.1%	0.8%	-0.5	-17.2%	p=0.10

Source: CJR evaluation team analysis of Medicare claims and enrollment data 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 2018 (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 99%, 95%, or 90% significance level are indicated by red, orange, and yellow shaded cells, respectively.

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

DiD = difference-in-differences, LEJR = lower extremity joint replacement, OASIS = Outcome and Assessment Information Set, PY = performance year.

Appendix K: Factors Associated with Earning Reconciliation Payments

Exhibit K-1: Factors related to hospital level of average reconciliation payment per episode, always mandatory hospitals (n=807), PY1-PY3

Domain	Reference category	Measure	Adjusted beta coefficient	90% Confidence Interval	p-value
Intercept	na	Intercept	-$2,191	-$2,550 to -$1,831	<0.01
Performance year	PY1	PY2	$331	$233 to $430	<0.01
		PY3	-$160	-$322 to $2	0.10
Difference between hospital historical average payment and PY target price	Above median	Below median percent difference between hospital historical average payment and PY target price	$198	$13 to $384	<0.10
PY performance quality category	Below acceptable	Acceptable	$758	$516 to $1,000	<0.01
		Good	$951	$786 to $1,116	<0.01
		Excellent	$974	$776 to $1,172	<0.01
PY average quarterly volume	<15 episodes	15-49 episodes	$499	$322 to $675	<0.01
		50 or more episodes	$619	$407 to $830	<0.01
Ownership	For profit	Not for profit	$510	$278 to $743	<0.01
		Government	$260	-$66 to $586	0.19
Patient characteristics	Above median	Below median average HCC score for PY episodes	$258	$88 to $427	<0.05
	Above median	Below median percent of PY episodes age 80 years or older	$210	$60 to $360	<0.05
	Above median	Below median percent of PY episodes dual eligible	$260	$70 to $450	<0.05
	Above median	Below median percent of PY episodes with prior institutional stay	$323	$173 to $473	<0.01
	Above median	Below median percent of PY episodes non-Hispanic Black	$100	-$40 to $240	0.24
	Above median	Below median percent of PY episodes with disability, no ESRD	-$26	-$202 to $150	0.81
	Above median	Below median percent of PY episodes MS-DRG 470 elective	-$147	-$321 to $27	0.16
	Above median	Below median percent of PY episodes that are female	$32	-$106 to $171	0.70

Domain	Reference category	Measure	Adjusted beta coefficient	90% Confidence Interval	p-value
Hospital characteristics	Above median	Below median hospital DSH patient percentage	$243	$29 to $456	<0.10
	Above median	Below median bed count	$155	-$45 to $354	0.20
	No affiliation	Any affiliation with a medical school	$108	-$74 to $290	0.33
	Never participated	Ever participated in BPCI LEJR	$150	-$93 to $392	0.31
Census region	West	Northeast	$137	-$199 to $472	0.50
		South	$57	-$233 to $347	0.75
		Midwest	-$123	-$507 to $260	0.60

Source: Lewin's 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 always mandatory CJR participant hospital(s) in PY1 (episodes initiated during or after April 2016 that ended by December 2016), PY2 (episodes ending in 2017), and PY3 (episodes ending in 2018).

***Notes*:** PY1 and PY2 NPRA data are final, while the PY3 NPRA data are preliminary and will be finalized spring 2020.

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 at the 99%, 95%, or 90% significance levels, as indicated by red, orange, and 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. Always 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, 224; DSH patient percentage, 26.9%; difference between hospital historical average payments and PY target price, 3.0%; dual eligible, 14.3%; average HCC score, 1.59; MS-DRG 470 elective, 79.6%; 80 years or older, 26.6%; prior institutional stay, 5.3%; disability, 16.1%; non-Hispanic Black, 3.9%; and female, 66.1%.

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.