Supplemental material

Modifications of diagnosis-related group payment

for hospital care: a systematic review

Table of content

List of tables 2
List of abbreviations 3
Supplement 1. Electronic search strategies 5
Supplement 2. Study characteristics 8
Supplement 3. Characteristics of included payment reforms 57
Supplement 4. Outcome-based study findings 60
Supplement 5. Risk of bias 81
Supplement 6. Assessment of the relative importance of outcomes 82
Supplement 7. Certainty assessment of evidence 83
BPCI (Model 2) compared to standard traditional Medicare payment 83
CJR compared to standard traditional Medicare payment 84
BPCI-A compared to standard traditional Medicare payment 85
ACE Demonstration compared to standard Medicare DRG-based payment 86
Carrum Health Direct Payment program compared to usual payment arrangements 87
Patient choice reform compared to usual DRG-based hospital payment 88
Supplement 8. Summary of Findings (SoF) using plain language summaries 89

List of tables

Table S1. Search strategy for MEDLINE (via PubMed) 5
Table S2. Search strategy for Web of Science 5
Table S3. Search strategy for EconLit (via EBSCOhost) 6
Table S4. Search string in WHOLIS (WHO Library Database) 7
Table S5. Extracted study characteristics of Barnett et al. (2019) 8
Table S6. Extracted study characteristics of Chen et al. (2018) 10
Table S7. Extracted study characteristics of Chopra et al. (2023) 12
Table S8. Extracted study characteristics of Cook et al. (2020) 13
Table S9. Extracted study characteristics of Dummit et al. (2016) 14
Table S10. Extracted study characteristics of Finkelstein et al. (2018) 16
Table S11. Extracted study characteristics of Goude et al. (2021) 17
Table S12. Extracted study characteristics of Goude et al. (2022) 18
Table S13. Extracted study characteristics of Haas et al. (2019) 20
Table S14. Extracted study characteristics of Hammond et al. (2022) 21
Table S15. Extracted study characteristics of Joynt Maddox et al. (2018) 23
Table S16. Extracted study characteristics of Joynt Maddox et al. (2019) 24
Table S17. Extracted study characteristics of Joynt Maddox et al. (2021) 26
Table S18. Extracted study characteristics of Kallies et al. (2022) 28
Table S19. Extracted study characteristics of Kiani et al. (2024) 29
Table S20. Extracted study characteristics of Kim et al. (2020) 30
Table S21. Extracted study characteristics of Liao et al. (2022) 32
Table S22. Extracted study characteristics of Martin et al. (2018) 34
Table S23. Extracted study characteristics of Maughan et al. (2019) 35
Table S24. Extracted study characteristics of McNeely et al. (2020) 38
Table S25. Extracted study characteristics of Meyers et al. (2019) 41
Table S26. Extracted study characteristics of Murphy et al. (2019) 42
Table S27. Extracted study characteristics of Navathe et al. (2017) 43
Table S28. Extracted study characteristics of Navathe et al. (2018) 45
Table S29. Extracted study characteristics of Navathe et al. (2020) 46
Table S30. Extracted study characteristics of Navathe et al. (2021) 48
Table S31. Extracted study characteristics of Oddleifson et al. (2024) 49
Table S32. Extracted study characteristics of Rolnick et al. (2020) 51
Table S33. Extracted study characteristics of Shashikumar et al. (2023) 53
Table S34. Extracted study characteristics of Stern et al. (2024) 55
Table S35. Extracted study characteristics of Whaley et al. (2021) 56
Table S36. Description of bundled payment programs included in the review 57
Table S37. Description of DRG system refinements included in the review 59
Table S38. Description of combined patient choice and bundled payment reform included in the review 59
Table S39. BPCI (Model 2) compared to standard traditional Medicare payment 60
Table S40. CJR compared to standard traditional Medicare payment 68
Table S41. BPCI-A compared to standard traditional Medicare payment 74
Table S42. ACE model compared to standard Medicare DRG-based payment 78
Table S43. Bundled payment period (ACE + BPCI) compared to standard traditional Medicare payment 79
Table S44. Carrum Health Direct Payment Program compared to usual commercial payment arrangements 79
Table S45. Patient choice and bundled payment reform compared to standard Swedish DRG-based payment 80

Table S46. Introduction of MS-DRGs compared to previous DRG-classification..... 80
Table S47. Certainty assessment of evidence for each outcome: BPCI (Model 2) 83
Table S48. Certainty assessment of evidence for each outcome: CJR 84
Table S49. Certainty assessment of evidence for each outcome: BPCI-A 85
Table S50. Certainty assessment of evidence for each outcome: ACE Demonstration..... 86
Table S51. Certainty assessment of evidence for each outcome: Carrum Health direct payment program 87
Table S52. Certainty assessment of evidence for each outcome: Patient choice reform 88
Table S53. Summary of Findings with plain language summary statements for the BPCI (Model 2) initiative..... 89
Table S54. Summary of Findings with plain language summary statements for the CJR model..... 91
Table S55. Summary of Findings with plain language summary statements for the BPCI-A model... 93
Table S56. Summary of Findings with plain language summary statements for the ACE Demonstration 94
Table S57. Summary of Findings with plain language summary statements for the Carrum Health direct payment program..... 96
Table S58. Summary of Findings with plain language summary statements for the patient choice reform 97

List of abbreviations

ACE: Acute Care Episode

AE: Adverse Events

AMI: Acute Myocardial Infarction

ASA: American Society of Anesthesiologists

BHS: Baptist Health System

BPCI: Bundled Payments for Care Improvement

BPCI-A: Medicare Bundled Payments for Care Improvement Advanced

CABG: Coronary Artery Bypass Grafting

CBA: Controlled Before-After study

CC: Complication or coexisting Condition

CCW: the Chronic Conditions data Warehouse score (indicates the number of major coexisting chronic conditions; out of 27)

CJR: Comprehensive Care for Joint Replacement

CMS: Centers for Medicare and Medicaid Services

COPD: Chronic Obstructive Pulmonary Disease

DiD: Difference-in-Differences (aDiD = adjusted)

DDD: Difference-in-Difference-in-Differences (aDDD = adjusted)

DRG: Diagnosis Related Group

ED: Emergency Department

ESRD: End-Stage Renal Disease

FFS: Fee-for-Service

FY: Fiscal Year

GWTG: Get With The Guidelines

HCC: Hierarchical Condition Category

HHA: Home health agency

IPPS: Inpatient Prospective Payment System

IRF: Inpatient Rehabilitation Facility

ITS: Interrupted time series

LEJR: Lower Extremity Joint Replacement

LOS: Length of stay

LTCF: Long-term care facility

LTCH: Long-term care hospital

MA: Medicare Advantage

MCC: Major Complication or coexisting Condition

MSA: Metropolitan Statistical Area

MS-DRG: Medicare Severity-Diagnosis Related Group

NR: Not Reported

OR: Odds Ratio (aOR = adjusted)

PAC: Post-Acute Care

PCI: Percutaneous coronary Intervention

PGP: Physician Group Practice

pp: percentage points

PPS: Prospective Payment System

PT: Physical Therapy

PY: Performance year

SMSA: Standard Metropolitan Statistical Area

SNF: Skilled Nursing Facility

THA: Total Hip Arthroplasty

TKA: Total Knee Arthroplasty

TJR: Total Joint Replacement

TM: Traditional Medicare

Supplement 1. Electronic search strategies

Table S1. Search strategy for MEDLINE (via PubMed)

PubMed search up to 31 August 2024		
intervention/ comparison	#1	(DRG[tiab] OR Prospective Payment System[mesh] OR diagnosis related group[tiab] OR diagnosis related groups[tiab] OR bundled payment[tiab] OR patient classification system[tiab] OR patient classification systems[tiab] OR ((reimbursement[tiab] OR prospective payment[tiab] OR prospective payments[tiab] OR prospective pricing[tiab] OR capitation[tiab]) AND (case based[tiab] OR disease group based[tiab] OR diagnosis group[tiab] OR diagnosis groups[tiab] OR diagnostic group[tiab] OR diagnostic groups[tiab] OR diagnostic category[tiab] OR diagnostic categories[tiab] OR diagnostic cost group[tiab] OR diagnostic cost groups[tiab] OR clinical risk group[tiab] OR clinical risk groups[tiab])))
setting	#2	(hospital[mesh] OR hospital[tiab] OR hospitals[tiab] OR Hospitals[mesh] OR ((medical[tiab] OR inpatient[tiab] OR healthcare[tiab] OR health[tiab]) AND (clinic[tiab] OR center[tiab] OR centre[tiab] OR clinics[tiab] OR centers[tiab] OR centres[tiab] OR facility[tiab] OR facilities[tiab] OR institution[tiab] OR institutions[tiab])))
study design (NRS)	#3	(cohort[all] OR (control[all] AND study[all]) OR (control[tw] AND group*[tw]) OR epidemiologic studies[mh] OR program[tw] OR clinical trial[pt] OR comparative stud*[all] OR evaluation studies[all] OR statistics as topic[mh] OR survey*[tw] OR follow-up*[all] OR time factors[all] OR ci[tw]) NOT ((animals[mh:noexp] NOT humans[mh:noexp]) OR comment[pt] OR editorial[pt] OR review[pt] OR meta analysis[pt] OR case report[tw] OR consensus[mh] OR guideline[pt] OR history[sh])
study design (ITS)	#4	(Interrupted Time Series Analysis[Mesh] OR interrupted time series[tiab] OR interrupted times series[tiab] OR"segmented regression"[tiab:~3] OR integrat* moving average[tiab] OR slope change[tiab] OR"piecewise regression"[tiab:~3] OR"piece-wise regression"[tiab:~3] OR ((time series[tiab] OR timesseries[tiab]) AND ("pre post"[tiab:~5] OR"before after"[tiab:~5] OR quasi-experiment*[tiab] OR quasiexperiment*[tiab] OR natural experiment*[tiab] OR ARIMA[tiab] OR autoregress*[tiab] OR auto-regress*[tiab] OR segmented[tiab] OR segments[tiab] OR piecewise[tiab] OR piece-wise[tiab] OR interrupt*[tiab] OR implement*[tiab] OR guideline*[tiab] OR prescription*[tiab] OR prescrib*[tiab] OR stewardship[tiab] OR rates[tiab] OR intervention[tiab]))) NOT (Animals[Mesh] NOT Humans[Mesh])
study design	#5	#3 OR #4
total	**#6**	**#1 AND #2 AND #5**
filter: Publication date (up to 31 August 2024)		

Table S2. Search strategy for Web of Science

Web of Science search up to 31 August 2024		
Intervention/ Comparison	#1	(TS=DRG* OR (AK="Prospective Payment*" OR KP="Prospective Payment*") OR (AK="Prospective Reimbursement*" OR KP="Prospective Reimbursement*") OR (AK="Discretionary Adjustment*" OR KP="Discretionary Adjustment*") OR (AK="Prospective Pricing*" OR KP="Prospective Pricing*") OR (TS="diagnosis related group*") OR (TI="bundled payment*" OR AB="bundled payment*") OR (TI="patient classification system*" OR AB="patient classification system*") OR (((TI=reimbursement* OR AB=reimbursement*) OR (TI="prospective payment*" OR AB="prospective payment*") OR (TI="prospective pricing*" OR AB="prospective pricing*") OR (TI=capitation* OR AB=capitation*)) AND ((TI="case based" OR AB="case based") OR (TI="disease group based" OR AB="disease group based") OR (TI="diagnosis group" OR AB="diagnosis group") OR (TI="diagnosis groups" OR AB="diagnosis groups") OR (TI="diagnostic group" OR AB="diagnostic group") OR (TI="diagnostic groups" OR AB="diagnostic

		groups") OR (TI="diagnostic category" OR AB="diagnostic category") OR (TI="diagnostic categories" OR AB="diagnostic categories") OR (TI="diagnostic cost group" OR AB="diagnostic cost group") OR (TI="diagnostic cost groups" OR AB="diagnostic cost groups") OR (TI="clinical risk group" OR AB="clinical risk group") OR (TI="clinical risk groups" OR AB="clinical risk groups"))))
Setting	#2	(TS="hospital" OR TS="hospitals" OR (AK="care facility" OR KP="care facility") OR (AK="care facilities" OR KP="care facilities") OR (AK=hospice* OR KP=hospice*) OR (AK=surgicenter OR KP=surgicenter) OR (AK="mobile health unit*" OR KP="mobile health unit*") OR (AK="care center" OR KP="care center") OR (AK="care centers" OR KP="care centers") OR (AK="care centre" OR KP="care centre") OR (AK="care centres" OR KP="care centres") OR (((TI="medical" OR AB="medical") OR (TI="inpatient" OR AB="inpatient") OR (TI="healthcare" OR AB="healthcare") OR (TI="health" OR AB="health")) AND ((TI="clinic" OR AB="clinic") OR (TI="clinics" OR AB="clinics") OR (TI="center" OR AB="center") OR (TI="centre" OR AB="centre") OR (TI="centers" OR AB="centers") OR (TI="centres" OR AB="centres") OR (TI="facility" OR AB="facility") OR (TI="facilities" OR AB="facilities") OR (TI="institution" OR AB="institution") OR (TI="institutions" OR AB="institutions"))))
Study design (NRS)	#3	((ALL=cohort OR (ALL=control AND (ALL=study)) OR (TS=control AND TS=group*) OR (AK=epidemiology OR KP=epidemiology) OR TS="program" OR TS="clinical trial" OR ALL="comparative stud*" OR ALL="evaluation stud*" OR TS=survey* OR ALL=follow-up* OR ALL="time factors" OR TS="ci") NOT (((AK=animal OR KP=animal) NOT (AK=human OR KP=human)) OR DT=Note OR DT=Letter OR DT="Editorial Material" OR DT=Review OR DT="Review Article" OR (AK="Case Report*" OR KP="Case Report*")))
study design (ITS)	#4	(((TS="interrupted time* series" OR TS="ITS") OR (TI=("segmented" NEAR/3 "regression") OR AB=("segmented" NEAR/3 "regression")) OR (TI= "integrat* moving average" OR AB= "integrat* moving average") OR (TI= "slope change" OR AB= "slope change") OR (TI=("piecewise" NEAR/3 "regression") OR AB=("piecewise " NEAR/3 "regression")) OR (TI=("piece-wise" NEAR/3 "regression") OR AB=("piece-wise" NEAR/3 "regression")) OR (((TI="time series" OR AB="time series")) AND ((TI=("pre" NEAR/5 "post") OR AB=("pre" NEAR/5 "post")) OR (TI=("before" NEAR/5 "after") OR AB=("before" NEAR/5 "after")) OR (TI=quasi-experiment* OR AB=quasi-experiment*) OR (TI=quasiexperiment* OR AB=quasiexperiment*) OR (TI="natural experiment*" OR AB="natural experiment*") OR (TI="ARIMA" OR AB="ARIMA") OR (TI=autoregress* OR AB=autoregress*) OR (TI=auto-regress* OR AB=auto-regress*) OR (TI="segmented" OR AB="segmented") OR (TI="segments" OR AB="segments") OR (TI="piecewise" OR AB="piecewise") OR (TI="piece-wise" OR AB="piece-wise") OR (TI=interrupt* OR AB=interrupt*) OR (TI=implement* OR AB=implement*) OR (TI=guideline* OR AB=guideline*) OR (TI=prescription* OR AB=prescription*) OR (TI=prescrib* OR AB=prescrib*) OR (TI="stewardship" OR AB="stewardship") OR (TI="rates" OR AB="rates") OR (TI="intervention" OR AB="intervention")))) NOT ((AK=animal OR KP=animal) NOT (AK=human OR KP=human)))
study design	#5	#3 OR #4
Total	**#6**	**#1 AND #2 AND #5**
Date	**#7**	**#6 AND DOP=(1900-01-01/2024-08-31)**

Table S3. Search strategy for EconLit (via EBSCOhost)

EconLit search up to 31 August 2024		
Intervention/ Comparison	#1	(XX (DRG* OR "diagnosis related group*") OR (ZW ("prospective payment system" OR "prospective payment, cost reimbursement, fee for service, quality, cost reduction" OR "prospective payments") or (SU "prospective payment*")) OR (TI "bundled payment*" OR AB "bundled payment*") OR (TI "patient classification system*" OR AB "patient classification system*") OR (((TI reimbursement* OR AB reimbursement*) OR (TI "prospective payment*" OR AB "prospective payment*") OR (TI "prospective pricing*" OR AB "prospective pricing*") OR (TI capitation OR AB capitation))

		AND ((TI "case based" OR AB "case based") OR (TI disease group based OR AB disease group based) OR (TI diagnos* group* OR AB diagnos* group*) OR (TI diagnostic categor*" OR AB diagnostic categor*) OR (TI clinical risk group* OR AB clinical risk group*))))
Setting	#2	((ZW hospital*) OR (SU hospital*) OR (TI hospital*) OR (AB hospital*) OR (SU care facilit* OR hospice* OR "mobile health unit*" OR "care center*" OR "care centre*") OR (((TI medical OR AB medical) OR (TI inpatient OR AB inpatient) OR (TI health care OR AB healthcare) OR (TI health OR AB health)) AND ((TI clinic OR AB clini c) OR (TI center* OR AB center*) OR (TI centre* OR AB centre*) OR (TI clinics O R AB clinics) OR (TI facility OR AB facility) OR (TI facilities OR AB facilities) OR (TI institution OR AB institution) OR (TI institutions OR AB institutions))))
Study design (NRS)	#3	(XX cohort* OR (XX control* AND XX (study* OR studies)) OR TX (control AND group*) OR (SU epidemiolog* OR ZW epidemiolog*) OR TX program OR XX ("clinical trial*" OR "comparative stud*" OR "evaluation stud*") OR TX survey* OR ALL follow-up* OR ALL "time factor*" OR TX ci) NOT (((SU animal* OR ZW animal*) NOT (SU human* OR ZW human*)) OR SU review OR PT "book review" OR SU "meta analysis" OR TI "case report" OR AB "case report" OR ZW "case report" OR SU "case report"))
study design (ITS)	#4	(TX "interrupted time* series" OR TX ITS OR TI (segmented N3 regression) OR AB (segmented N3 regression) OR TI "integrat* moving average" OR AB "integrat* moving average" OR TI "slope change" OR AB "slope change" OR TI (piecewise N3 regression) OR AB (piecewise N3 regression) OR TI (piece-wise N3 regression) OR AB (piece-wise N3 regression) OR ((TI "time series" OR AB "time series" OR TI "timeseries" OR AB "timeseries") AND (TI (pre N5 post) OR AB (pre N5 post) OR TI (before N5 after) OR AB (before N5 after) OR TI quasi-experiment* OR AB quasi-experiment* OR TI quasiexperiment* OR AB quasiexperiment* OR TI "natural experiment*" OR AB "natural experiment*" OR TI ARIMA OR AB ARIMA OR TI autoregress* OR AB autoregress* OR TI auto-regress* OR AB auto-regress* OR TI segmented OR AB segmented OR TI segments OR AB segments OR TI piecewise OR AB piecewise OR TI piece-wise OR AB piece-wise OR TI interrupt* OR AB interrupt* OR (TI implement* OR AB implement*) OR (TI guideline* OR AB guideline*) OR (TI prescription* OR AB prescription*) OR (TI prescrib* OR AB prescrib*) OR (TI ste wardship OR AB stewardship) OR (TI rates OR AB rates) OR (TI intervention OR A B intervention)))) NOT ((SU animal* OR ZW animal*) NOT (SU human* OR ZW human*))
study design	#5	#3 OR #4
Total	**#6**	**#1 AND #2 AND #5**
Date		Limiters - Publication Date: -20240831

Table S4. Search string in WHOLIS (WHO Library Database)

WHOLIS search (iris.) up to 31 August 2024		
Intervention/ Comparison	#1	((dc.title:DRG OR dc.description.abstract:DRG OR dc.subject:DRG) OR (dc.title:"diagnosis group*"~1 OR dc.description.abstract:"diagnosis group*"~1 OR dc.subject:"diagnosis group*"~1) OR (dc.title:"diagnostic group*"~1 OR dc.description.abstract:"diagnostic group*"~1 OR dc.subject:"diagnostic group*"~1) OR (dc.title:"prospective payment*" OR dc.description.abstract:"prospective payment*" OR dc.subject:"prospective payment*") OR (dc.title:"case payment*" OR dc.description.abstract:"case payment*" OR dc.subject:"case payment*") OR (dc.title:"prospective pricing*" OR dc.description.abstract:"prospective pricing*" OR dc.subject:"prospective pricing*") OR (dc.title:"patient classification system*" OR dc.description.abstract:"patient classification system*" OR dc.subject:"patient classification system*") OR (dc.title:"reimbursement mechanism*" OR dc.description.abstract:"reimbursement mechanism*" OR dc.subject:"reimbursement mechanism*") OR (dc.title:"capitat*" OR dc.description.abstract:"capitat*" OR dc.subject:"capitat*") OR (dc.title:"disease group*" OR

		dc.description.abstract:"disease group*" OR dc.subject:"disease group*") OR (dc.title:"clinical risk group*" OR dc.description.abstract:"clinical risk group*" OR dc.subject:"clinical risk group*") OR (dc.title:"bundled payment*" OR dc.description.abstract:"bundled payment*" OR dc.subject:"bundled payment*") OR dc.subject.mesh:"Prospective Payment System" OR dc.subject.mesh:"Diagnosis Related Groups" OR dc.subject.mesh:"Reimbursement Mechanisms") AND dc.date.issued:[1900-01 TO 2024-08]

Supplement 2. Study characteristics

Table S5. Extracted study characteristics of Barnett et al. (2019)

Study ID	**Barnett (2019)**
Methods	**Study design:** CBA
	Allocation: Natural experiment (random assignment of metropolitan statistical areas [MSAs] to CJR participation)
	Selection of hospitals (intervention/control): Hospitals in MSAs randomly assigned to mandatory CJR participation/Hospitals in 121 MSAs not selected for CJR
	Data source: - Medicare claims and enrollment data - Medicare enrollment files for patients' characteristics
	Study period (before/after): Jan 2015–Dec 2015/Jul 2016–Sep 2017 (data extended through December 2017 to complete 90-day follow-up)
	Statistical analyses: - DiD with mixed-effects linear regression - Adjusted for patient and episode characteristics, hospital and MSA random effects - Sensitivity analyses included: hospital fixed effects with robust variance estimators, generalized linear models with log link, and logistic-regression models
Participants	**Country/region/setting:** USA/196 MSAs/acute and post-acute care
	Health system characteristics: Medicare payment system. In April 2016, Medicare introduced the mandatory Comprehensive Care for Joint Replacement (CJR) bundled-payment model for hip and knee replacement. Under CJR, usual Medicare fee-for-service payments to providers continued during the episode of care, followed by annual retrospective reconciliation of hospital-level episode spending against a hospital-specific benchmark.
	Number of included hospitals (intervention/control): 803/962

	Characteristics of hospitals (before, intervention/control): No. of beds (%): <100: 14.1/20.5 100–249: 34.2/34.5 >250: 51.7/45.0 Ownership (%): Government: 13.3/10.8 Non-profit: 63.3/63.1 Profit: 23.3/26.1 Teaching Status (%): Major: 19.2/18.4 Minor: 18.7/16.2 Non: 62.1/65.4 Location: NR Geographic region (%): North: 17.2/15.5 South: 20.9/26.5 Midwest: 30.1/38.8 West: 31.8/17.1 Puerto Rico: 0.0/2.1
	Characteristics of MSA (intervention/control): Counties (weighted N): 216/298 Mean county Population: 458,326/385,136 Medicare Advantage Penetration: 32.2/30.6 Mean SNF total beds in county: 2,078/1,754
	Patient condition/procedure: inpatient primary hip- or knee-replacement procedures (DRG 469 or DRG 470 at discharge)
	Number of patients included in the analysis (intervention/control): 280,161/377,278
	Characteristics of patients (before, intervention/control): Age (y), mean: 74,5/74,3 Male sex (%): 35.9/36.0 Race or ethnic group (%): Non-Hispanic white: 89.8/90.8 Non-Hispanic black: 5.9/ 5.6 Hispanic: 1.1/0.7 Asian: 1.0/0.6 Other: 2.2/2.2 Medicaid (%): 11.4/10.3 Disability (%): 15.7/15.8 Mean number of chronic conditions: 7.1/7.0
Intervention and Control	**Description of intervention:** - Mandatory CJR bundled-payment model for inpatient hip and knee replacement - Episode includes index hospitalization + 90 days post-discharge - Usual fee-for-service payments during the episode - Annual retrospective reconciliation against a hospital-specific benchmark - Shared savings below benchmark; penalties above target from 2017 - Savings/losses adjusted for performance on quality measures
	Control: Hospitals in MSAs not selected for mandatory CJR participation

Outcome(s)	**Description of outcome(s)** Primary: - Institutional Spending - Composite Rate of Complications - Proportion of Episodes in Top Quartile of Patient Risk Secondary: - Total Medicare Spending (20% sample) - Utilization measures (incl. the use of post-acute care [PAC] services according to facility type, length of stay in PAC facilities, and readmission or visits to an emergency department within 90 days after discharge) - 90-Day Mortality - complications (excluding hip fractures)
	Measurement time points: Quarterly data from Q1 2015 – Q3 2017 with exclusion of Q1–Q2 2016 (transition)
Notes	**Funding:** Commonwealth Fund, National Institute on Aging of the National Institutes of Health (NIH grants K23 AG058806 and P01 AG032952).
	Conflict of interest: Authors disclosed no relevant conflicts.
	Other comments: none

Table S6. Extracted study characteristics of Chen et al. (2018)

Study ID	**Chen (2018)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary enrollment in Medicare's Acute Care Episode [ACE] Demonstration Program)
	Selection of hospitals (intervention/control): hospitals participating in the ACE Demonstration/matched non-participating hospitals
	Data source: - Medicare MedPAR and Master Beneficiary Summary files - American Hospital Association Annual Survey for hospital characteristics
	Study period (before/after): 2007–2008/2011–2012
	Statistical analyses: - DiD using multivariate generalized linear models with log link, and multivariate logistic regression models for dichotomous outcomes) - Adjustment factors: age, gender, race, 29 Elixhauser comorbidities, hospital characteristics (bed size, ownership, teaching status, region), year of procedure - Stratified exact matching to identify control hospitals
Participants	**Country/region/setting:** USA/Texas, Oklahoma, Colorado, New Mexico/acute care
	Health system characteristics: Medicare payment system. In April 2009, the Centers for Medicare and Medicaid Services (CMS) launched the ACE Demonstration, introducing bundled payments covering Medicare Part A and Part B services provided during an inpatient stay.
	Number of included hospitals: - Cardiac: 4/9 - Orthopedic: 5/22

	Characteristics of cardiac hospitals (after matching, intervention/ control): No. of beds, %: <250: 39/9 250–499: 23/48 ≥500: 37/44 Ownership, %: For-Profit: 77/12 Nonprofit: 23/88 Nonteaching, %: 100/71 Urban, %: 100/100 Geographic region, %: Northeast: 0/0 West: 25/33 Midwest: 0/14 South: 75/53 **Characteristics of orthopedic hospitals (after matching, intervention/ control):** Bed size, %: <250: 17/29 250–499: 2/45 ≥500: 81/26 Ownership, %: For-Profit: 98/10 Nonprofit: 2/90 Nonteaching, %: 100/84 Urban, %: 100/88 Geographic region, %: Northeast: 0/16 West: 17/19 Midwest: 0/44 South: 83/21
	Patient condition/procedure: Cardiac: cardiac valve procedures and/or coronary artery bypass grafting - MS-DRGs: 216 to 221; 231 to 236 - ICD-9 codes (inclusions): 35.21 to 35.24; 36.10 to 36.19 Orthopedic: hip or knee replacement or revision - MS-DRGs: 461, 462, 466 to 470, 488, 489 - ICD-9 codes (inclusions): 81.51 to 81.55 - ICD-9 codes (exclusions): 81.56, 84.26 to 84.28
	Number of patients included in the analysis: - Cardiac: 5,017/9,617 - Orthopedic: 10,462/42,312
	Characteristics of cardiac patients (after matching, intervention/control): Age, y, mean (SD): 74 (6.1)/75 (6.4) Male, %: 65/65 White, %: 90/89 Comorbidities, %: 0: 10/10 1: 28/28 ≥2: 62/62 **Characteristics of orthopedic patients (before, intervention/control):** Age, y, mean (SD): 75 (7.1)/75 (7.2) Male, %: 34/37 White, %: 88/95 Comorbidities, %: 0: 7/11 1: 21/28 ≥2: 72/61

Intervention and Control	**Description of intervention:** - Participation in the ACE Demonstration, a bundled-payment model for 28 cardiovascular and 9 orthopedic inpatient services - Episode covers inpatient stay only - Bundled payment covering Medicare Part A and Part B services provided - Participating organizations received discounted payments and shared in savings if quality reporting and monitoring requirements were met - Optional physician gainsharing
	Control: Hospitals not participating in ACE
Outcome(s)	**Description of outcome(s)** **Primary:** - 30-day Total Episode Payment **Quality outcomes:** - 30-day mortality - 30-day serious complications (defined via validated claims-based algorithm) - 30-day all-cause readmissions
	Measurement time points: yearly from 2007– 2011
Notes	**Funding:** This work was supported by the National Institute on Aging (R01AG039434). Prior to completion of the study Andrew Ryan was supported by K01HS18546 from AHRQ, Terry Shih was supported by 5T32HL07612309 from the National Institutes of Health, and Lena Chen was supported by K08HS020671 from the Agency for Healthcare Research and Quality (AHRQ). Prior to the completion of the study, Lena Chen was also supported by R01HS024698 from AHRQ, 5P01AG019783 from the NIH/NIA, and 2059.11 from the Blue Cross Blue Shield of Michigan Foundation Investigator Initiated Research Program.
	Conflict of interest: Justin Dimick, MD, MPH has equity interest in ArborMetrix, a company that profiles hospital quality and episode cost efficiency. The company played no role in this work. Lena Chen currently receives support from the Department of Health and Human Services Office of the Assistant Secretary for Planning and Evaluation for her work there. That support played no role in this work.
	Other comments: none

Table S7. Extracted study characteristics of Chopra et al. (2023)

Study ID	**Chopra (2023)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary BPCI Model 2 participation)
	Selection of hospitals (intervention/control): hospitals participating in BPCI for LEJR, CABG, or colectomy/hospitals not participating in BPCI
	Data source: - 100% Medicare claims data (MedPAR, carrier, outpatient institutional, home health, DME) - hospital characteristics from AHA annual surveys
	Study period (before/after): Jan 2010–Sep 2013/Oct 2013–Nov 2016
	Statistical analyses: - DiD analyses using fixed effects models with inverse probability weights - Control group is derived using generalized synthetic control method - Adjustment factors: patient age, sex, race, comorbidity count, hospital size, teaching status, profit status, urbanicity, nurse-to-bed ratio, proportion of Medicare/Medicaid days
Participants	**Country/region/setting:** USA/acute and post-acute care hospitals excluding Maryland, Puerto Rico, and other US territory hospitals
	Health system characteristics: Medicare payment system. In 2013, CMS established the Bundled Payments for Care Improvement (BPCI) program as a voluntary, episode-based payment model.

	Number of included hospitals (intervention/control): 312/1,977
	Characteristics of hospitals (intervention/control): No. of beds, n (%): <200: 92,285 (18.5)/ 673,058 (32.4) 200–349: 142,645 (28.6)/ 598,017 (28.8) 350–499: 100,491 (20.2)/ 355,924 (17.1) ≥500: 163,278 (32.7)/449,907 (21.7) Profit Status, n (%): For-profit: 58,709 (11.8)/ 347,770 (16.7) Non-profit: 424,693 (85.2)/1,508,303 (72.6) Other: 15,297 (3.1)/220,833 (10.7) Teaching status, n (%): Non-teaching: 353,693 (70.92)/1,790,063 (86.19) Teaching: 145,006 (29.08)/ 286,843 (13.81) Urban, n (%):497,837 (99.8)/2,052,729 (98.8) Region: NR
	Patient condition/procedure: LEJR, CABG, and colectomy
	Number of patients included in the analysis (intervention/control): 498,699/2,076,906 (episodes) (454,369 cases for LEJR; 107,307 cases for CABG; and 73,717 cases for colectomy)
	Characteristics of patients (before, intervention/control): Age, mean (SD): 74.6 (6.5)/74.6 (6.4) Female, n (%): 300,495 (60.3)/1,241,001 (59.8) Race, (%): Non-White: 49,880 (10.0)/ 180,354 (8.7) White: 448,819 (90.0)/1,896,552 (91.3) Count of Elixhauser comorbidities, mean (SD): 2.33 (1.58)/2.30 (1.58)
Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI Model 2, a retrospective bundled payment program covering all services (excluding drugs) within 30-, 60-, or 90-days post-discharge - study includes only 30-day episodes for LEJR, CABG, and colectomy
	Control: Hospitals not participating in BPCI Model 2
Outcome(s)	**Description of outcome(s)** Primary outcomes: - 30-day total episode spending (price-standardized - 30-day institutional PAC spending Secondary outcomes: - 30-day readmission spending - Index hospitalization spending - Outlier payment spending - Physician spending
	Measurement time points: quarterly data from Q1 2010 – Q4 2016
Notes	**Funding:** Supported by NIH, AHRQ, Blue Cross Blue Shield of Michigan Foundation, and the University of Michigan
	Conflict of interest: J.B. Dimick is a cofounder of ArborMetrix, Inc., which had no role in the study
	Other comments: - Participation in BPCI Model 2 was voluntary - Study included only hospitals with uninterrupted participation over 28 quarters

Table S8. Extracted study characteristics of Cook et al. (2020)

Study ID	**Cook (2020)**
Methods	**Study design:** CBA
	Allocation: natural experiment (Medicare DRG restructuring)

	Selection of hospitals (intervention/control): Hospitals treating patients assigned to DRG groups that were expanded in the 2007 reform / hospitals treating patients in DRG groups that were not expanded
	Data source: - Healthcare Cost and Utilization Project's (HCUP) national inpatient sample (NIS) from the years 2005–2010
	Study period (before/after): 2005–2007/2008–2010
	Statistical analyses: - DiD models with hospital, year, and DRG fixed effects - Adjustment factors: age, sex, race
Participants	**Country/region/setting:** USA/nationwide/community hospitals
	Health system characteristics: Medicare hospital payment under the Prospective Payment System (PPS), with fixed prospective payments based on DRGs and DRG-specific relative weights.
	Number of included hospitals (intervention/control): NR (≈1,000 hospitals in total)
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedure: all
	Number of patients included in the analysis: 26,904,427 total discharges
	Characteristics of patients (before, intervention/control): NR
Intervention and Control	**Description of intervention:** - Introduction of the Medicare Severity Diagnosis-Related Group (MS-DRG) classification system, effective October 2007 (Version 25) - Expansion of the previous system of up to 540 DRGs to 745 MS-DRGs - 335 base MS-DRGs, with selected groups further subdivided according to the presence of complications or comorbidities (CCs) or major complications or comorbidities (MCCs) - Reimbursement differentiated by patient severity within MS-DRGs, with lower payments for less severe cases and higher payments for more severe cases
	Control: Control group patients are assigned to DRGs that were not affected by the MS-DRG expansion and thus continue to follow the original DRG system, which does not differentiate by severity (i.e., does not account for complications or comorbidities).
Outcome(s)	**Description of outcome(s):** DRG weight (as a proxy for upcoding)
	Measurement time points: annually from 2005 to 2010
Notes	**Funding:** None
	Conflict of interest: NR
	Other comments: - No descriptive statistics on patient or hospital characteristics - No clinical validation of whether upcoding reflects real changes in care or just coding behavior

Table S9. Extracted study characteristics of Dummit et al. (2016)

Study ID	**Dummit (2016)**
Methods	**Study design:** CBA
	Allocation: natural experiment (participation in BPCI Model 2)
	Selection of hospitals (intervention/control): hospitals voluntarily participating in BPCI Model 2 for LEJR/matched non-participating hospitals
	Data source: Medicare Part A and B claims data; beneficiary survey
	Study period (before/after): Oct 2011–Sep 2012 / Oct 2013–Jun 2015

	Statistical analyses: - DiD approach using multivariate regression models - propensity score matching on volume, spending, case mix, and region - Adjustment for patient demographics, comorbidities, hospital characteristics - Sensitivity analyses conducted for robustness
Participants	**Country/region/setting:** USA /multiple U.S. regions/acute and post-acute care
	Health system characteristics: - Medicare fee-for-service (FFS) payment system; hospitals were paid under Medicare's Inpatient Prospective Payment System (IPPS) - BPCI Model 2 introduced retrospective bundled payments with financial reconciliation over a 90-day episode
	Number of included hospitals (intervention/control): 176/841
	Characteristics of hospitals[1] (after matching, intervention/control): No. of beds, mean (SD): 299 (213)/314 (267) Ownership, n (%): Nonprofit: 130 (73.9)/692 (75.6) For-profit: 40 (22.7)/186 (20.3) Government: 6 (3.4)/37 (4.1) Teaching, n (%): 85 (48.3)/459 (50.2) Location, n (%) Urban: 168 (95.5)/870 (95.0) Rural: 8 (4.6)/45 (5.0) Region: NR
	Patient condition/procedure: LEJR (MS-DRG: 469 or 470)
	Number of patients included in the analysis: 17,641/61,557 (LEJR episodes)
	Characteristics of patients (baseline, intervention/control): Age, y, n (%): 65–79: 18,195 (61,8)/18,220 (61.9) ≥80: 8,702 (29.6)/8,786 (29.8) Female, n (%): 19,764 (67.1)/19,466 (66.1) Race: NR Medicaid, n (%): 3,744 (12.7)/3,733 (12.7) Disabled with no end-stage renal disease, n (%): 3,086 (10.5)/2,990 (10.2) Hierarchical condition categories, mean (SD): 1.2 (1.4)/1.2 (1.4)
Intervention and Control	**Description of intervention:** - Voluntary participation in the BPCI initiative Model 2: retrospective bundled payment model covering hospitalization and all Medicare-covered services for 90 days post-discharge. - Total episode payments were retrospectively reconciled against a target price based on discounted historical spending, with shared savings below the target and potential repayment to CMS above the target.
	Control: Non-BPCI hospitals paid under Medicare's Inpatient Prospective Payment System (IPPS)
Outcome(s)	**Description of outcome(s)** Primary outcomes: - Total Medicare payments during index hospitalization and 90-d post-discharge period Secondary outcomes: - Patients discharged to institutional PAC (SNF, IRF, LTCH) - Total inpatient LOS during index hospitalization - Total institutional PAC days in 90-d post-discharge period - 30- and 90-day emergency department (ED) use - Mortality rate by post-discharge period (30-d and 90-d) - and survey-reported functional outcomes and satisfaction

[1] Baseline characteristics refer to all matched comparison hospitals (n = 915), whereas the outcome analysis included only hospitals that contributed eligible episodes during the intervention period (n = 841)

	Measurement time points: - Claims: pre/post comparison - Survey: conducted at 3 time points (Oct 2014, Feb 2015, May 2015)
Notes	**Funding:** Funded by CMS
	Conflict of interest: Several authors reported owning stock in UnitedHealth Group and were employed by The Lewin Group, which was contracted by CMS to evaluate the BPCI initiative. The Lewin Group, a subsidiary of Optum (which is owned by UnitedHealth Group), operates with analytic and editorial independence. 2 authors conducted the work at Abt Associates, a subcontractor to The Lewin Group under a CMS contract. Other authors were CMS employees. No other disclosures were reported.
	Other comments: - Baseline characteristics of included hospitals refer to all matched comparison hospitals (n = 915). However, the outcome analysis includes only those hospitals that contributed eligible episodes during the intervention period (n = 841) - Selection bias possible due to voluntary program participation - Analysis limited to hospitals participating continuously in BPCI during the post-period

Table S10. Extracted study characteristics of Finkelstein et al. (2018)

Study ID	**Finkelstein (2018)**
Methods	**Study design:** CBA
	Allocation: natural experiment (mandatory CJR bundled payment model, randomization of MSAs by CMS)
	Selection of hospitals (intervention/control): acute care hospitals in MSAs randomized to the CJR model and meeting inclusion criteria (not in BPCI, under IPPS)/hospitals in MSAs not assigned to CJR
	Data source: - Medicare FFS claims data - Hospital Compare data - CMS reconciliation payment data
	Study period (baseline/after): Apr 1, 2014–Sep 15, 2014 / Apr 1, 2016–Sep 15, 2016 (data extended through December 2016 to complete 90-day follow-up)
	Statistical analyses: - Instrumental variable regression analysis - MSAs were divided into 8 strata and the regression analysis included strata fixed-effects - Sensitivity: intention-to-treat, regressions without lag controls
Participants	**Country/region/setting:** USA/196 MSAs/acute and post-acute care
	Health system characteristics: Traditional Medicare was transitioning from fee-for-service (FFS) reimbursement toward alternative payment models; by 2016, mandatory participation for eligible hospitals in CJR regions.
	Number of included hospitals (intervention/control): 67 MSAs/121 MSAs (hospital count: 1,633 total)
	Characteristics of hospitals (all eligible hospitals): Hospital beds, mean (SD): 289.1 For-profit hospitals, %, mean (SD): 25.3 (25.8) Teaching hospitals, %, mean (SD): 9.3 (16.1) Location: NR Region: NR **Characteristics of eligible MSAs:** Acute care hospitals, mean (SD): 14.8 (16.0) Institutional PAC facilities, mean (SD): 50.9 (53.3) Home health agencies, mean (SD): 57.8 (117.1)
	Patient condition/procedure: LEJR (MS-DRGs 469 and 470)

	Number of patients included in the analysis: 131,285 procedures among 130,343 patients
	Characteristics of patients (all eligible LEJR patients): Age, y, mean (SD): 72.5 (0.91) Female, %, mean (SD): 64.8 (0.03) White, %, mean (SD): 90.0 (0.07) Medicaid, %, mean (SD): 11.7 (0.06) Elixhauser Comorbidity Index, mean (SD): 2.4 (0.29)
Intervention and Control	**Description of intervention:** - Mandatory participation in the CJR bundled-payment model for eligible hospitals in selected MSAs - Episode includes index hospitalization + 90 days post-discharge - Medicare FFS payments continued during the episode, followed by annual retrospective reconciliation against a CMS target price based on historical hospital and regional spending - Hospitals received shared savings for spending below the target if minimum quality standards were met and repaid CMS for spending above the target, subject to stop-gain/stop-loss limits (+5% stop-gain and no downside risk in year 1; both scheduled to reach 20% of the target price by year 5) - Introduced in April 2016 and designed as a 5-year program
	Control: Hospitals in control MSAs not selected for mandatory CJR participation
Outcome(s)	**Description of outcome(s)** Primary: share of LEJR admissions discharged to institutional PAC Secondary: - number of days in institutional PAC (SNF, LTCH and IRF) - discharges to other locations - Medicare FFS spending during the episode - Medicare net spending during the episode (incl. any reconciliation payments) - LEJR patient volume - patient comorbidity severity (Elixhauser Comorbidity Index) several quality measures: - Composite Quality Score (0–18 scale) - 90-day emergency department (ED) visit rate - 90-day all-cause readmission rate - 90-day complication rate for total hip and total knee arthroplasty
Notes	**Funding:** J-PAL North America; NIH/National Institute on Aging (P01AG019783-15)
	Conflict of interest: Dr. Skinner is an investor in a clinical pathway startup and consultant to Sutter Health; no other disclosures
	Other comments: None

Table S11. Extracted study characteristics of Goude et al. (2021)

Study ID	**Goude (2021)**
Methods	**Study design:** CBA
	Allocation: natural experiment (regional payment reform in Stockholm)
	Selection of hospitals (intervention/control): Hospitals in Region Stockholm/hospitals in other Swedish regions
	Data source: - Swedish Hip Arthroplasty Register (SHAR) - National Patient Register - Swedish Longitudinal Integrated Database for Health Insurance and Labour Market Studies
	Study period (before/after): 2008/2009–2012

	Statistical analyses: - DiD analysis with weighted regression model - Adjusted for: age, sex, ASA grade, BMI, Charnley classification, comorbidity (Elixhauser index), surgical approach, education, civil status, baseline PROMs - Sensitivity analyses using inverse probability weighting with propensity scores
Participants	**Country/region/setting:** Sweden/Stockholm vs. other regions/orthopedic care setting
	Health system characteristics: - Tax-funded and decentralized system - DRG-based reimbursement in all regions - Stockholm introduced bundled payment, patient choice, and new provider entry in 2009
	Number of included hospitals (intervention/control): NR
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedure: elective primary total hip replacements
	Number of patients included in the analysis: 6,563/30,064
	Characteristics of patients (before, intervention/control): Age, y, mean: 68.0/68.2 Female, mean: 0.66/0.57 Race or ethnicity: NR Elixhauser Index (%): 0: 76/61 1: 18/25 2: 4/10 3: 2/4 Education (%): Low: 27/40 Middle: 39/40 High: 34/20
Intervention and Control	**Description of intervention:** - Reform included patient choice of provider, accreditation-based market entry, and bundled payment - Bundled payment model: Lump sum covered full care chain including revisions up to 2 years post-op - Performance-based payment of a few percentages awarded based on 1-year PROM improvement
	Control: Standard care under DRG-based reimbursement model
Outcome(s)	**Description of outcome(s)** Primary outcomes: - Change in EQ-5D index - Change in Health Status (VAS) - Reduction in Pain (VAS) - Satisfaction with surgery outcome (VAS)
	Measurement time points: Baseline (pre-surgery), 1-year, and 6-year post-surgery Compared between pre-reform (2008) and post-reform (2009–2012)
Notes	**Funding:** Unrestricted grant from Region Stockholm
	Conflict of interest: 2 authors employed or contracted by Region Stockholm
	Other comments: The reform is limited to low-risk profile patients (ASA-grade 1 or 2)

Table S12. Extracted study characteristics of Goude et al. (2022)

Study ID	**Goude (2022)**
Methods	**Study design:** CBA
	Allocation: natural experiment (regional payment reform in Stockholm)

	Selection of hospitals (intervention/control): Hospitals in Region Stockholm/hospitals in other Swedish regions
	Data source: - Swedish Hip Arthroplasty Register (SHAR) - National Patient Register - Swedish Longitudinal Integrated Database for Health Insurance and Labour Market Studies
	Study period (before/after): 2005–2008/2009–2012
	Statistical analyses: - DiD analyses using linear regression with generalized estimating equations (GEE) to account for clustering of patients within hospitals - Adjustments: age, sex, comorbidity (Elixhauser Index), education, civil status
Participants	**Country/region/setting:** Sweden/Stockholm vs. other regions/orthopedic departments
	Health system characteristics: - The Swedish healthcare system is decentralized and primarily financed through regional income taxes - Orthopedic providers are reimbursed using a DRG-based model, applied either as activity-based funding or as the basis for a global budget - In January 2009, Region Stockholm implemented a patient choice reform and introduced bundled payment as the new reimbursement model for elective hip replacement surgery
	Patient condition/procedure: elective total hip and knee replacement surgery
	Number of included hospitals (intervention/control): NR
	Characteristics of hospitals (before, intervention/control): No. of beds: NR Hospital type, n (%) University: 859 (13.5)/2,379 (7.3) Central: 2,272 (35.6)/11,853 (36.1) Local: 1,644 (25.8)/18,249 (55.6) Private: 1,605 (25.2)/318 (1.0) Location: NR Region: NR
	Number of patients included in the analysis: 14,570/70,705 (surgeries)
	Characteristics of patients (before, intervention/control): Female, n (%):4028 (63.1)/18,653 (56.9) Age, n (%): 18–54: 561 (8.8)/ 2735 (8.3) 55–64: 1663 (26.1)/ 7884 (24) 65–74: 2181 (34.2)/ 11,804 (36) 75–84: 1656 (26)/ 9038 (27.6) 85+: 319 (5.0)/ 1338 (4.1) Race or ethnicity: NR Elixhauser Comorbidity Index, n (%) 0: 4222 (66.2)/20,482 (62.4) 1: 1274 (20.0)/ 7439 (22.7) 2: 526 (8.2)/ 3246 (9.9) 3+: 358 (5.6)/ 1632 (5.0) Education level, n (%): Low: 1862 (29.2)/14,708 (44.8) Middle: 2610 (40.9)/12,305 (37.5) High: 1908 (29.9)/ 5786 (17.6)
Intervention and Control	**Description of intervention:** - The reform consists of three central elements: 1) free choice of provider for patients, 2) free entry of new providers and 3) bundled payment model for hip replacement surgery (fixed payment per episode for care chain, excluding rehab and primary care) - A complication guarantee held providers financially responsible for adverse events within 2 years

	Control: Other Swedish regions continuing pre-existing DRG-based hospital reimbursement system
Outcome(s)	**Description of outcome(s)** - LOS of surgical admission (OP_LOS): (discharge − admission + 1) - Adverse Events (AE_90D): Death, reoperation, and severe complications within 90 days post-surgery - Patient satisfaction (SATISF): VAS scale (ranging from 0–100), score ≤ 40 = satisfied, measured 1 year post-op (available from 2007 only)
	Measurement time points: - OP_LOS and AE_90D: yearly data from 2005 to 2012 - SATISF: yearly data from 2007 to 2012
Notes	**Funding:** Unrestricted grant from Region Stockholm
	Conflict of interest: 2 authors employed or contracted by Region Stockholm
	Other comments: The analysis was limited to low- risk profile patients as these were the only ones covered by the reform

Table S13. Extracted study characteristics of Haas et al. (2019)

Study ID	**Haas (2019)**
Methods	**Study design:** CBA
	Allocation: natural experiment (Randomization of MSAs by CMS)
	Selection of hospitals (intervention/control): All hospitals in MSAs randomly assigned to participate in CJR (excluding BPCI participants)/ all hospitals in MSAs not assigned to CJR
	Data source: 100% Medicare Part A and 5% Part B fee-for-service claims data
	Study period (before/after): Jan 2014–Jul 2015/Apr 2016–Dec 2017 (data extended through March 2018 to complete 90-day follow-up)
	Statistical analyses: - Regression analysis using DiD models - Covariates: age, sex, CMS-HCC score, hip fracture, quarter-year indicators - Propensity score matching
Participants	**Country/region/setting:** USA/CJR-selected and control MSAs/acute and post-acute care
	Health system characteristics: Medicare FFS payment system. In April 2016, CMS introduced the mandatory CJR bundled-payment model, covering services from the index hospitalization through 90 days post-discharge.
	Patient condition/procedure: LEJR (typically knee and hip replacements; MS-DRG 469 or 470)
	Number of included hospitals (intervention/control): 815/955
	Characteristics of hospitals (intervention/control): No. of beds, mean: 313/ 313 Ownership: NR Teaching status, n (%): 367 (53.7)/375 (51.7) Location or region: NR
	Number of patients included in the analysis: 157,828/180,594 (episodes)
	Characteristics of patients (all eligible LEJR patients): Age, y, mean (SD): 72.8 (8.9)/72.6 (8.8) Female, No. (%) 101 641 (64.4) 115 580 (64.0) Race: NR CMS-HCC risk score, mean (SD): 1.09 (1.03)/1.08 (1.02)

Intervention and Control	**Description of intervention:** - Mandatory participation in the CJR bundled-payment model - Episode includes index hospitalization + 90 days post-discharge - Medicare FFS payments continued during the episode, followed by retrospective reconciliation against a CMS target price based on a combination of hospital historical and regional spending - Hospitals meeting CMS quality thresholds could receive shared savings for spending below the target; upside risk was capped at 5% of the target price in 2016 and 10% in 2017, while downside risk was 0% in 2016 and up to 5% in 2017 - CJR introduced in April 2016; this study evaluated the first 2 years of the program
	Control: Hospitals in control areas not selected for mandatory CJR participation
Outcome(s)	**Description of outcome(s)** Primary: - Medicare Part A spending per episode (total, index admission, and 90-day PAC) - Medicare Part B spending (in 5% subsample) - total 90-day spending (in 5% subsample) Quality indicators: - Hospitalization LOS - 90-day readmission rate - 90-day complication rate (THA/TKA) - 30-day and 90-day mortality rate Use and selection outcomes: - Volume of episodes (per 1,000 beneficiaries) - Patient selection (age, sex, HCC risk score, hip fracture episodes)
	Measurement time points: - Medicare spending visualized monthly from Jan 2014–Mar 2018 (excluding transition period) - All other outcomes aggregated pre/post
Notes	**Funding:** The Office of the Director, National Institutes of Health (1DP5OD024564)
	Conflict of interest: Mr Haas and Ms Zhang reported personal fees and stock ownership from Avant-garde Health outside the submitted work and have patent 15/973,081 pending. Dr Kaplan reported personal fees and nonfinancial support from Avant-garde Health and personal fees and nonfinancial support from Medtronic outside the submitted work. Dr Song reported grants from the National Institutes of Health during the conduct of the study. No other disclosures were reported.
	Other comments: none

Table S14. Extracted study characteristics of Hammond et al. (2022)

Study ID	**Hammond (2022)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary participation in BPCI-A)
	Selection of hospitals (intervention/control): BPCI-A participating hospitals/non-BPCI-A program hospitals
	Data source: - Medicare claims data - American Hospital Association data and Area Health Resources File data to characterize hospitals and geographic markets
	Study period (before/after): Jan 2017–Sep 2018/Oct 2018–Sep 2019 (data extended through Dec 2019 to complete 90-day follow-up)
	Statistical analyses: - Segmented regression model with control group - Adjustment factors: hospital fixed effects, patient demographic, clinical and socioeconomic characteristics, community characteristics

Participants	**Country/region/setting:** USA/nationwide/~10% of U.S. hospitals
	Health system characteristics: - Medicare payment system increasingly incorporating value-based and alternative payment models. In 2018, Medicare introduced BPCI-A as an episode-based payment model for hospitalized patients.
	Number of included hospitals (intervention/control): 826/2016
	Characteristics of hospitals (intervention white/black vs comparison white/black): Hospital size, %: Small: 4.9/3.4 vs 15.6/9.8 Medium: 60.2/56.2 vs 55.6/48.7 Large: 34.7/40.3 vs 28.7/41.5 Ownership, %: For profit: 23.3/23.8 vs 11.7/12 Not for profit: 71/ 68.3 vs 75/70.2 Public: 5.6/7.9 vs 13.3/17.8 Teaching status, %: Major teaching: 18.1/ 26 vs 13.5/25 Minor teaching: 40.2/ 37.3 vs 33.5/32.1 Nonteaching: 41.6/ 36.5 vs 53/42.9 Rural location, %: 1/0.7 vs 3.4/3.5 Region, %: Northeast: 22.0/18.6 vs 19.9/12.9 Midwest: 25.3/24.7 vs 23.9/13.5 South: 37.9/46.5 vs 41.9/67.3 West: 14.6/10.0 vs 14.3/6.3
	Patient condition/procedure: any of the 29 inpatient conditions covered under BPCI-A
	Number of patients included in the analysis: 1,461,222/5,229,114 (episodes)
	Characteristics of patients (intervention white/black vs comparison white/black): Age, y (%): <65: 9.2/23.7 vs 10.5/23.9 65–80: 46.2/45.8 vs 49/47 ≥80: 44.7/30.5 vs 40.5/29.1 Female (%): 56.0/58.2 vs 56.2/57.4 Race or ethnicity: NR Medicaid (%): 21/53.1 vs 22.8/52.4 Disabled, %: 21.6/43.9 vs 23.9/44.3 Total No. of CCWs: 6.21/6.60 vs 5.78/6.08
Intervention and Control	**Description of intervention:** - Participation in the BPCI-A program starting October 2018 - Episode: hospitalization-triggered 90-day episode of care - Episode costs retrospectively reconciled against target prices - Hospitals retained savings below the target and repaid Medicare when costs exceeded the target
	Control: Hospitals not participating in BPCI-A during study period, paid under the inpatient prospective payment system
Outcome(s)	**Description of outcome(s)** Primary outcomes (measured quarterly): - 90-day readmission rate - 90-day mortality rate - Healthy days at home - Proportion of Black patients of each outcome mentioned above (proxy for access)
	Measurement time points: quarterly from Q1 2017 – Q3 2019
Notes	**Funding:** NR

	Conflict of interest: Dr Orav reported receiving grants from the National Institutes of Health (NIH) for salary support during the conduct of the study. Dr Epstein reported receiving grants from the National Heart, Lung, and Blood Institute (NHLBI) during the conduct of the study; grants from the National Institute on Aging outside the submitted work. Dr Joynt Maddox reported receiving grants from the NIH/NHLBI during the conduct of the study; grants from National Institute on Aging, grants from the NIH/National Institute of Nursing Research, a contract from Humana Research, and personal fees from Centene Health Policy Advisory Council outside the submitted work. No other disclosures were reported.
	Other comments: None

Table S15. Extracted study characteristics of Joynt Maddox et al. (2018)

Study ID	**Joynt Maddox (2018)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary participation in BPCI)
	Selection of hospitals (intervention/control): BPCI participating hospitals for one or more of five common medical bundles (CHF, pneumonia, COPD, sepsis, and AMI)/matched non participating hospitals
	Data source: - Medicare claims data - data from the American Hospital Association file and the Area Resource File
	Study period (before/after): - Baseline period: 9 to 3 months before hospital's start date (6-month period) - Intervention period: 3 to 9 months after hospital's start date, according to its enrollment date - Study uses data from Jan 2013–Sep 2015 (data extended through Dec 2019 to complete 90-day follow-up)
	Statistical analyses: - DiD with marginal models using generalized estimating equations (GENMOD) - Adjustment factors: DRGs, Chronic Conditions Warehouse scores, hospital and market characteristics
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare FFS payment system. In 2013, the Center for Medicare and Medicaid Innovation (CMMI) launched the BPCI initiative
	Number of included hospitals (intervention/matched control): 492/898
	Characteristics of hospitals (intervention/matched control): No. of beds, mean: 322/315 Ownership (%): Nonprofit: 74.6/76.5 Public: 2.4/3.6 For-profit: 23.0/19.9 Teaching (%): 47.2/48.7 Urban (%): 100/100 Geographic region (%): Northeast: 28.9/30.2 Midwest: 14.6/16.8 South: 27.2/24.9 West: 29.3/28.1
	Patient condition/procedure: congestive heart failure (CHF; MS-DRGs 291-293), pneumonia (MS-DRGs 177-179), chronic obstructive pulmonary disease (COPD; MS-DRGs 190-192, 202, 203), sepsis (MS-DRGs 870-872), and acute myocardial infarction (AMI; 280-282)
	Number of patients included in the analysis: 39,442/70,780 (episodes)

	Characteristics of patients (after matching, baseline, intervention/control: Age (%): ≤64 y.: 8.7/9.7 65-79 y.: 35.8/35.7 ≥80 y.: 55.5/54.6 Female (%): 55.1/53.9 Race or ethnic group (%): White: 83.9/83.6 Black: 10.7/10.9 Hispanic: 2.3/1.9 Unknown or other: 3.1/3.6 Medicaid beneficiary (%): 25.1/24.5 Disabled without ESRD (%): 21.3/23.1 Mean Chronic Conditions Warehouse score: 8.91/8.94
Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI Model 2 - Episode includes 30, 60, or 90 days after hospitalization, selected by the hospital - Hospitals continued to receive Medicare FFS payments; allowed payments were retrospectively reconciled against target prices on a quarterly basis - Target prices were based on historical spending minus a 2–3% discount, depending on episode length - Hospitals retained a portion of savings below the target and reimbursed Medicare for part of the difference when costs exceeded the target
	Control: Matched non-BPCI hospitals paid under Medicare IPPS, excluding hospitals participating in other bundled-payment models or BPCI for other conditions.
Outcome(s)	**Description of outcome(s)** Primary: - Changes in Medicare payments per episode of care (defined as the hospitalization plus 90 days after discharge) **Secondary:** - Changes in case mix (e.g. age, Medicaid, Disabled, Race or ethnic group, CCW score, Clinical complexity) - Component costs (incl. total Medicare spending, Index hospitalization, Readmission, Skilled nursing, Inpatient rehabilitation, Long-term care hospital, Home health agency, Physician fees, Non physician outpatient fees) - 30- and 90-day readmissions - Length of stay - Emergency department use - 30- and 90-day mortality
	Measurement time points: single pre/post comparison using one baseline and one intervention measurement
Notes	**Funding:** The Commonwealth Fund
	Conflict of interest: None
	Other comments: voluntary nature of participation; only inclusion of beneficiaries who were continuously enrolled in Medicare Parts A and B during their episode of care, and exclusion of those with end-stage renal disease

Table S16. Extracted study characteristics of Joynt Maddox et al. (2019)

Study ID	**Joynt Maddox (2019)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary participation in BPCI)
	Selection of hospitals (intervention/control): BPCI hospitals/matched non-BPCI hospitals

	Data source: - Medicare inpatient files - Outpatient and carrier files (20% sample)
	Study period (before/after): - Baseline period: 9 to 3 months before hospital's start date - Intervention period: 3 to 9 months after hospital's start date - Burnout period: 3 months before start date - Study uses data from Jan 2013 to Jan 2017
	Statistical analyses: - DiD with marginal models using generalized estimating equations (GENMOD) - Models controlled for patient comorbidities using CCW - sensitivity analysis examining total Medicare parts A and B payments
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare FFS payment system. Hospitals could enroll in BPCI between October 2013, and October 2015.
	Number of included hospitals (intervention/matched control): 264/439
	Characteristics of hospitals (intervention/matched control): No. of beds, mean: 326.4/342.4 Ownership (%): Nonprofit: 79.5/82.0 Public: 4.2/2.5 For-profit: 16.3/15.5 Teaching: 46.2/49.4 Urban (%): 99.2/99.3 Geographic region (%): Northeast: 21.6/21.6 Midwest: 22.3/20.3 South: 36.4/38.0 West: 19.7/20.0
	Patient condition/procedure: LEJR (DRGs 469-470)
	Number of patients included in the analysis: 30,307/47,842 (discharges)
	Characteristics of patients (baseline, intervention/control): Age (%): <65 y.: 7.7/7.0 65-79 y.: 67.2/66.9 ≥80 y.: 25.1/26.0 Female (%): 65.0/64.8 Race or ethnic group (%): White: 89.3/89.2 Black: 6.5/6.8 Hispanic: 1.0/0.9 Unknown or other: 3.2/3.1 Medicaid (%): 10.7/9.9 HCC mean (20% sample): 2.63/2.66
Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI Model 2 - Episode includes index hospitalization + 30, 60, or 90 days post-discharge - FFS payments retrospectively reconciled against hospital-specific target prices based on historical spending minus a 2–3% discount - Hospitals retained part of savings below the target and repaid part of excess spending above the target
	Control: Matched non-BPCI hospitals paid under Medicare IPPS, excluding hospitals that participated in BPCI at any time

Outcome(s)	**Description of outcome(s):** **Primary:** - initial discharge destination - days of PAC delivered - Dispersion of PAC - quality of PAC providers **Secondary:** - total Medicare part A payments - institutional PAC payments
	Measurement time points: quarterly data from Q1 2017–Q3 2019 (only for 90-Day Readmission Rate reported)
Notes	**Funding:** The Commonwealth Fund
	Conflict of interest: Dr Joynt Maddox does contract work for the US Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation
	Other comments: voluntary nature of participation; only inclusion of beneficiaries who were continuously enrolled in parts A and B and exclusion of those with end-stage renal disease

Table S17. Extracted study characteristics of Joynt Maddox et al. (2021)

Study ID	**Joynt Maddox (2021)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary participation in BPCI-A)
	Selection of hospitals (intervention/control 1/control 2): BPCI-A participating hospitals/ late-joining hospitals (in Jan 2020)/ non-joining BPCI-A hospitals
	Data source: - Medicare claims data - data from the American Hospital Association (AHA) 2017 annual survey and the Area Health Resources File
	Study period (before/after): - Baseline: 6 quarters before Oct 2018 - Intervention: 4 quarters after Oct 2018 (early joining hospitals) - second control group comprised hospitals that joined BPCI-A in January 2020 (late-joining hospitals) - Study uses inpatient files from Jan 2017–Sep 2019
	Statistical analyses: - Modified segmented regression analysis - Generalized estimating equation (GEE) models to account for within-hospital correlation over time - Adjustment factors: indicator variables for month, DRG (indicating condition and complexity), individual patient-level Chronic Conditions Warehouse (CCW) coexisting conditions, hospital characteristics (ownership, teaching status, rural location, region) - Sensitivity analyses: log link with gamma distribution, models with hospital random effects
Participants	**Country/region/setting:** USA/nationwide/acute, outpatient, and post-acute care
	Health system characteristics: Medicare FFS payment system transitioning toward alternative payment models. BPCI-A, launched in October 2018 as the successor to the BPCI program, retained FFS payments with subsequent episode-based reconciliation against target prices.
	Number of included hospitals (intervention/ control 1/control 2): 826/334/2016

	Characteristics of hospitals (intervention/control 1/ control 2): No. of beds: NR Ownership (%): - Nonprofit: 68.6/69.3/58.1 - Public: 4.1/9.0/19.6 - For-profit: 27.3/21.8/22.4 Teaching (%): - Major: 11.5/5.3/6.1 - Minor: 36.1/32.7/22.4 - Non: 52.1/60.8/71.5 Rural (%): 1.5/2.4/12.3 Region: NR
	Patient condition/procedure: all 29 BPCI inpatient conditions
	Number of patients included in the analysis (intervention/control 1/control 2): 1,633,448/288,288/5,692,561
	Characteristics of patients (intervention/control 1/control 2): Age (%): <65 y.: 11.1/12.5/12.0 65-80 y.: 46.2/46.3/49.0 ≥80 y.: 42.7/41.1/38.9 Female (%): 56.1/55.3/56.1 Race or ethnic group (%): White: 79.4/81.6/83.0 Black: 10.0/10.0/8.9 Hispanic: 6.5/4.7/4.5 other: 4.1/3.8/3.7 Medicaid (%): 28.0/29.5/27.9 CCW score: 6.3/6.2/5.8
Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI-A - Episode includes the index hospitalization or procedure and extends for 90 days thereafter - FFS payments were reconciled quarterly against target prices, with hospitals receiving reconciliation payments below the target and making payments to Medicare above the target - Reconciliation payments were adjusted according to quality performance
	Control 1: Late-joining hospitals: Joined BPCI-A in January 2020 (post-intervention period) **Control 2:** Non-joining hospitals: Never participated in BPCI-A; continued traditional payments
Outcome(s)	**Description of outcome(s):** **Primary:** - Changes in quarterly trends in 90-day Medicare payments - 90-day readmission rates **Secondary:** - individual components of payment (quarterly trend) - 30-day readmission - 30- and 90-day mortality - Healthy days at home - Composite of 90-day readmission or mortality - Composite of 30-day readmission or mortality
	Measurement time points: quarterly from Q1 2017 – Q3 2019 (for primary outcomes)
Notes	**Funding:** Supported by a grant (R01HL143421) from the National Heart, Lung, and Blood Institute
	Conflict of interest: none

	Other comments: voluntary nature of participation

Table S18. Extracted study characteristics of Kallies et al. (2022)

Study ID	**Kallies (2022)**
Methods	**Study design:** CBA
	Allocation: natural experiment (mandatory bundled payment model)
	Selection of hospitals (intervention/control): all hospitals in 67 randomly assigned SMSAs (CJR hospitals)/hospitals in areas exempt from CJR
	Data source: - CMS's Medicare FFS Research Identifiable Inpatient, Outpatient and Carrier files
	Study period (before/after): 2014–2015 (pre-CJR)/2017 (post-CJR)
	Statistical analyses: - DiD analyses with multivariate generalized linear models with log link, and multivariate logistic regression - Adjusted for patients' age, sex, race/ethnicity, poverty status (as measured by dual enrollment in Medicaid or a state buy-in program), number of comorbidities (0, 1, 2, 3, 4+), an indicator for whether the hip replacement was secondary to a hip fracture, prior SNF stay within 6 months from the THA surgery, and census division of the acute care treatment hospital
Participants	**Country/region/setting:** USA/CJR-selected and control SMSAs/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2016, CMS implemented the CJR Model as a mandatory episode-based bundled payment model for joint replacement.
	Number of included hospitals (intervention/ control): NR
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedure: primary THA (DRGs 469 and 470)
	Number of patients included in the analysis (intervention/control): 229,828/128,016 (episodes)
	Characteristics of patients (all eligible): Age, y, n (%): 66–69: 62,874 (17.6) 70–74: 82,562 (23.1) 75–79: 74,734 (20.9) 80–84: 62,565 (17.5) 85–89: 46,620 (13.0) ≥90: 28,489 (8.0) Sex, n (%): Female: 236,914 (66.2) Male: 120,930 (33.8) Race, n (%): Non-Hispanic White: 330,473 (92.4) Black/African American: 13,544 (3.8) Hispanic: 7605 (2.1) Other: 6222 (1.7) Medicaid eligible, n (%): 25,615 (7.2) Number of comorbidities, n (%): 0: 55,666 (15.6) 1: 77,658 (21.7) 2: 72,458 (20.2) 3: 52,807 (14.8) ≥4: 99,255 (27.7)
Intervention and Control	**Description of intervention:** - Mandatory participation in CJR for hospitals in 67 randomly selected SMSAs - Episode includes the inpatient stay and 90 days post-discharge, covering relevant Medicare costs across PAC settings - Hospitals assumed financial risk for episode costs, intended to optimize PAC

	Control: Hospitals exempt from CJR participation
Outcome(s)	**Description of outcome(s):** - Escalation in care to institutionalization (admission to an IRF or SNF during 90 days post-discharge for those initially discharged to the community) - Return to the community at the end of the episode of care among those initially discharged to an institutional setting
Notes	**Funding:** Supported in part by a research grant from the National Institute on Aging (NIA) to (5-R01-AG058718)
	Conflict of interest: None
	Other comments: only data from beneficiaries continuously enrolled in Medicare Parts A and B were included; data from Medicare Part C (Medicare Advantage) were not considered

Table S19. Extracted study characteristics of Kiani et al. (2024)

Study ID	**Kiani (2024)**
Methods	**Study design:** CBA
	Allocation: natural experiment (mandatory bundled payment program)
	Selection of hospitals (intervention/control): CJR-participating hospitals/non-CJR hospitals (stratified analysis by insurance type; Medicare vs non-Medicare)
	Data source: New York Statewide Planning and Research Cooperative System (SPARCS) database
	Study period (before/after): Jul 2014–Mar 2016 (pre-CJR)/Apr 2016–Dec 2017 (post-CJR)
	Statistical analyses: - DiD analysis with linear regression models - Poisson distribution with log link (for binary outcomes) - Regression models were adjusted for all available relevant covariates (specific variables not explicitly listed)
Participants	**Country/region/setting:** USA/New York/New York State hospitals
	Health system characteristics: Medicare payment system. In April 2016, CMS introduced the CJR bundled-payment program for LEJR.
	Number of included hospitals (intervention/ control): 49/95
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedure: primary elective THA or TKA surgery (for THA: ICD-9 81.51 and ICD-10 0SR9x or 0SRBx; for TKA: ICD-9 81.54 and ICD-10 0SRCx or 0SRDx)
	Number of patients included in the analysis (intervention/control): 58,610/78,728
	Characteristics of patients (before, intervention/control): Age, mean (SD): 65.7 (10.8)/65.1 (10.8) Sex, n (%): Female: 13,058 (60.3)/ 21,766 (58.9) Male: 8,596 (39.7)/15,190 (41.1) Race, n (%): White: 15,382 (71.0)/29,855 (80.8) African American: 2490 (11.5)/3526 (9.5) Native American: 40 (0.2)/72 (0.2) Asian Indian: 33 (0.2)/36 (0.1) Chinese: 58 (0.3)/36 (0.1) Other Asian: 370 (1.7)/277 (0.7) Other: 3244 (15.0)/2759 (7.5) Multiracial: 37 (0.2)/395 (1.1) Medicaid, n (%): 854 (3.9)/911 (2.5) Charlson-Deyo Index, n (%): 0: 12,616 (58.3)/21,645 (58.6) 1: 5927 (27.4)/10,080 (27.3) 2: 1948 (9.0)/3202 (8.7)

	>2: 1163 (5.4)/ 2029 (5.5)
Intervention and Control	**Description of intervention:** - Participation in the CJR bundled-payment program - Bundled payment covering services related to LEJR - Financial risk shifted to the hospital system
	Control: Hospitals not participating in the CJR program
Outcome(s)	**Description of outcome(s):** Primary: - percentage of patients with a low comorbidity burden as measured by the Charlson-Deyo Comorbidity Index - patient age Secondary: - inflation-adjusted cost of hospitalization - rate of discharge to institutional PAC - percentage of patients with a hospital stay longer than 3 days
	Measurement time points: - primary outcomes measured monthly from Jun 14 to Sep 17
Notes	**Funding:** The author(s) received no financial support for the research, authorship, and/or publication of this article
	Conflict of interest: CM reports relationships with 3M, Biocomposites, and Smith & Nephew. The remaining authors declare no potential conflicts of interest. The data used to produce this publication were purchased from or provided by the New York State Department of Health (NYSMDOH). However, the conclusions derived and the views expressed herein are those of the author(s) and do not reflect the conclusions or views of NYSMDOH. NYSMDOH, its employees, officers, and agents make no representation, warranty, or guarantee as to the accuracy, completeness, currency, or suitability of the information provided here.
	Other comments: None

Table S20. Extracted study characteristics of Kim et al. (2020)

Study ID	**Kim (2020)**
Methods	**Study design:** CBA
	Allocation: natural experiment (random assignment of MSAs to CJR participation)
	Selection of hospitals (intervention/control): treatment MSAs selected for CJR participation/control MSAs not participating in CJR
	Data source: - Medicare inpatient, outpatient, skilled nursing facility, home health agency, and carrier claims - patient demographic characteristics from the master beneficiary summary files - CMS Provider of Service files and Provider Specific files for hospital and PAC information
	Study period (before/after): Jan 2013–Jun 2015 (pre-CJR)/ Apr 2016–Sep 2017 (post-CJR)
	Statistical analyses: - Difference-in-differences-in-differences using linear regression models - Adjustment factors: patient age, sex, surgery type - propensity score weighting and instrumental variable analysis used as sensitivity analyses
Participants	**Country/region/setting:** USA/CJR-selected and control MSAs/acute and post-acute care
	Health system characteristics: Medicare Payment system. In April 2016, mandatory CJR bundled-payment model was implemented in 67 randomly selected MSAs.
	Number of included hospitals (intervention/ control): NR (67/103 MSAs)

	Characteristics of hospitals (before, White/Black/Hispanic) [2]**:** No. of beds: NR Ownership, n (%): For profit: 70,294 (17.2)/3,135 (16.4)/2,507 (20.7) Not for profit: 280,082 (68.4)/12,715 (66.4)/7,859 (65.0) Public: 51 115 (12.5)/2,950 (15.4)/1,551 (12.8) Other: 7,823 (1.9)/352 (1.8)/174 (1.4) Major teaching, n (%): 87,104 (21.3)/5,813 (30.4)/2,350 (19.4) Location: NR Region: NR
	Patient condition/procedure: inpatient joint replacement surgeries (MS-DRG 469 and 470)
	Number of patients included in the analysis (intervention/control): 688,346 patients (747,098 joint replacement surgeries)
	Characteristics of patients (before, White/Black/Hispanic): Age, y, n (%): 66–70: 115,295 (28.2)/6,774 (35.4)/3,983 (32.9) 71–75: 107,993 (26.4)/5,278 (27.6)/3,370 (27.9) 76–80: 84,895 (20.7)/3,669 (19.2)/2,451 (20.3) >80: 101,130 (24.7)/3,430 (17.8)/2,286 (18.9) Female, n (%):261,703 (63.9)/13,714 (71.6)/ 8025 (66.4) Medicaid, n (%): 14,862 (3.6)/ 3,180 (16.4)/3,847 (31.8) Medically complex, n (%): 40,173 (9.8)/2,817 (14.7)/1,437 (11.9)
Intervention and Control	**Description of intervention:** - Mandatory participation in the CJR bundled-payment model - Episode includes the index hospitalization through 90 days after discharge - Hospitals are accountable for episode spending and quality - Hospitals repay a portion of excess spending above a quality-adjusted spending limit and receive a bonus when spending falls below the limit
	Control: Hospitals in control MSAs not selected for CJR
Outcome(s)	**Description of outcome(s):** Primary: - Total spending - Discharge to institutional PAC - 90-day readmission Secondary: - discharges to home or home health care - days of institutional PAC - length of index hospitalization stay - spending incurred at each care setting - post-discharge complication - 90-day emergency department use - 90-day all-cause mortality
Notes	**Funding:** funded by grant R01MD011403 from the National Institute on Minority Health and Health Disparities and grant K24AR055259 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (Dr Ibrahim)
	Conflict of interest: None
	Other comments: Exclusion of joint replacements occurring from July 2015 to March 2016 because hospitals may have started to change their practice patterns after the CMS announcement of the CJR model in July 2015

[2] Numbers refer to the proportion of patients treated in hospitals with the respective attribute, not to the proportion of hospitals with that attribute

Table S21. Extracted study characteristics of Liao et al. (2022)

Study ID	**Liao (2022)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary participation in bundled payment model 2)
	Selection of hospitals (intervention/control): - BPCI-participating physician group practices (PGPs)/non-BPCI-participating PGPs - BPCI-participating hospitals/non-BPCI-participating hospitals - BPCI-participating PGPs/BPCI-participating hospitals
	Data source: - Medicare claims data - American Hospital Association Annual Survey to capture hospital characteristics - Medicare Provider of Service, Beneficiary Summary, American Community Survey, and Accountable Care Organization files to obtain market and additional hospital characteristics - Physician Compare, and Compendium of US Health Systems to obtain physician characteristics
	Study period (before/after): Jan 2011–Sep 2013/Oct 2013–Dec 2017
	Statistical analyses: - DiD using generalized linear models with identity links and normal distributions - Adjustment factors: Covariates for patient demographics (age, sex, comorbidities), market characteristics (Medicare population size, Medicare Advantage penetration), and hospital-specific indicators (teaching status, ownership) - Propensity-matching
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare Payment system
	Number of included hospitals (after matching, BPCI hospitals/non-BPCI hospitals): 379/1,441 **physicians in BPCI-participating PGPs/matched physicians in non−BPCI-participating PGPs:** 6,405/24,758
	Characteristics of hospitals (after matching, BPCI hospitals/non-BPCI hospitals): Total beds, mean (SD): 315.4 (217.1)/301.8 (211.8) Ownership, n (%): Not-for-profit: 902 (75%)/2700 (74.9%) For-profit: 269 (22.4%)/822 (22.8%) Government: 31 (2.6%)/84 (2.3%) Rural, n (%): 69 (5.7%)/167 (4.6%) Teaching status, n (%): Non-teaching: 579 (48.2%)/1771 (49.1%) Minor teaching: 487 (40.5%)/1456 (40.4%) Major teaching: 136 (11.3%)/379 (10.5%) Region: NR **Characteristics of Physicians (after matching, BPCI PGPs/Non-BPCI PGPs):** Female, n (%): 4,094 (24.9)/12,376 (25.1) Age, mean (SD): 51.4 (8.5)/ 51.4 (9.0) Annual services provided, Mean (SD): 1,582.0 (1,294.8/1,527.4 (1,484.3) Rural, n (%): 347 (2.1)/1,004 (2.0)

	Patient condition/procedure: any of the 10 highest-volume episodes (5 medical and 5 surgical) included in the BPCI initiative for Medicare patients medical: - CHF - COPD, Bronchitis, or Asthma - Pneumonia and Pleurisy - Respiratory Infections and Inflammations - Sepsis surgical: - Hip and Femur Procedures except Major Joint - LEJR - PCI - Spinal fusion - Upper extremity joint replacement
	Number of patients included in the analysis (intervention/control): 696,710/592,071
	Characteristics of patients (baseline, non-BPCI/BPCI PGP/BPCI Hospital/BPCI Both): **Patients admitted for Medical Episodes:** Age, y, mean (SD): 76.5 (12.1)/76.2 (12.1)/77.8 (12.0)/77.2 (12.3) Sex, n (%): Female: 176,766 (56.6)/25,252 (56.3)/19,929 (56.7)/2,206 (55.2) Male: 135,785 (43.4)/19,579 (43.7)/15,243 (43.3)/1,788 (44.8) Race, n (%): White: 262,328 (83.9)/ 37,757 (84.2)/ 28,460 (80.9)/ 3,334 (83.5) Black: 27,702 (8.9)/ 3,653 (8.2%)/ 3,518 (10.0%)/ 326 (8.2%) Other: 22,521 (7.2)/ 3,421 (7.6)/ 3,194 (9.1)/ 334 (8.4) Dual Eligible, n (%): 84,864 (27.2)/12,633 (28.2)/9,368 (26.6)/ 1,029 (25.8) Disabled, n (%): 42,910 (13.7)/6403 (14.3)/4107 (11.7)/533 (13.4) Elixhauser Comorbidity Index, mean (SD): 20.09 (14.1)/19.46 (14.0)/20.50 (14.0)/20.35 (14.0) **Patients admitted for surgical episodes:** Age, y, mean (SD): 74.9 (9.6)/74.7 (9.0)/75.1 (9.5)/74.9 (8.7) Sex, n (%): Female: 131,081 (63.1)/17,555 (63.5)/22,846 (64.4)/1,951 (65.7) Male: 76,635 (36.9)/10,103 (36.5)/12,634 (35.6)/1,020 (34.3) Race, n (%): White: 184,823 (89.0)/25,057 (90.6)/30,890 (87.1)/2,640 (88.9) Black: 11,166 (5.4)/1,293 (4.7)/2,170 (6.1)/196 (6.6) Other: 11,727 (5.7)/1,308 (4.7)/2,420 (6.8)/135 (4.5) Medicare/Medicaid dual eligible, n (%): 27,923 (13.4)/3,041 (11.0)/4,643 (13.1)/298 (10.0) Disabled, n (%): 18,987 (9.1)/2051 (7.4)/2924 (8.2)/194 (6.5) Elixhauser comorbidity index, mean (SD): 7.2 (11.9)/5.8 (11.1)/6.5 (11.4)/5.9 (11.2)
Intervention and Control	**Description of intervention:** - Participation in BPCI Model 2 by either hospitals or physician group practices (PGPs) - Episode spans the index hospitalization and up to 90 days of PAC - Participants assumed accountability for episode quality and costs
	Control: - Matched non-BPCI PGPs that never participated in any BPCI episode - Matched non-BPCI hospitals that never participated in any BPCI episode
Outcome(s)	**Description of outcome(s):** Primary: 90-day total episode spending Secondary: 90-day readmissions, 90-day mortality, exploratory spending and utilization outcomes
	Measurement time points: - Quarterly from Q1 2011 – Q2 2017 - Exclusion of transition period (Jan 2013 – Sep 2013)

Notes	**Funding:** funded by grants from the US National Institutes of Health's National Institute on Minority Health and Health Disparities (Nos. R01MD013859, R01MD015055) and the Agency for Healthcare Research and Quality (No. R01HS027595).
	Conflict of interest: Dr Liao reported personal fees/honoraria from Marcus Evans, Comagine Health, and Brown University outside the submitted work. Dr Linn reported grants from Commonwealth Fund and Robert Wood Johnson Foundation and personal fees from Correlation One outside the submitted work. Dr Navathereported grants from the Hawaii Medical Service Association, the Commonwealth Fund, the Robert Wood Johnson Foundation, the Donaghue Foundation, the Pennsylvania Department of Health, the Veterans Affairs Administration, Ochsner Health System, United Healthcare, Blue Cross Blue Shield of North Carolina, Blue Shield of California, and Humana; personal fees from Navvis Healthcare, Yale New Haven Hospital/Center for Outcomes Research and Evaluation, Maine Health Accountable Care Organization, Singapore Ministry of Health, Elsevier Press, Medicare Payment Advisory Commission, Cleveland Clinic, the Analysis Group, the VBID Health, Advocate Physician Partners, the Federal Trade Commission, and the Catholic Health Services Long Island; equity from Nava Health and Clarify Health; and noncompensated board membership from Integrated Services, all outside the submitted work. No other disclosures were reported.
	Other comments: None

Table S22. Extracted study characteristics of Martin et al. (2018)

Study ID	**Martin (2018)**
Methods	**Study design:** CBA
	Allocation: natural experiment (linked to hospital participation in the BPCI program)
	Selection of hospitals (intervention/control): Hospitals grouped based on BPCI participation status: nonparticipants, preparatory, or risk-bearing
	Data source: - Medicare Inpatient Research Identifiable File (RIF) to identify admissions - Master Beneficiary Summary file to define cohort eligibility and to describe patient demographics - American Hospital Association Annual Survey (2012) to summarize hospital bed count and total discharges
	Study period (before/after): Jan 2012–Sep 2012/Jan 2013–Sep 2013
	Statistical analyses: - Generalized estimating equation regression - Adjustment factors: Patient age, sex, race, comorbidity (Charlson index), log-hospital bed size, and log-total discharges
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2013, CMS introduced the voluntary BPCI program, with risk-bearing contracts beginning in October 2013.
	Number of included hospitals (Risk-Bearing/Preparatory/Nonparticipant hospitals): 102/671/1,277
	Characteristics of hospitals (2012, Risk-Bearing /Preparatory /Nonparticipant hospitals): No. of beds, mean: 443/315/284 Ownership: NR Teaching: NR Location: NR Region: NR
	Patient condition/procedure: lumbar fusion
	Number of patients included in the analysis (2012, Risk-Bearing /Preparatory/ Nonparticipant hospitals): 3,715/19,038/29,707

	Characteristics of patients (2012, Risk-Bearing /Preparatory/ Nonparticipant hospitals): Age, mean: 73.6/73.5/73.4 Female, %: 59.5/59.7/59.6 Race, %: White: 91.8/92.2/92.3 Black: 5.5/4.4/4.4 Other: 2.7/3.4/3.2 Charlson Comorbidity Index, %: None: 48.7/49.7/48.7 One: 29.1/28.8/28.8 Two +: 22.2/21.6/22.6
Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI Model 2 - Episode includes the triggering hospitalization and a 30-, 60-, or 90-day post-acute period selected by the hospital - Hospitals continued to receive Medicare FFS payments, which were subsequently reconciled against a prenegotiated target price discounted by 2–3% - Hospitals assumed financial risk for keeping episode spending at or below the target
	Control: Hospitals not participating in the BPCI program
Outcome(s)	**Description of outcome(s):** Primary: 90-day episode-of-care payments Secondary: - Lumbar Fusion procedure volume - All-cause 90-day readmission - Repeat lumbar spine surgery
	Measurement time points: annually (2012 vs. 2013)
Notes	**Funding:** - Agency for Healthcare Research and Quality (AHRQ), grant 5U19HS024075-02 and 5K12HS021695-02 - National Institute of Arthritis, Musculoskeletal, and Skin Disease, grant 4P60AR062799-05
	Conflict of interest: NR
	Other comments: Participation in BPCI Model 2 was voluntary

Table S23. Extracted study characteristics of Maughan et al. (2019)

Study ID	**Maughan (2019)**
Methods	**Study design:** CBA
	Allocation: natural experiment (linked to hospital participation in the BPCI program)
	Selection of hospitals (intervention/control): BPCI-participating hospitals/matched non-BPCI hospitals
	Data source: Medicare claims and enrollment data
	Study period (before/after): Oct 2011–Sep 2012/Oct 2013–Dec 2016
	Statistical analyses: - DiD with logistic regression - Adjustment factors: Beneficiaries' demographic characteristics, health status, prior care use, hospital and regional characteristics, and other health care initiatives - propensity score matching
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2013, CMS implemented the BPCI initiative to test episode-based payment across Medicare-covered services.

	Number of included hospitals (after matching, intervention/control): Clinical Episode: - Urinary Tract infection: 83/556 - Stroke: 77/503 - COPD, Bronchitis, Asthma: 133/817 - Major LEJR: 304/1,160 - Congestive Heart Failure: 173/897 - AMI: 93/640 - Hip and Femur Procedures Except Major Joint: 101/602 - Sepsis: 119/731 - Simple Pneumonia and Respiratory Infections: 132/817 - Renal Failure: 75/518 - Coronary Artery Bypass Graft: 43/250 - Percutaneous Coronary Intervention: 45/268 - Cardiac Valve: 32/173
	Characteristics of hospitals (after matching, intervention/control): Clinical Episode: Urinary Tract Infection Bed Count, mean: 272/266 Ownership: Non-Profit (%): 69/70 For-Profit (%): 28/26 Government (%): 4/4 Urban (%): 98/97 Teaching (%): 45/46 Average Case Weight of MS-DRG Discharges, mean: 1.58/1.56 Clinical Episode: Stroke Bed Count, mean: 380/392 Ownership: Non-Profit (%): 70/69 For-Profit (%): 21/21 Government (%): 9/10 Urban (%): 99/98 Teaching (%): 58/60 Average Case Weight of MS-DRG Discharges, mean: 1.67/1.67 Clinical Episode: COPD, Bronchitis, Asthma Bed Count, mean: 289/313 Ownership: Non-Profit (%): 75/75 For-Profit (%): 20/19 Government (%): 5/6 Urban (%): 94/93 Teaching (%): 50/53 Average Case Weight of MS-DRG Discharges, mean: 1.60/1.58 Clinical Episode: Major LEJR Bed Count, mean: 302/323 Ownership: Non-Profit (%): 76/77 For-Profit (%): 17/15 Government (%): 7/7 Urban (%): 94/93 Teaching (%): 46/48 Average Case Weight of MS-DRG Discharges, mean: 1.62/1.64 Clinical Episode: Congestive Heart Failure Bed Count, mean: 313/309 Ownership: Non-Profit (%): 80/82 For-Profit (%): 13/10

	Government (%): 6/8 Urban (%): 91/90 Teaching (%): 45/45 Average Case Weight of MS-DRG Discharges, mean: 1.60/1.56 Clinical Episode: AMI Bed Count, mean: 316/292 Ownership: Non-Profit (%): 76/75 For-Profit (%): 17/20 Government (%): 6/5 Urban (%): 94/93 Teaching (%): 43/44 Average Case Weight of MS-DRG Discharges, mean: 1.60/1.60 Clinical Episode: Hip and Femur Procedures Except Major Joint Bed Count, mean: 313/315 Ownership: Non-Profit (%): 72/73 For-Profit (%): 23/22 Government (%): 5/5 Urban (%): 98/98 Teaching (%): 51/52 Average Case Weight of MS-DRG Discharges, mean: 1.62/1.61 Clinical Episode: Sepsis Ownership: Non-Profit (%): 68/64 For-Profit (%): 28/31 Government (%): 4/5 Urban (%): 95/95 Teaching (%): 45/48 Bed Count, mean: 273/282 Average Case Weight of MS-DRG Discharges, mean: 1.61/1.58 Clinical Episode: Simple Pneumonia and Respiratory Infections Bed Count, mean: 272/268 Ownership: Non-Profit (%): 74/78 For-Profit (%): 20/17 Government (%): 5/5 Urban (%): 95/96 Teaching (%): 42/44 Average Case Weight of MS-DRG Discharges, mean: 1.58/1.58 Clinical Episode: Renal Failure Bed Count, mean: 287/284 Ownership: Non-Profit (%): 69/73 For-Profit (%): 27/24 Government (%): 4/3 Urban (%): 97/97 Teaching (%): 41/44 Average Case Weight of MS-DRG Discharges, mean: 1.61/1.58 Clinical Episode: Coronary Artery Bypass Graft Bed Count, mean: 467/483 Ownership: Non-Profit (%): 86/85 For-Profit (%): 9/9 Government (%): 5/7

	Urban (%): 98/99 Teaching (%): 74/76 Average Case Weight of MS-DRG Discharges, mean: 1.75/1.72 Clinical Episode: Percutaneous Coronary Intervention Bed Count, mean: 335/316 Ownership: Non-Profit (%): 62/63 For-Profit (%): 36/36 Government (%): 2/1 Urban (%): 96/96 Teaching (%): 56/54 Average Case Weight of MS-DRG Discharges, mean: 1.68/1.70 Clinical Episode: Cardiac Valve Bed Count, mean: 551/545 Ownership: Non-Profit (%): 84/79 For-Profit (%): 3/5 Government (%): 13/16 Urban (%): 94/92 Teaching (%): 71/70 Average Case Weight of MS-DRG Discharges, mean: 1.77/1.77
	Patient condition/procedure: 13 (10 common + 3 cardiovascular episodes) selected out of a total of 48 possible BPCI conditions (incl. urinary tract infection, stroke, COPD/Bronchitis/Asthma, MJRLE, CHF, AMI, hip and femur procedures, sepsis, pneumonia and respiratory infections, renal failure, CABG, PCI, cardiac valve) among patients with dementia, dual eligibility and recent institutional care
	Number of patients included in the analysis (intervention/control): 105,458 (episodes)
	Characteristics of patients (intervention/control): NR
Intervention and Control	**Description of intervention:** - Participation in BPCI Model 2 - Episode begins with an anchor hospitalization at a BPCI-participating acute care hospital or involving a physician from a BPCI-participating physician group practice and continues for up to 90 days after discharge - Participants were financially rewarded when episode payments were below a target price and could be required to make payments to CMS when episode payments exceeded the target
	Control: Matched control hospitals not participating in BPCI
Outcome(s)	**Description of outcome(s):** Primary: - Changes in emergency department visits - unplanned hospital readmissions - and all-cause mortality within 90 days of discharge
Notes	**Funding:** funded by the CMS (Contract Nos. HHSM-500-2011-00001I and HHSM-500-2011-00001T)
	Conflict of interest: NR
	Other comments: - Hospitals participated in BPCI voluntarily and could exit the initiative at various times - The result may be a biased estimate because the BPCI and comparison populations had significantly different baseline trends ($p< 0.05$) for this outcome

Table S24. Extracted study characteristics of McNeely et al. (2020)

Study ID	**McNeely (2020)**
Methods	**Study design:** CBA
	Allocation: natural experiment (linked to hospital participation in the BPCI program)

	Selection of hospitals (intervention/control): BPCI participating hospitals (PCI and CABG)/matched non participating hospitals
	Data source: - Medicare inpatient files - 2014 American Hospital Association file for hospital and market characteristics - The Area Resource File for PAC services and median income levels
	Study period (before/after): - Baseline: 9 to 3 months (a 6-month period) before each hospital's start date - Intervention: starting immediately after the start date (varied per hospital, mean intervention period of 16 months) - Study uses data from Jan 2013 – Sep 2016
	Statistical analyses: - DiD with mixed-effects linear regression - Covariates based on hospital characteristics and market characteristics - propensity score matching
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2013, the Center for Medicare and Medicaid Innovation launched the BPCI initiative as an episode-based alternative to traditional fee-for-service payment.
	Number of included hospitals (BPCI/matched non-BPCI): PCI: 42/115 CABG: 45/122
	Characteristics of hospitals (after matching, BPCI/non-BPCI): PCI hospitals: No. of beds, mean: 321/307 No. of cases: 108.1/81.8 Ownership: Non-Profit (%): 64.3/62.6 Public (%): 0/1.7 For-Profit (%): 35.7/35.7 Region: Northeast (%): 21.4/20.0 Midwest (%): 11.9/12.2 South (%): 40.5/39.1 West (%): 26.2/28.7 Urban (%): 100/100 Teaching (%): 42.9/43.5 Region: NR CABG hospitals: No. of beds, mean: 512/450 No. of cases: 51.5/54.5 Ownership: Non-Profit (%): 84.8/78.7 Public (%): 4.3/6.6 For-Profit (%): 10.9/14.8 Region: Northeast (%): 43.5/38.5 Midwest (%): 17.4/19.7 South (%): 28.3/29.5 West (%): 10.9/10.7 Urban (%): 100/100 Teaching (%): 80.4/74.6 Region: NR
	Patient condition/procedure: PCI or CABG

	Number of patients included in the analysis (intervention/control): - PCI baseline period: 2,054/4,339 (episodes) - PCI intervention period: 3,799/8,657 (episodes) - CABG baseline period: 1,245/2,254 (episodes) - CABG intervention period: 3,898/7,744 (episodes)
	Characteristics of patients (baseline, intervention/control): **PCI patients:** Age, y, %: ≤ 64: 10.9/10.0 65–79: 60.4–59.8 80+: 28.7/30.1 Female, %: 39.6/39.2 Race, %: White: 87.1/88.5 Black: 6.0/6.2 Hispanic: 2.1/1.3 Unknown/other: 4.7/3.9 Medicaid, %: 19.2/16.7 Disabled, %: 22.7/23.0 CCW mean: 7.69/7.73 **CABG patients:** Age, y, %: ≤ 64: 8.0/8.5 65–79: 76.5/75.3 80+: 15.5/16.2 Female, %: 28.8/26.4 Race, %: White: 85.1/91.1 Black: 7.1/4.6 Hispanic: 1.4/0.9 Unknown/other: 6.5/3.4 Medicaid, %: 15.0/11.7 Disabled, %: 18.3/19.8 CCW mean: 7.75/7.59
Intervention and Control	**Description of intervention:** - Participation in BPCI Model 2 - Episode includes the inpatient hospital stay and related services for 30, 60, or 90 days after discharge; this study analyzed 90-day episodes only - Total episode payments were compared with target prices based on historical data and risk-adjustment models - Hospitals retained a portion of savings below the target and reimbursed Medicare for part of the difference when payments exceeded the target
	Control: Hospitals not participating in the BPCI program
Outcome(s)	**Description of outcome(s):** Primary: Change in standardized Medicare allowed payments per 90-day episode Secondary: - changes in hospital case mix (based on the mean Chronic Conditions Warehouse score, ranges from 0 to 27) - discharge to PAC - per-hospital case volume - proportion of patients dually enrolled in Medicare and Medicaid - proportion of patients with disabling conditions - changes in the individual components of payment - LOS - emergency department use (30-days, 90-days) - readmissions (30-days, 90-days) - and mortality (30-days, 90-days)

Notes	**Funding:** Supported by the National Heart, Lung, and Blood Institute (R01HL143421).
	Conflict of interest: Dr Joynt Maddox receives research support from the National Heart, Lung, and Blood Institute (R01HL143421), National Institute on Aging (R01AG060935), and Commonwealth Fund and previously did contract work for the US Department of Health and Human Services. Dr Orav does contract work for the US Department of Health and Human Services. The other authors report no conflicts.
	Other comments: Participation in BPCI Model 2 was voluntary; only beneficiaries continuously enrolled in Medicare Parts A and B during the episode of care were included, and those eligible for Medicare because of end-stage renal disease were excluded; physician services and other outpatient care were excluded from the payment analysis because claims were available only for a 20% sample of beneficiaries; the authors noted that secondary end points and analyses should be considered exploratory.

Table S25. Extracted study characteristics of Meyers et al. (2019)

Study ID	**Meyers (2019)**
Methods	**Study design:** CBA
	Allocation: natural experiment (mandatory bundled payment program for joint replacement surgical procedures in randomly selected MSAs)
	Selection of hospitals (intervention/control): Hospitals in MSAs randomly assigned to mandatory CJR participation/Hospitals in 121 MSAs not selected for CJR
	Data source: - Medicare Provider Analysis and Review (MedPAR) files - claims and PAC assessment data - data from annual Medicare Master Beneficiary Summary files (patient demographic characteristics and MA enrollment status)
	Study period (before/after): Jan 2013 – Mar 2016/April 2016 – Sept 2017 (Analyses compared changes in outcomes before and after 2 points: July 14, 2015, when the hospitals that would participate were announced, and April 1, 2016, when mandatory bundled payments were implemented)
	Statistical analyses: - DiD with linear regression and Poisson model, stratified by MA status - Adjustment factors: Sex, Race/Ethnicity, Medicaid eligibility, Quarter, Hospital fixed effects
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2016, CMS introduced the mandatory CJR bundled-payment model in randomly selected MSAs.
	Number of included hospitals (intervention/control): NR (75/121 MSAs)
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedure: hip and knee joint replacement operations (MS-DRG 469 or 470)
	Number of patients included in the analysis (intervention/control): 1,536,387 total (493,977 MA patients, 1,042,410 TM patients[3])
	Characteristics of patients (before, intervention/control): **TM patients:** Age, mean (SD), y: 73.4 (8.8)/73.2 (8.8) Women, n (%): 171,882 (63.7)/212,213 (63.5) Race/ethnicity, n (%): White: 340,938 (88.6)/429,906 (89.9) Black: 24,683 (6.4)/30,314 (6.3) Asian: 4,518 (1.2)/3,208 (0.7) Hispanic: 5,322 (1.4)/4,725 (1.0)

[3] TM = Traditional Medicare; MA = Medicare Advantage. TM beneficiaries are covered under FFS Medicare and were included in bundled payment programs when treated by participating hospitals. MA beneficiaries are enrolled in private managed care plans and are generally excluded from CMS bundled payment initiatives.

	Native American or American Indian: 1,029 (0.3)/1,792 (0.4) Other or unknown: 8,141 (2.1)/8,495 (1.8) Dually eligible, n (%): 30,755 (11.4)/34,842 (10.4) Elixhauser comorbidities, mean (SD): 2.4 (1.7)/2.3 (1.6) **MA patients:** Age, mean (SD), y: 73.4 (8.4)/73.2 (8.3) Women, n (%): 82,729 (65.4)/96,233 (64.7) Race/ethnicity, n (%): White: 145,294 (84.2)/176,190 (87.1) Black: 16,093 (9.3)/17,612 (8.7) Asian: 2,296 (1.3)/1,465 (0.7) Hispanic: 4,626 (2.7)/2,968 (1.5) Native American or American Indian: 271 (0.2)/501 (0.3) Other or unknown: 4,077 (2.4)/3,536 (1.8) Dually eligible, n (%): 17,628 (13.9)/17,280 (11.6) Elixhauser comorbidities, mean (SD): 2.4 (1.6)/2.4 (1.6)
Intervention and Control	**Description of intervention:** - Mandatory participation in CJR for hospitals in randomly selected MSAs - Episode includes the index surgery and all Medicare Parts A and B care through 90 days after surgery - Annual spending benchmarks were set by CMS - Hospitals paid a penalty when spending exceeded the benchmark and shared in savings when spending was below the benchmark
	Control: Hospitals in MSAs not selected for mandatory CJR participation
Outcome(s)	**Description of outcome(s):** Primary: - discharge to any institutional PAC and days spent in institutional PAC within 90 days after surgery Secondary: - discharge to any institutional PAC setting - estimated total spending - initial hospital LOS - discharge to home with and without home health care - and the number of days in each posthospital setting
	Measurement time points: Quarterly data from Q1 2013 – Q3 2017
Notes	**Funding:** Supported by grant P01AG027296 from the National Institute on Aging of the National Institutes of Health
	Conflict of interest: Dr Mor reported chairing the scientific advisory committee for naviHealth, which manages post-acute care for Medicare Advantage plans and hospitals. Dr Trivedi reported receiving grants from the US Department of Veterans Affairs Health Services Research and Development during the conduct of the study. No other disclosures were reported.
	Other comments: none

Table S26. Extracted study characteristics of Murphy et al. (2019)

Study ID	**Murphy (2019)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary bundled payment program for THA)
	Selection of hospitals (intervention/control 1/control 2): BPCI participating PGPs (PGP-run)/BPCI participating hospitals (Hospital-run)/BPCI non participating hospitals
	Data source: CMS Limited Data Set (LDS) for FFS claims
	Study period (before/after): Jan 2013–Mar 2016

	Statistical analyses: - Log-transformed linear regression and logistic regression - Adjustment factors: age, race, gender, comorbidities, geography, and date of surgery
Participants	**Country/region/setting:** USA/nationwide (excluding Maryland)/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2013, the CMS Innovation Center launched the BPCI program as an alternative to traditional fee-for-service payment.
	Number of included hospitals (intervention/control 1/control 2): 74 (PGP-run)/222 (hospital-run)/NR
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedure: elective THA without major comorbidity (DRG 470)
	Number of patients included in the analysis (intervention/control 1/control 2): 44,662 PGP-run episodes/42,922 hospital-run episodes/284,002 non BPCI episodes
	Characteristics of patients (before, PGP-BPCI/hospital BPCI/Non-BPCI): Age, mean: 71.89/71.56/71.74 Sex: NR Race or ethnic group: NR Medicaid: NR Comorbidity sum, mean: 2.03/2.14/2.10
Intervention and Control	**Description of intervention:** - Participation in BPCI with either physician group practices (PGPs) or hospitals acting as episode initiators - Episode payments were assessed from the date of surgery through 90 days postoperatively - Episode initiators assumed comprehensive responsibility and financial risk for the episode of care
	Control: - Hospital-initiated BPCI episodes vs PGP-initiated BPCI episodes - Fee-for-service elective THA episodes outside BPCI used to estimate the background payment trend
Outcome(s)	**Description of outcome(s):** Primary: 90-day payment reduction Secondary: 90-day readmissions and mortality
Notes	**Funding:** Supported by Harvard Catalyst, The Harvard Clinical and Translational Science Center (BR; National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health Award UL1 TR001102), and financial contributions from Harvard University and its affiliated academic healthcare centers.
	Conflict of interest: NR
	Other comments: - Data analysis was restricted to elective THA without major comorbidity Diagnosis-Related Group (DRG) 470 episodes. - Data analysis was limited to Medicare Part A payment

Table S27. Extracted study characteristics of Navathe et al. (2017)

Study ID	**Navathe (2017)**
Methods	**Study design:** Longitudinal pre–post study with interrupted trend analysis
	Allocation: natural experiment (participation in the voluntary ACE and BPCI demonstration projects)
	Selection of hospitals (intervention): Baptist Health System (BHS) hospitals participated in ACE and BPCI programs
	Data source: - Medicare claims data - BHS internal cost data

	Study period (before/after): The study defined four study periods as follows: - ACE baseline period: July 2008 – Dec 2008 (non-bundled FFS) - ACE period: July 2009 – June 2012 (orthopedic bundles for acute hospitalization only) - Transition period: July 2012 – Sept 2013 (preparation for BPCI Model 2) - BPCI period: Oct 2013 – June 2015 (bundles spanning acute hospitalization and 30 days of PAC)
	Statistical analyses: - GLM for episode payments and hospital costs - GEE models for PAC spending - Logistic regression for quality outcomes (readmissions, ED visits, prolonged LOS) - Mann–Kendall trend tests
Participants	**Country/region/setting:** USA/Baptist Health System (BHS), a clinically integrated network of 5 hospitals in San Antonio, Texas
	Health system characteristics: Medicare payment system. CMS introduced the voluntary ACE demonstration in 2009 and expanded bundled payments through the voluntary BPCI program in 2013.
	Number of included hospitals: 5 BHS hospitals in total
	Characteristics of hospitals: NR
	Patient condition/procedure: LEJR surgery (DRG 469 and 470 e.g. MJRLE with and without major complications or comorbidities)
	Number of patients included in the analysis: 3,942 (in total)
	Characteristics of patients (ACE year 1/ACE year 2/ACE year 3/BPCI year 1/BPCI Year 2): Age, y, mean (SD): 73.2 (8.5)/72.4 (8.6)/72.1 (8.8)/72.3 (7.9)/71.9 (8.1) Male, %: 33.5/34.1/33.9/33.9/36.9 Race or ethnic group: NR Medicaid: NR Elixhauser comorbidity score, mean (SD): 1.6 (4.6)/1.5 (4.8)/1.4 (4.9)/1.4 (5.1)/1.3 (4.8)
Intervention and Control	**Description of intervention:** - Implementation of bundled payment models for MJRLE at Baptist Health System (BHS) - The intervention period includes ACE Year 3 (2009-2012), BPCI Year 1 (2013-2014), and BPCI Year 2 (2014-2015) - Under the ACE program (2009–2010), bundles covered the acute hospitalization only, covering hospital facility and physician professional fees - Under the BPCI Model 2 program (2013–2015), bundles included the acute hospitalization and 30 days of PAC, based on the arrangement selected by BHS
	Control: Nonbundled fee-for-service care before ACE participation (July–December 2008)
Outcome(s)	**Description of outcome(s):** - Hospital savings (total and within components) - PAC spending (by components) - readmissions - ER visits - prolonged LOS
Notes	**Funding:** Supported in part by The Commonwealth Fund
	Conflict of interest: Dr Emanuel is a frequent paid event speaker at numerous conventions, committee meetings and professional healthcare gatherings and is a venture partner with Oak HC/FT. Dr Navathe serves as advisor to Navvis and Company, Navigant Inc, Lynx Medical, Indegene Inc, and Sutherland Global Services; receives an honorarium from Elsevier Press; and receives grant funding from Oscar Health Insurance and Hawaii Medical Services Association. No other conflicts are reported.
	Other comments: Data were not available from the transition period between ACE and BPCI (July 2012 to September 2013)

Table S28. Extracted study characteristics of Navathe et al. (2018)

Study ID	Navathe (2018)
Methods	**Study design:** CBA
	Allocation: natural experiment (hospital BPCI participation for LEJR)
	Selection of hospitals (intervention/control): BPCI Markets/Non-BPCI Markets
	Data source: - 100% Medicare provider analysis and review files from 2011 and 2015 - Data from the American Hospital Association (AHA), Hospital Compare, and Medicare's provider of service, beneficiary summary, and accountable care organization (ACO) provider-level research-identifiable files were used to obtain market and hospital characteristics
	Study period (before/after): Jan 2011 – Sep 2013/Oct 2013 – Dec 2015
	Statistical analyses: - DiD method - Unconditional negative binomial models (for market volume) - Linear regression with hospital and quarter fixed effects (for case mix) - Analyses incorporated covariates that reflected market-level characteristics of patients (age, sex, race), hospitals (number of beds), and market structure (hospital and skilled nursing facility concentration, ACO penetration, Medicare Advantage penetration, and presence of physician group practice)
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2013, CMS initiated the voluntary BPCI program. The predominant model in BPCI is Model 2, which bundles payment for acute hospitalization and up to 90 days of PAC.
	Number of included hospitals (intervention/control): 322/1,340
	Characteristics of hospitals (intervention/control): No. of beds, median (IQR): 267 (160-401)/158 (80-278) Total admissions: 12,939/6,796 Ownership, n (%): For-profit: 67 (21)/ 278 (21) Not-for-profit: 244 (76)/876 (65) Government: 11 (4)/186 (14) Teaching status, n (%): Major: 46 (14)/85 (6) Minor: 196 (61)/561 (42) None: 80 (25)/694 (53) Urban, n (%): 306 (95)/964 (72) Region: NR
	Patient condition/procedure: LEJR (MS-DRG 469 and 470; with and without major complications or comorbidities)
	Number of patients included in the analysis: 462,161/580,043 (beneficiaries)
	Characteristics of patients (intervention/control): Age, y, mean (SD): 75 (1)/75 (1) Women, n (%): 65.1 (2.3)/64.4 (3.0) Race/ethnicity, % (SD) Black: 4.5 (4.7)/4.5 (5.8) Other: 95.5 (4.7)/95.5 (5.8) Dual-eligible, n (%): 9.1 (6.7)/8.4 (4.8) Elixhauser score, mean (SD): 1.9 (0.7)/2.0 (0.8)
Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI Model 2 - Episode includes the acute hospitalization and up to 90 days of PAC
	Control: Markets without BPCI-participating hospitals

Outcome(s)	**Description of outcome(s):** Primary: - Changes in market-level LEJR volume (quarterly from Q1 2011 – Q4 2015) - Changes in hospital-level case mix based on demographic, socioeconomic, clinical, and utilization factors (12 months prior to LEJR hospitalization)
Notes	**Funding:** Supported by The Commonwealth Fund
	Conflict of interest: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Navathe reports that he has received grant support from Hawaii Medical Service Association and Oscar Health; personal fees from Navvis and Co, Navigant Inc, Lynx Medical, Indegene Inc, and Sutherland Global Services; and honoraria from Elsevier Pressper. Dr Emanuel reports that he has received speakers fees from Partners Connected Health, North Texas Roundtable, Valence Health, AmeriHealth Carotas, athenahealth, American Health Lawyers Association, Novo Nordisk, Klick Health, Marcus Evans, Merrill Lynch, Sound Physicians, National Council for Behavioral Health, CVS Caremark, University of Michigan, 92nd Street Y, JP Morgan, CareFIRST, Ingalls, NY eHealth Collaborative, JCC Atlanta, Boston Children's Hospital, Prix Galien, JBF Food Conference, EOCI Pharmacomm, British Broadcasting Co, Mizzou Med, Florida Hospital Association, Endocrine Society, Quad Med, VHA Mid-atlantic, Wells Fargo, Cigna, Northwestern University, Minnesota Ideas Fest, ASCO, Society for Clinical Trials, Revolution, Beth Israel Hospital, Genetech, Robert Wood Johnson Foundation, Thomson Reuters, Association of Professors of Medicine, Lurie Children's Hospital, Princeton University, Hamline University, Oak Investment Partners, Verizon, Healthcare Financial Management Association, Ohio State University, Association of American Cancer Institutes, Kaiser Permanente, Optum Health, Gerontological Society of America, and McKesson; owns stock in Nuna; and is an investment partner of Oak HC/FT. No other disclosures were reported.
	Other comments: none

Table S29. Extracted study characteristics of Navathe et al. (2020)

Study ID	**Navathe (2020)**
Methods	**Study design:** CBA
	Allocation: natural experiment (BPCI participation)
	Selection of hospitals (intervention/control): BPCI-hospitals (model 2)/non-BPCI hospitals
	Data source: - Medicare claims data - American Hospital Association (AHA) - Hospital Compare - 2017 CMS IMPACT Act file
	Study period (before/after): Jan 2011–Sep 2013/Oct 2013–Dec 2016
	Statistical analyses: - Generalized linear models with a log link and gamma distribution, and ordinary least squares regression for secondary spending and binary quality outcomes - Adjustment factors: Patient demographics, clinical conditions, hospital and market characteristics, time-varying market characteristics - Propensity score matching
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2013, CMS expanded voluntary bundled payment nationwide through the BPCI initiative.
	Number of included hospitals (intervention/control): 244/244

	Characteristics of hospitals (intervention/control): No. of beds, median: 261/235 Ownership status (%): For profit: 16.4/17.2 Not for profit: 78.7/77.1 Government: 4.9/5.7 Teaching status (%): Major teaching: 12.7/13.5 Minor teaching: 34.4/31.6 Nonteaching: 52.9/54.9 Urban (%): 99.2/99.2 Region: NR
	Patient condition/procedure: LEJR
	Number of patients included in the analysis: 10,757 per quarter (early BPCI period); 10,949 per quarter (late BPCI period)/2,819 per quarter (non-BPCI hospitals, early period); 2,375 per quarter (non-BPCI patients, late period)
	Characteristics of patients (before, intervention/control): Mean age, y: 73.3/73.4 Black race (%): 7.6/8.7 Female (%): 66.1/64.4 Dually eligible (%): 15.5/13.3 Mean Elixhauser comorbidity index: 5.4/5.5
Intervention and Control	**Description of intervention:** Participation in the BPCI program Model 2 for LEJR, with bundled payments for episodes starting at hospitalization and covering up to 90 days of PAC
	Control: Hospitals that never participated in BPCI Model 2 and were located in markets with no BPCI-participating hospitals
Outcome(s)	**Description of outcome(s):** Primary: - Change in average Medicare payment per LEJR episode Secondary: - proportion of average episode spending attributable to the index hospitalization - readmissions - SNF care - IRF care - home health agency services - professional services - and all remaining components Quality outcomes: - hospital admissions - mortality rate - unplanned readmission rate - ED visit rate - LEJR-specific complication rate
Notes	**Funding:** funded by the Commonwealth Fund and the Leonard Davis Institute of Health Economics at the University of Pennsylvania.
	Conflict of interest: Amol Navathe received personal fees from Navvis Healthcare Inc., the National University Health System of Singapore, and Agathos Inc.; personal fees and equity from NavaHealth; personal fees for service as a commissioner of the Medicare Payment Advisory Commission; speaking fees from the Cleveland Clinic; and an honorarium from Elsevier Press, none of which were related to this article. He also serves without compensation as a board member of Integrated Services Inc.
	Other comments: none

Table S30. Extracted study characteristics of Navathe et al. (2021)

Study ID	**Navathe (2021)**
Methods	**Study design:** CBA
	Allocation: natural experiment (voluntary BPCI participation in model 2)
	Selection of hospitals (intervention/control): BPCI-hospitals/non-BPCI hospitals
	Data source: - data from the Health Care Cost Institute - Medicare hospital claims from the Medicare Provider Analysis and Review file - the American Hospital Association annual survey for hospital characteristics - the Medicare Provider Analysis and Review and Provider of Service files for market characteristics - Health Care Cost Institute data for patient characteristics
	Study period (before/after): Jul 2011 – Sep 2013/Oct 2013 – Sep 2016
	Statistical analyses: - DiD regression models - Adjustment factors: Patient characteristics - Propensity score matching using geographic region and baseline patient, hospital, and market characteristics from 2011
Participants	**Country/region/setting:** USA/ nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system, predominantly based on fee-for-service reimbursement, alongside Medicare Advantage (MA), which provides coverage through private health plans and does not overlap with fee-for-service Medicare. Between October 2013 and September 2018, Medicare tested the voluntary BPCI initiative as an alternative payment model.
	Number of included hospitals (intervention/control): 281/562
	Characteristics of hospitals (after matching, intervention/control): High bed size, n (%): 190 (68)/361 (64) Ownership, n (%): For-Profit: 55 (20)/123 (22) Not-for-Profit: 2017 (77)/422 (75) Teaching hospital, n (%): 41 (15)/66 (12) Urban, n (%): 280 (99.6)/560 (99.6) Region, n: Midwest 59/129 Northwest: 66/108 South: 106/230 West: 50/95
	Patient condition/procedure: LEJR (MS-DRG 469 and 470); spillover effects assessed in Medicare Advantage and commercially insured patients
	Number of patients included in the analysis (intervention/control): 68,693/116,229
	Characteristics of patients (after matching, intervention/control): Age: NR Female, n (%):41,180 (60.0)/69,106 (59.5) Race or ethnic group: NR Dually eligible for Medicare and Medicaid, n (%): 1,957 (2.9)/2,797 (2.4) Elixhauser Comorbidity Index score, mean (SD): 0.22 (4.57)/0.13 (4.2) Comorbid conditions, n (%): COPD: 8,073 (11.8)/12,623 (10.9) Obesity: 10,057 (14.6)/15,988 (13.8) Diabetes: 10,910 (15.9)/18,242 (15.7) Diabetes with complication: 5,528 (8.1)/8,177 (7.0) Hypertension: 14,751 (21.5)/21,519 (18.5)

Intervention and Control	**Description of intervention:** - Voluntary hospital participation in BPCI Model 2 - Episode includes the index hospitalization and up to 90 days of PAC - Participating hospitals assumed financial accountability for episode costs and quality - This study evaluated potential spillover effects of BPCI participation among Medicare Advantage and commercially insured patients
	Control: Non-BPCI hospitals, excluding hospitals participating in CJR and episodes attributed to BPCI physician group participants
Outcome(s)	**Description of outcome(s):** Primary: - Changes in 90-day total episode spending on LEJR episodes - 90-day readmission rate Secondary: - LEJR complication rates - index hospital LOS - PAC spending - discharge to PAC providers - hospital decisions to discharge patients to home with home health agency services
	Measurement time points: for primary outcomes: quarterly data from Q3 2011–Q3 2016
Notes	**Funding:** by the Leonard Davis Institute of Health Economics at the University of Pennsylvania
	Conflict of interest: NR
	Other comments: None

Table S31. Extracted study characteristics of Oddleifson et al. (2024)

Study ID	**Oddleifson (2024)**
Methods	**Study design:** CBA
	Allocation: natural experiment (participation in BPCI model 2)
	Selection of hospitals (intervention/control): BPCI-hospitals/same-state non-BPCI hospitals
	Data source: - American Heart Association Get With The Guidelines–Heart Failure (GWTG-HF) registry - Medicare claims data
	Study period (before/after): Nov 2008–Sep 2013/Oct 2013–Aug 2018
	Statistical analyses: - DiD with multivariable logistic regression using GEE - Adjustment factors: Year as fixed effect, patient- and hospital-level characteristics - Sensitivity analysis compared BPCI hospitals with all non-BPCI hospitals, irrespective of state
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2013, the CMS Innovation Center launched the BPCI program, which continued until 2018 and was then replaced by BPCI Advanced.
	Number of included hospitals (intervention/control): 18 BPCI hospitals (23 GWTG sites)/211 non-BPCI hospitals (224 GWTG sites)

	Characteristics of hospitals[4] (intervention/control): No. of beds, mean (SD): 314.3 (194.0)/370.8 (207.1) Ownership, n (%): Investor-owned, for-profit: 1,478 (17.0)/13,546 (14.3) Nongovernment, not-for-profit: 4,314 (49.5)/69,532 (73.6) Government, nonfederal: 657 (7.5)/7,243 (7.7) Teaching hospital, n (%): 7,090 (81.3)/70,384 (74.5) Rural location, n (%): 657 (7.5)/2,577 (2.7) Geographic region, n (%): West: 930 (10.7)/ 971 (1.0) South: 2,626 (30.1)/30,712 (32.5) Midwest: 1,125 (12.9)/12,487 (13.2) Northeast: 4,040 (46.3)/50,360 (53.3)
	Patient condition/procedure: Heart failure (HF)
	Number of patients included in the analysis (intervention/control): 8,721/94,530
	Characteristics of patients (intervention/control): Age ≤ 75y, n (%): 2,564 (29.4)/28,159 (29.8) Female, n (%): 4,750 (54.5)/49,879 (52.8) Race/ethnicity, n (%): Asian: 192 (2.2)/842 (0.9) Black: 1201 (13.8)/7,814 (8.3) Hispanic (any race): 237 (2.7)/6,248 (6.6) White: 6,836 (78.4)/76,742 (81.2) Other (includes UTD): >188 (>2.2)/1,871 (2.0) Insurance status, n (%): None: <21 (<0.2) 233 (0.3) Medicare: 5412 (62.1)/57,987 (61.3) Medicaid: 454 (5.2)/5,098 (5.4) Other: 1,921 (22.0)/20,239 (21.4)
Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI Model 2 - Episode includes acute and PAC for 30, 60, or 90 days after hospital discharge - FFS payments continued during the episode, with total episode costs reconciled against hospital-specific target prices based on historical costs minus a 2–3% discount - Hospitals shared in savings below the target and reimbursed Medicare when episode costs exceeded the target
	Control: Non-BPCI hospitals from the same states as BPCI-participating hospitals
Outcome(s)	**Description of outcome(s):** Primary: - Evidence-based β-blocker at discharge - angiotensin-converting enzyme inhibitors/angiotensin receptor blockers (ACEIs/ARBs) or angiotensin receptor-neprilysin inhibitor (ARNI) at discharge - aldosterone antagonist at discharge - cardiacresynchronizationtherapy (CRT) defibrillator or CRT pacemaker placed or prescribed at discharge - implantable cardioverter-defibrillator (ICD) counseling or ICD placed or prescribed at discharge - heart failure education - follow-up visit within 7 days Secondary: - LOS longer than 4 days - referral to cardiac rehabilitation - in-hospital mortality - discharge to hospice or home hospice - discharge to skilled nursing facility

[4] Numbers for hospital characteristics indicate the proportion of patients treated in hospitals with the respective characteristic

	- 30-day and 90-day all-cause readmission rate - 30-day and 90-day all-cause mortality rate (without in-hospital deaths)
Notes	**Funding:** The Get With The Guidelines–Heart Failure (GWTG-HF) program is provided by the American Heart Association. GWTG-HF is sponsored in part by Novartis, Boehringer Ingelheim, Novo Nordisk, AstraZeneca, Bayer, Tylenol, and Alnylam Pharmaceuticals.
	Conflict of interest: Dr Wadhera reported consultant fees from Chambercardio, CVS Health, and Abbott and research support from the National Heart, Lung, and Blood Institute (NHLBI) outside the submitted work. Dr Allen reported consultant fees from ACI Clinica, Boston Scientific, Cytokinetics, Medscape, Novartis, Quidel, Story Health, and UpToDate during the conduct of the study; grants from the Patient-Centered Outcomes Research Institute and NHLBI; and serving as an editor for the American Heart Association outside the submitted work. Dr Greene reported research support from the Duke University Department of Medicine Chair's Research Award, American Heart Association (929502), Amgen, AstraZeneca, Bristol Myers Squibb, Cytokinetics, Merck, Novartis, Pfizer, and Sanofi; serving on advisory boards or as consultant for Amgen, AstraZeneca, Bayer, Boehringer Ingelheim/Lilly, Bristol Myers Squibb, Corteria, CSL Vifor, Cytokinetics, Lexicon, Roche Diagnostics, Merck, PharmaIN, Sanofi, scPharmaceuticals, Tricog Health, and Urovant Pharmaceuticals; and speaker fees from Bayer, Boehringer Ingelheim, Cytokinetics, Lexicon, and Roche Diagnostics. Dr Fonarow reported consultant fees from Abbott, Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Cytokinetics, Eli Lilly, Johnson & Johnson, Medtronic, Merck, Novartis, and Pfizer outside the submitted work. Dr Spatz reported grant funding from the Centers for Disease Control and Prevention (20042801-Sub01), National Institute on Minority Health and Health Disparities (U54MD010711-01), US Food and Drug Administration to support projects within the Yale-Mayo Clinic Center of Excellence in Regulatory Science and Innovation (CERSI, U01FD005938), National Institute of Biomedical Imaging and Bioengineering (R01 EB028106-01), and NHLBI (R01HL151240). Dr Desai reported being under contract with the Centers for Medicare & Medicaid Services to develop and maintain performance measures used for public reporting and pay for performance programs and reported research grants and consulting for Amgen, AstraZeneca, Bristol Myers Squib, Boehringer Ingelheim, Cytokinetics, Merck, MyoKardia, Relypsa, Novartis, scPharmaceuticals, and Vifor outside the submitted work. No other disclosures were reported.
	Other comments: None

Table S32. Extracted study characteristics of Rolnick et al. (2020)

Study ID	**Rolnick (2020)**
Methods	**Study design:** CBA
	Allocation: natural experiment (hospital participation in BPCI model 2)
	Selection of hospitals (intervention/control): BPCI hospitals/matched non-BPCI hospitals
	Data source: Medicare claims data
	Study period (before/after): Jan 2011–Sep 2013/Oct 2013–Dec 2016
	Statistical analyses: - DiD models, generalized linear models with various distributions depending on the outcome - Adjustment factors: age, sex, race, disability status, comorbidities, diagnosis-related group, market characteristics, ACO penetration, Medicare Advantage penetration
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. BPCI applied to traditional Medicare, a national government program covering older (≥65 years) or disabled Americans. Traditionally, services were paid separately without adjustment for total episode spending.
	Number of included hospitals (intervention/matched control): 226/700

	Characteristics of hospitals (after matching, intervention/control): No. of hospital beds, mean (SD): 316 (238)/292 (220) Ownership, n (%): For profit: 94 (18.3)/256 (17.5) Government: 15 (2.9)/50 (3.4) Not-for-profit: 405 (78.8)/1,154 (79.0) Teaching status, n (%): Major teaching: 58 (11.3)/155 (10.6) Minor teaching: 199 (38.7)/559 (38.3) Non-teaching: 257 (50.0)/746 (51.1) Urban, n (%): 476 (92.6)/1,335 (91.4) Geographic region (not used in propensity score matching), n (%): Midwest: 41 (18.1)/222 (31.7) North east: 74 (32.7)/69 (9.9) South: 74 (32.7)/230 (32.9) West: 37 (16.4)/179 (25.9)
	Patient condition/procedure: AMI (MS-DRGs 280-282), CHF (MS-DRGs 291-293), COPD (MS-DRGs 190-192, 202, 203), and pneumonia (MS-DRGs 177-179 and 193-195)
	Number of patients included in the analysis (intervention/control): baseline period: 213,853/166,884 (beneficiaries) treatment period: 82,842/68,871
	Characteristics of patients (Baseline period, intervention/control): Age (y), mean (SD): 77.2 (12.3)/ 75.7 (12.7) Men, n (%): 115,853 (44.4)/94,069 (44.5) Race/ethnicity, n (%): White: 216,355 (82.8)/169,623 (80.3) Black: 30,979 (11.9)/32,475 (15.4) Other: 13,829 (5.3)/9,110 (4.3) Elixhauser index score, mean (SD): 19.4 (13.8)/19.4 (13.8) Dual eligibility, n (%): 69,540 (26.6)/55,786 (26.4) Disabled, n (%): 33,122 (12.7)/32,850 (15.6)
Intervention and Control	**Description of intervention:** - Voluntary hospital participation in BPCI Model 2 - Episode includes the index hospitalization and extends for up to 90 days after discharge - The participating hospital assumed accountability for total episode spending and quality outcomes - Hospitals received a bonus or penalty based on episode spending relative to their historical episode spending
	Control: Non-BPCI hospitals from markets with no BPCI hospital participants
Outcome(s)	**Description of outcome(s):** Primary: - Total spending on episodes - Death 90-days after discharge Secondary: - 90-day readmission - 90-day ED visit - PAC use - Index discharge to SNF and number of SNF days - Spending by category (hospital, all institutional care, readmission, outpatient professional fees, durable medical equipment, SNF care, home health services, outpatient hospital care)
	Measurement time points: quarterly data from Q1 2011 to Q3 2016
Notes	**Funding:** Funded by the Commonwealth Fund (grant No 20170816)
	Conflict of interest: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: support from the Commonwealth Fund for the submitted work; JAR reported consulting fees from Tuple Health. ASN reported receiving grants from Hawaii Medical Service Association, Anthem Public Policy Institute, the Healthcare Research and Education Trust, Cigna, Ochsner Health System, United

	Healthcare Blue Cross Blue Shield of NC, and Oscar Health; personal fees from Navvis Healthcare, Navigant, National University Health System – Singapore, Singapore Ministry of Health, the Medicare Payment Advisory Commission, and Agathos; personal fees and equity from NavaHealth; equity from Embedded Healthcare; speaking fees from the Cleveland Clinic; serving as a board member of Integrated Services. without compensation, and an honorarium from Elsevier Press, none of which are related to this manuscript. Since 2016, EJE reports speaker's fees from Tanner Healthcare System, Mid-Atlantic Permanente Group, American College of Radiology, Marcus Evans, Loyola University, Oncology Society of New Jersey, Good Shepherd Community Care, Remedy Partners, Medzel, Kaiser Permanente Virtual Medicine, Wallace H Coulter Foundation, Lake Nona Institute, Allocation, Partners Chicago, Pepperdine University, Huron, American Case Management Association, Philadelphia Chamber of Commerce, Blue Cross Blue Shield Minnesota, UnitedHealth Group, Futures Without Violence, CHOP, Washington State Hospital Association, Association of Academic Health Centers, Blue Cross Blue Shield of Massachusetts, American Academy of Ophthalmology, Lumeris, Roivant Sciences, Medical Specialties Distributors, Vizient University Healthcare System, Center for Neuro-Degenerative Research, Colorado State University, Genentech Oncology, Council of Insurance Agents and Brokers, Grifols Foundation, America's Health Insurance Plans, Montefiore Physician Leadership Academy, Greenwall Foundation, Medical Home Network, HFMA, Ecumenical Center – UT Health, American Association of Optometry, Associação Nacional de Hospitais Privados, National Alliance of Healthcare Purchaser Coalitions, Optum Laboratories, Massachusetts Association of Health Plans, District of Columbia Hospital Association, Washington University. Since January 2017, EJE has been a venture partner with Oak HC/FT, a firm that invests in health services but not pharmaceuticals or devices. EJE is on the board of two start-ups: Village MD and Oncology Analytics. EJE has investments in Gilead, Allergan, Amgen, Baxter, and UnitedHealth Group.
	Other comments: None

Table S33. Extracted study characteristics of Shashikumar et al. (2023)

Study ID	**Shashikumar (2023)**
Methods	**Study design:** CBA
	Allocation: natural experiment (hospital participation in BPCI-A)
	Selection of hospitals (intervention/control): BPCI-A participating hospitals/matched non-BPCI-A participating control hospitals
	Data source: - Medicare claims data - 2017 American Hospital Association Annual Survey and Area Health Resources File
	Study period (before/after): Jan 2017–Sep 2018/Oct 2018–Sep 2019
	Statistical analyses: - DiD models - one-to-one matching algorithm - Adjustment factors: Age, sex, primary diagnosis-related group, comorbidities, disability status
Participants	**Country/region/setting:** USA/ nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. In 2018, CMS launched the voluntary Bundled Payments for Care Improvement Advanced (BPCI-A) model as an alternative payment model.
	Number of included hospitals (intervention/matched control): 540/540

	Characteristics of hospitals (after matching, intervention/control): No. of beds: NR Profit status (%): For-profit: 29/23 Non-profit: 65/72 Public (%): 5/5 Teaching hospital (%): 59/63 Urban (%): 99/98 Geography (%): Northeast: 19/19 Midwest: 23/23 South: 39/36 West: 19/22
	Patient condition/procedure: cardiovascular events (incl. AMI, arrhythmia, and CHF) or cardiac procedures (incl. CABG, defibrillator implantation, pacemaker insertion, PCI, and valve procedures)
	Number of patients included in the analysis (intervention/control): 389,349/252,258
	Characteristics of patients (before, intervention/control): **Cardiac medical bundles:** Age, n (%): <65 y: 27,565 (8.5)/19,419 (8.8) 65–80 y: 131,245 (41.0)/90,881 (41.3) >80 y: 164,131 (50.8)/109,717 (49.9) Female, n (%): 175,019 (54.2)/118,210 (53.7) Race/ethnicity, n (%): White: 252,949 (78.3)/177,118 (80.5) Black: 39,340 (12.2)/24,805 (11.3) Hispanic: 18,984 (5.9)/11,193 (5.1) Asian/Pacific Islander, American Indian/Alaska Native, Other, and Unknown: 11,668 (3.6)/6,901 (3.1) Medicaid, n (%):83 697 (25.9)/57,440 (26.1) Disabled, n (%): 71,267 (22.1)/ 50,292 (22.9) Complications, n (%): Major: 193,102 (59.8)/131,689 (59.9) Minor: 44,461 (13.8)/29,721 (13.5) No complication: 85,378 (26.4)/58,607 (26.6) **Cardiac procedural bundles:** Age, n (%): <65 y: 5,788 (8.7)/3,275 (10.2) 65–80 y: 41,055 (61.8)/19,608 (60.8) >80 y: 19,565 (29.5)/9,358 (29.0) Female, n (%): 24,976 (37.6)/12,328 (38.2) Race/ethnicity, n (%): White: 55,774 (84.0)/27,466 (85.2) Black: 3,904 (5.9)/1,983 (6.2) Hispanic: 3,335 (5.0)/1,526 (4.7) Asian/Pacific Islander, American Indian/Alaska Native, Other, and Unknown: 3,395 (5.1%)/1,266 (3.9) Medicaid, n (%): 11,005 (16.6)/5,736 (17.8) Disabled, n (%): 13,245 (19.9)/7,098 (22.0) Complications, n (%): Major: 22,948 (34.6)/10,622 (32.9) Minor: 36,453 (54.9)/18,672 (57.9) No complication: 7,007 (10.6)/2,947 (9.1)

Intervention and Control	**Description of intervention:** - Voluntary participation in BPCI-A - Episode includes the index hospitalization and extends for 90 days after discharge - Hospitals were assigned target spending prices for episodes - Hospitals received a financial bonus when episode spending was below the target and paid a penalty to CMS when spending exceeded the target
	Control: Matched BPCI-A-non-participating hospitals
Outcome(s)	**Description of outcome(s):** Primary: - Changes in mean episode spending for the aggregate of cardiac medical and cardiac procedural bundles Secondary: - Episode spending (overall and by individual payment components) - Selection (proportion of medically or socially complex patients) - Care utilization (PAC use and LOS) - Quality measures (90-day mortality, readmissions, and healthy days at home)
Notes	**Funding:** Funded by the National Institutes of Health National Heart, Lung, and Blood Institute (R01HL143421) and National Institute on Aging (R01AG060935)
	Conflict of interest: Dr Shashikumar has received grants from the Sarnoff Cardiovascular Research Foundation outside the submitted work. Dr Orav receives grants from the National Heart, Lung, and Blood Institute and the National Institute on Aging. Dr Epstein receives grants from the National Heart, Lung, and Blood Institute, the National Institute on Aging, and the Commonwealth Fund. Dr Joynt Maddox receives grants from the National Heart, Lung, and Blood Institute, the National Institute on Aging, and the National Institute of Nursing Research; participates in research projects with Humana; and receives personal fees from the Centene Corporation (St. Louis, MO) for service on the health policy advisory committee. Dr Zheng reports no conflicts.
	Other comments: None

Table S34. Extracted study characteristics of Stern et al. (2024)

Study ID	**Stern (2024)**
Methods	**Study design:** CBA
	Allocation: natural experiment (mandatory CJR participation)
	Selection of hospitals (intervention/control): Hospitals in MSAs randomly assigned to mandatory CJR participation/Non-CJR hospitals
	Data source: Medicare claims data (Inpatient, Outpatient, Carrier, Home Health, Skilled Nursing Facility claims files)
	Study period (before/after): Jan 2013–Sep 2015/Oct 2016–Sep 2019
	Statistical analyses: - Generalized linear models using DiD - Adjustment factors: Age, sex, race, comorbidities (using the Charlson–Deyo comorbidity index score)
Participants	**Country/region/setting:** USA/nationwide/acute and post-acute care
	Health system characteristics: Medicare payment system. CJR was implemented on April 1, 2016 as a mandatory bundled-payment model for hospitals in selected MSAs.
	Number of included hospitals (intervention/control): 34/42 (MSAs)
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedure: elective THA or TKA
	Number of patients included in the analysis (intervention/control): THA: 3,110/3,217 (episodes) TKA: 5,072/5,692 (episodes)

	Characteristics of patients (before, intervention/control): **THA patients:** Age, median (IQR): 72 (68-76)/71 (67-75) Female, n (%): 567 (55.0)/633 (55.0) Race (non-White), n (%): 66 (6.4)/85 (7.4) Charlson−Deyo Comorbidity Index score, n (%) 0: 431 (41.8)/503 (43.7) 1: 227 (22.0)/262 (22.7) 2: 187 (18.1)/192 (16.7) ≥3: 186 (18.0)/195 (16.9) **TKA patients:** Age, median (IQR): 71 (67-76)/71 (67-75) Female, n (%): 1,098 (59.5)/1,432 (58.8) Race (non-White), n (%): 184 (10.0)/204 (8.4) Charlson-Deyo Comorbidity Index score, n (%) 0: 742 (40.2)/922 (37.9) 1: 454 (24.6)/637 (26.2) 2: 322 (17.5)/388 (15.9) ≥3: 327 (17.7)/ 87 (20.0)
Intervention and Control	**Description of intervention:** - Mandatory participation in CJR - Episode includes the inpatient hospital admission and extends through 90 days post-discharge - The bundle included post-discharge care across inpatient post-acute facilities, home health, and outpatient settings
	Control: Acute care IPPS hospitals located in 42 control MSAs
Outcome(s)	**Description of outcome(s):** Primary: - Changes in the use of home health physical therapy (PT) services within 90 days - Changes in the use of Outpatient PT services within 90 days - total number of per-episode home health, outpatient, and composite community-based PT visits within 90 days
	Measurement time points: Quarterly data from Q1 2013 to Q3 2019
Notes	**Funding:** NR
	Conflict of interest: Dr Hayden reports consulting fees from Depuy/Johnson & Johnson and Heraeus outside of this work. Dr Chen reports consulting fees for Stryker and PolarisAR outside of this work. The other authors have nothing to disclose.
	Other comments: None

Table S35. Extracted study characteristics of Whaley et al. (2021)

Study ID	**Whaley (2021)**
Methods	**Study design:** CBA
	Allocation: natural experiment (based on the differential implementation of the direct payment program among self-insured employers)
	Selection of hospitals (intervention/control): Hospitals that delivered procedures reimbursed through the direct-payment bundled-care program/Hospitals that performed the same procedures for patients insured by employers not participating in the program
	Data source: Medicare claims data
	Study period (before/after): 2014–2016/2016–2020
	Statistical analyses: - Multivariate regression models - Adjustment factors: Patient characteristics (comorbidity score, age, sex), employer fixed-effects, DRG fixed-effects, time trends

Participants	**Country/region/setting:** USA/nationwide/acute and perioperative care across hospitals and ambulatory surgical centers
	Health system characteristics: In the United States, FFS is the predominant reimbursement model. Medicare has introduced bundled payment models such as BPCI and CJR. In the commercially insured market, high and variable provider prices created incentives for self-insured employers to adopt direct employer-to-provider bundled payment models aimed at reducing spending by directing patients toward lower-priced, high-quality providers.
	Number of included hospitals (intervention/control): 15/NR
	Characteristics of hospitals (intervention/control): NR
	Patient condition/procedures: spinal fusion (DRG 455), major joint replacement (DRG 470), and bariatric surgery (DRG 621)
	Number of patients included in the analysis (pre/post non-direct payment/post direct payment): 1,638/577/157 (procedures)
	Characteristics of patients (pre/post non-direct payment/post direct payment) Age, mean (SD): 52.0 (10.1)/53.6 (9.5)/51.2 (10.3) Percent male, mean (SD): 34.4 (47.5)/39.0 (48.8)/36.3 (48.2) Race or ethnic group: NR Medicaid: NR Comorbidities or chronic conditions: NR
Intervention and Control	**Description of intervention:** - Direct employer-to-provider prospective bundled payment program developed by Carrum Health and implemented among self-insured employers - Episode covers the surgical procedure, 3 days of preoperative care, and at least 30 days of postoperative care, including treatment of related complications and readmissions - Patients receiving care from participating providers had no cost sharing - Participating providers were selected based on quality criteria and assumed financial responsibility for the bundled episode
	Control: Acute and perioperative care across hospitals and ambulatory surgical centers
Outcome(s)	**Description of outcome(s):** Changes in episode costs (incl. episode prices, Employer paid amount, and patient cost-sharing)
	Measurement time points: >2 years before, 2 years before, 1 year after, 2+ years after
Notes	**Funding:** National Institute on Aging (Grant No. 1K01AG061274 to Christopher Whaley)
	Conflict of interest: Dena Bravata and Christoph Dankert are shareholders of Carrum Health. Dena Bravata received payment for consulting services.
	Other comments: None

Supplement 3. Characteristics of included payment reforms

Table S36. Description of bundled payment programs included in the review

Program	BPCI (Model 2)	CJR Model	BPCI-A	ACE Demonstration	Employer-Provider Direct Payment (Carrum Health)
Country	USA	USA	USA	USA	USA
Implementation period	2013–2018	2016–2024	2018–2025	2009–2011	2016–2020
Hospital payment base before reform	MS-DRGs	MS-DRGs	MS-DRGs	MS-DRGs	NR
Description of Modification	Retrospective bundled payment for hospitalization + 90 days post-discharge	Retrospective bundled payment model for LEJR	Retrospective bundled payment model with prospective target prices	Prospective bundled/global payment, covering hospital and physician	Direct payment program for self-insured employers with prospective bundled payment contracts

				services during the inpatient stay	negotiated with selected providers
Bundled scope/ coverage	48 clinical inpatient episodes; episode includes index hospitalization, post-acute care, readmissions, and related services for 30, 60, or 90 days post-discharge	LEJR episodes (incl. inpatient hip, knee, and ankle replacements, and outpatient hip/knee replacements from PY6); episode extends 90 days after discharge or outpatient procedure	8 clinical episode service lines groups with 29 inpatient, 3 outpatient and 2 multi-setting clinical episode categories; episode includes index admission + 90 days post-discharge	28 cardiovascular and 9 orthopedic inpatient episodes; episode includes Part A and Part B services (incl. physician services, pertaining to the inpatient stay)	Selected elective and outpatient procedures (e.g. joint replacement, colonoscopy, cardiac procedures); episode includes preoperative, surgical, and postoperative services within a 33-day perioperative period
Payment flow	FFS payments during episode; retrospective reconciliation against target price	FFS payments during episode; retrospective reconciliation against target price	FFS payments during episode; retrospective reconciliation against target price	Prospective global payment	Prospective bundled payment contracts
Mandatory/ Voluntary	Voluntary	Mandatory for selected MSAs	Voluntary	Voluntary	Voluntary
Level of application	Hospital or physician group practice as episode initiator	Hospital-level	Hospital or physician group practice as episode initiator	Hospital-level (in coordination with physicians)	Employer-provider contract
Payer/program administrator	Medicare/CMS	Medicare/CMS	Medicare/CMS	Medicare/CMS	Self-insured employers via Carrum Health
Provider(s)	Hospitals, physician group practices	Hospitals	Hospitals, physician group practices	Affiliation between at least one hospital and at least one physician group	Participating providers incl. academic medical centers, large for-profit health systems, faith-based nonprofit health systems, and surgeon groups
Risk sharing	Two-sided risk	Two-sided risk	Two-sided risk with stop-loss/stop-gain limits	Shared savings/ gainsharing permitted	Providers at risk for postoperative complications and readmissions
Adjustment mechanism	Target price based on historical FFS payments, including a discount	Target prices based on historical hospital-specific and regional spending, with regional component increasing over time	Target prices adjusted for patient case mix, peer group historical spending, peer group trend, and CMS discount; payment tied to quality measures	Competitive bids for each listed DRG; discount off base DRG payment amount	Negotiated prospective bundled payment contracts; provider screening based on facility and surgeon quality criteria
Primary objective(s)	Reduce Medicare costs, improve care coordination and quality	Improve quality and coordination of care, reduce episode spending, maintain patient choice	Reduce cost and improve care coordination; expand value-based care and quality incentives	Improve coordination, quality, and efficiency of care	Reduce costs and improve outcomes through provider quality screening, patient support, appropriateness screening, and reduced patient cost sharing

Table S36 summarizes key design features of bundled payment models evaluated in the included studies. Information was extracted from the included studies and, for CMS models, supplemented with official CMS program descriptions.

Table S37. Description of DRG system refinements included in the review

Program	MS-DRG introduction
Country	USA
Implementation period	FY 2008; effective since October 1, 2007–present
Hospital payment base before reform	DRGs under Medicare PPS/IPPS
Description of Modification	Replacement of DRGs with Medicare Severity (MS-)DRGs, increasing differentiation by severity of illness
Scope/coverage	Medicare inpatient hospital stays reimbursed under IPPS
Payment flow	Prospective
Mandatory/voluntary	Mandatory
Level of application	National
Payer/program administrator	Medicare/CMS
Provider(s)	Acute care hospitals treating Medicare beneficiaries
Risk sharing	Hospitals bear financial risk if actual inpatient costs exceed the fixed prospective payment
Adjustment mechanism	Severity-based classification using CC/MCC subgroups and MS-DRG relative weights
Primary objective(s)	Improve payment accuracy by better reflecting patient severity, complexity, and expected resource use

Table S37 summarizes key design features of the introduction of MS-DRGs. Information was extracted from the included study and supplemented with an official terminology source.

Table S38. Description of combined patient choice and bundled payment reform included in the review

Program	Patient choice and bundled payment reform
Country/region	Sweden/Stockholm
Implementation period	January 2009
Hospital payment base before reform	DRGs
Description of Modification	Reform combining three elements: (1) patient free choice of provider, (2) free entry of accredited new providers, and (3) bundled payment for elective total hip and knee replacement
Scope/coverage	Low-risk patients (ASA 1–2) undergoing elective total hip or knee replacement; bundle covered diagnostics, surgery, postoperative care/follow-up, implant costs, and complications/adverse events within 2 years
Payment flow	Lump-sum payment per patient; a small performance-based component was retained and paid only if quality targets were met
Mandatory/ voluntary	Mandatory regional reform; provider participation through accreditation
Level of application	Regional level (implemented by Region Stockholm Healthcare Administration)
Payer/program administrator	Region Stockholm (single public payer)
Provider(s)	Authorized providers, incl. region-owned hospitals and accredited private specialized centers
Risk sharing	Providers bear financial responsibility for complications within 2 years (“complication guarantee”)
Adjustment mechanism	Fixed bundled rate per patient (no patient-level risk adjustment reported); accreditation and quality reporting required; performance-based retention tied to quality indicators
Primary objective(s)	Improve quality and efficiency through competition, reduce waiting times, increase patient choice and access

Table S38 summarizes key design features of the patient choice reform in Region Stockholm as described in the included studies.

Supplement 4. Outcome-based study findings

Table S39. BPCI (Model 2) compared to standard traditional Medicare payment

Outcome domain	Study	Condition/ procedure	Follow-up	Effect		n hospitals (i/c)
Mortality	Dummit (2016)	LEJR	30d	aDiD −0.1 pp (95% CI −0.3 to 0.2)	↔	176/841
Mortality	Dummit (2016)	LEJR	90d	aDiD −0.0 pp (95% CI −0.3 to 0.3)	↔	176/841
Mortality	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	30d	Across combined conditions: aDiD –0.2 pp (p=0.64)	↔	492/898
Mortality	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD 0.3 pp (p=0.52)	↔	492/898
Mortality	Liao (2022)	Medical episodes	90d	aDiD −1.1 pp (95% CI −1.8 to −0.5)	↓	379/ 1,441[5]
Mortality	Liao (2022)	Surgical episodes	90d	aDiD −0.7 pp (95% CI −0.9 to −0.4)	↓	379/ 1,441[5]
Mortality	Maughan (2019)	13 conditions (orthopedic, cardiovascular, pulmonary) in vulnerable populations (dually eligible, dementia, recent institutional care)	90d	No significant changes across conditions/ subgroups (changes reported separately by condition/ subgroup)	↔ CI/p-values NR	1,410/ 7,932[6]
Mortality	McNeely (2020)	PCI	30d	aDiD 0.8 pp (p=0.170)	↔	42/115
Mortality	McNeely (2020)	PCI	90d	aDiD 1.1 pp (p=0.099)	↔	42/115
Mortality	McNeely (2020)	CABG	30d	aDiD 0.2 pp (p=0.691)	↔	46/122
Mortality	McNeely (2020)	CABG	90d	aDiD 0.4 pp (p=0.554)	↔	46/122
Mortality	Murphy (2019)	THA (uncomplicated)	90d	OR 1.12 (95% CI 0.88 to 1.41)[7]	↔	222/NR
Mortality	Navathe (2020)	LEJR	90d	Adj. diff. change –0.15 pp (95% CI –0.5 to 0.2)	↔	244/244
Mortality	Oddleifson (2024)	Heart failure	in-hospital	aOR 0.67 (95% CI 0.51 to 0.86)	↓	23/224

[5] The extracted effect refers only to the comparison between BPCI hospitals and non-BPCI providers, as this reflects our primary review question. Accordingly, the number of participating providers refers to BPCI hospitals and non-BPCI providers only; BPCI physician group practices analyzed in the study were not considered.
[6] Reported numbers are clinical department-level counts rather than distinct hospital counts. Because overlap across departments within the same hospital could not be determined, they may overestimate the number of unique hospitals.
[7] The study did not report formal DiD estimates; reported effects reflect simple before-after comparisons within the intervention group only.

Mortality	Oddleifson (2024)	Heart failure	30d	Adj. rate difference 0.02 (95% CI −0.04 to 0.07)	↔	23/224
Mortality	Oddleifson (2024)	Heart failure	90d	Adj. rate difference 0.05 (95% CI −0.08 to 0.17)	↔	23/224
Mortality	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD −0.1 pp (95% CI −0.5 to 0.2)	↔	226/700
Readmissions	Dummit (2016)	LEJR	30d	aDiD −0.1 pp (95% CI –0.6 to 0.4)	↔	176/841
Readmissions	Dummit (2016)	LEJR	90d	aDiD –0.4 pp (95% CI –1.1 to 0.3)	↔	176/841
Readmissions	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	30d	Across combined conditions: aDiD –0.1 pp (p=0.79)	↔	492/898
Readmissions	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD 0.0 pp (p=0.97)	↔	492/898
Readmissions	Liao (2022)	Medical episodes	90d	aDiD –0.8 pp (95% CI –1.6 to –0.02)	↓	379/ 1441[8]
Readmissions	Liao (2022)	Surgical episodes	90d	aDiD 0.1 pp (95% CI –0.42 to 0.7)	↔	379/ 1441[8]
Readmissions	Martin (2018)	lumbar fusion	90d	Percentage change (risk-bearing vs control): +2.7% vs. –10.7% (p=0.043)	↑	102[9]/ 1,277
Readmissions	Maughan (2019)	Dementia + MJRLE fracture	90d	DiD −3.0 pp; Significant decline reported	(↓); CI/p-value NR	1,410/ 7,932[10]
Readmissions	Maughan (2019)	13 conditions (orthopedic, cardiovascular, pulmonary) in vulnerable populations (dually eligible, dementia, recent institutional care)	90d	No meaningful differences for other clinical conditions than MJRLE fracture (changes reported separately by condition/ subgroup)	↔ CI/p-value NR	1,410/ 7,932[10]
Readmissions	McNeely (2020)	PCI	30d	aDiD 0.1 pp (p=0.912)	↔	42/115
Readmissions	McNeely (2020)	PCI	90d	aDiD –1.1 pp (p=0.367)	↔	42/115
Readmissions	McNeely (2020)	CABG	30d	aDiD 0.1 pp (p=0.939)	↔	46/122
Readmissions	McNeely (2020)	CABG	90d	aDiD 0.4 pp (p=0.785)	↔	46/122

[8] The extracted effect refers only to the comparison between BPCI hospitals and non-BPCI providers, as this reflects our primary review question. Accordingly, the number of participating providers refers to BPCI hospitals and non-BPCI providers only; BPCI physician group practices analyzed in the study were not considered.
[9] Risk-bearing hospitals only.
[10] Reported numbers are clinical department-level counts rather than distinct hospital counts. Because overlap across departments within the same hospital could not be determined, they may overestimate the number of unique hospitals.

Readmissions	Murphy (2019)	THA (uncomplicated)	90d	OR 0.98 (95% CI 0.92 to 1.04)[11]	↔	222/NR
Readmissions	Navathe (2020)	LEJR	90d	Adj. diff. change 0.15 pp (95% CI –0.5 to 0.8)	↔	244/244
Readmissions	Navathe (2021)	LEJR	90d	aDiD −0.47 pp (95% CI −1.0 to 0.06)	↔	281/562
Readmissions	Oddleifson (2024)	Heart failure	30d	Adj. rate difference −0.04 (95% CI −0.27 to 0.18)	↔	23/224
Readmissions	Oddleifson (2024)	Heart failure	90d	Adj. rate difference 0.45 (95% CI −0.40 to 1.30)	↔	23/224
Readmissions	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD 0.1 pp (95% CI –0.5 to 0.6)	↔	226/700
ED visits	Dummit (2016)	LEJR	30d	aDiD –0.1 pp (95% CI –0.7 to 0.5)	↔	176/841
ED visits	Dummit (2016)	LEJR	90d	aDiD 0.2 pp (95% CI –0.6 to 1.0)	↔	176/841
ED visits	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	30d	Across combined conditions: aDiD 1.0 pp (p=0.23)	↔	492/898
ED visits	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD 0.9 pp (p=0.44)	↔	492/898
ED visits	Maughan (2019)	13 conditions (orthopedic, cardiovascular, pulmonary) in vulnerable populations (dually eligible, dementia, recent institutional care)	90d	No significant changes reported across conditions/ subgroups (changes reported separately by condition/ subgroup)	↔ CI/p-value NR	1,410/ 7,932[12]
ED visits	McNeely (2020)	PCI	30d	Increased ED use: aDiD 7.7 pp (p=0.010)	↑	42/115
ED visits	McNeely (2020)	PCI	90d	Increased ED use: aDiD 7.2 pp (p=0.045)	↑	42/115
ED visits	McNeely (2020)	CABG	30d	Reduced ED use: aDiD –9.5 pp (p=0.001)	↓	46/122
ED visits	McNeely (2020)	CABG	90d	Reduced ED use: aDiD –12.1 pp (p=0.002)	↓	46/122

[11] The study did not report formal DiD estimates; reported effects reflect simple before-after comparisons within the intervention group only.

[12] Reported numbers are clinical department-level counts rather than distinct hospital counts. Because overlap across departments within the same hospital could not be determined, they may overestimate the number of unique hospitals.

ED visits	Navathe (2020)	LEJR	90d	Adj. diff. change −0.19 pp (95% CI –1.0 to 0.6)	↔	244/244
ED visits	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD –0.2 pp (95% CI –0.7 to 0.3)	↔	226/700
Complications (LEJR complication rate)	Navathe (2020)	LEJR	90d	Adj. diff. change 0.12 pp (95% CI –0.3 to 0.6)	↔	244/244
Complications (LEJR complication rate)	Navathe (2021)	LEJR	NR	No difference between hospital groups (estimate shown graphically only; not reported numerically)	↔	281/562
Complications (repeat lumbar spine surgery)	Martin (2018)	Lumbar fusion	NR	Percentage change (risk-bearing vs control): +30.6% vs. +7.1% (p=0.043)	↑	102[13]/ 1,277
Complications (device complications, wound problems, and life-threatening complications)	Martin (2018)	Lumbar fusion	NR	Changes in rates were similar and not statistically significant (reported narratively only)	↔ CI/p-value NR	102[13]/ 1,277
LOS	Dummit (2016)	LEJR	in-hospital	aDiD: −0.1 d (95% CI −0.2 to 0.0)	↔	176/841
LOS	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	in-hospital	Across combined conditions: DiD 0.0 d (p=0.92)	↔	492/898
LOS	McNeely (2020)	PCI	in-hospital	aDiD –0.04 d (p=0.760)	↔	42/115
LOS	McNeely (2020)	CABG	in-hospital	aDiD 0.3 d (p=0.277)	↔	46/122
LOS	Navathe (2021)	LEJR	in-hospital	aDiD −3.2% (95% CI −5.8% to −0.71%)	↓	281/562
LOS	Oddleifson (2024)	Heart failure	in-hospital	Prolonged LOS >4d aOR 0.95 (95% CI 0.83 to 1.09)	↔	23/224
Risk selection (case mix changes among 19 demographic, socio-economic, clinical, and prior utilization factors)	Navathe (2018)	LEJR	index admission	no significant associations among demographic, socioeconomic, clinical, and prior utilization factors except for prior SNF utilization	↔	322/ 1,340

[13] Risk-bearing hospitals only.

Risk selection (prior SNF use)	Navathe (2018)	LEJR	index admission	aDiD −0.53 pp (95% CI −0.96 to −0.10)	↓	322/ 1,340
Risk selection	McNeely (2020)	PCI or CABG	index admission	No changes in patient selection (changes in patient characteristics and risk factors reported separately)	↔	42/115 (PCI) 46/122 (CABG)
Risk selection (demographic/ social risk)	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	index admission	No statistically significant differences in social determinants or case complexity	↔	492/898
Total episode spending	Chopra (2023)	LEJR	30d	aDiD −$541.6 (95% CI −718.0 to −365.3)	↓	312/ 1,977
Total episode spending	Chopra (2023)	Colectomy	30d	aDiD −$582.1 (95% CI −927.3 to −236.8)	↓	312/ 1,977
Total episode spending	Chopra (2023)	CABG	30d	aDiD −$268.9 (95% CI −831.5 to 293.7)	↔	312/ 1,977
Total episode spending	Dummit (2016)	LEJR	90d	aDiD −$1,166 (95% CI −1,634 to −699)	↓	176/841
Total episode spending	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD $112 (p=0.79)	↔	492/898
Total episode spending (Part A Medicare payments)	Joynt Maddox (2019)	LEJR	90d	aDiD −$702 (p=0.015)	↓	264/439
Total episode spending	Liao (2022)	Medical episodes	90d	aDiD −$763 (95% CI −1,139 to −386)	↓	379/ 1,441[14]
Total episode spending	Liao (2022)	Surgical episodes	90d	aDiD –$1,010 (95% CI –1,345 to –675)	↓	379/ 1,441[14]
Total episode spending	Martin (2018)	Lumbar fusion	90d	Percentage change (risk-bearing vs control): –0.4% vs. –2.9% (p=0.044)[15]	↑	102[16]/ 1,277
Total episode spending	McNeely (2020)	PCI	90d	aDiD $122; p=0.768	↔	42/115
Total episode spending	McNeely (2020)	CABG	90d	aDiD −$300; p=0.730	↔	46/122
Total episode spending	Murphy (2019)	THA (uncomplicated)	90d	−4.44% (95% CI −4.58 to −4.30)[17]	↓	222/NR
Total episode spending	Navathe (2020)	LEJR	90d	Adj. change: –1.6% (–$377)	CI/p value NR for	244/244

[14] The extracted effect refers only to the comparison between BPCI hospitals and non-BPCI providers, as this reflects our primary review question. Accordingly, the number of participating providers refers to BPCI hospitals and non-BPCI providers only; BPCI physician group practices analyzed in the study were not considered.
[15] Spending decreased in both groups, but less in the intervention group than in the control group; therefore, the relative effect indicates higher spending.
[16] Risk-bearing hospitals only.
[17] The study did not report formal DiD estimates; reported effects reflect simple before-after comparisons within the intervention group only.

				Unadj. change: –2.7% (–$662; p<0.001)	adjusted analysis	
Total episode spending	Navathe (2021)	LEJR	90d	aDiD −2.2% (95% CI −3.6% to −0.71%)	↓	281/562
Total episode spending	Navathe (2017)	MJRLE (without complications)	30d	BPCI period vs. baseline: significant decrease (p<0.001)	↓	5 BHS hospitals in total
Total episode spending	Navathe (2017)	MJRLE (with complications)	30d	BPCI period vs. baseline: no significant change (p=0.75)	↔	5 BHS hospitals in total
Total episode spending	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD –1.2% (95% CI −2.3 to −0.2)	↓	226/700
PAC spending (SNF)	Dummit (2016)	LEJR	90d	aDiD −$546 (95% CI −892 to −199)	↓	176/841
PAC spending (IRF)	Dummit (2016)	LEJR	90d	aDiD −$445 (95% CI −811 to −79)	↓	176/841
PAC spending (HHA)	Dummit (2016)	LEJR	90d	aDiD $18 (95% CI −127 to 163)	↔	176/841
PAC spending	Chopra (2023)	LEJR	30d	aDiD −$465.6 (95% CI −594.8 to −336.5)	↓	312/ 1,977
PAC spending	Chopra (2023)	Colectomy	30d	aDiD −$308.4 (95% CI −551.9 to −64.9)	↓	312/ 1,977
PAC spending	Chopra (2023)	CABG	30d	aDiD −$128.1 (95% CI −389.1 to 132.9)	↔	312/ 1,977
PAC spending (SNF)	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD $45 (p=0.78)	↔	492/898
PAC spending (IRF)	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD −$40 (p=0.52)	↔	492/898
PAC spending (LTCH)	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD −$103 (p=0.30)	↔	492/898
PAC spending (HHA)	Joynt Maddox (2018)	CHF, pneumonia, COPD, sepsis, or AMI	90d	Across combined conditions: aDiD −$24 (p=0.47)	↔	492/898
PAC spending (SNF + IRF + LTCH)	Joynt Maddox (2019)	LEJR	90d	aDiD −$693 (p<0.001)	↓	264/439
PAC spending (institutional)	Liao (2022)	Medical episodes	90d	aDiD –$779 (p<0.05)	↓	379/ 1,441[18]
PAC spending (institutional)	Liao (2022)	Surgical episodes	90d	aDiD –$1,062 (p<0.05)	↓	379/ 1,441[18]
PAC spending (HHA)	Liao (2022)	Medical episodes	90d	aDiD $92 (not significant)	↔	379/ 1,441[18]

[18] The extracted effect refers only to the comparison between BPCI hospitals and non-BPCI providers, as this reflects our primary review question. Accordingly, the number of participating providers refers to BPCI hospitals and non-BPCI providers only; BPCI physician group practices analyzed in the study were not considered.

PAC spending (HHA)	Liao (2022)	Surgical episodes	90d	aDiD $52 (not significant)	↔	379/ 1,441[18]
PAC spending (SNF)	McNeely (2020)	PCI	90d	aDiD –$90 (p=0.607)	↔	42/115
PAC spending (SNF)	McNeely (2020)	CABG	90d	aDiD –$482 (p=0.097)	↔	46/122
PAC spending (IRF)	McNeely (2020)	PCI	90d	aDiD $127 (p=0.192)	↔	42/115
PAC spending (IRF)	McNeely (2020)	CABG	90d	aDiD –$71 (p=0.749)	↔	46/122
PAC spending (LTCH)	McNeely (2020)	PCI	90d	aDiD $83 (p=0.301)	↔	42/115
PAC spending (LTCH)	McNeely (2020)	CABG	90d	aDiD $0 (p=0.997)	↔	46/122
PAC spending (HHA)	McNeely (2020)	PCI	90d	aDiD –$3 (p=0.950)	↔	42/115
PAC spending (HHA)	McNeely (2020)	CABG	90d	aDiD –$57 (p=0.502)	↔	46/122
PAC spending	Navathe (2021)	LEJR	90d	aDiD –7.3% (95% CI –11.2 to –3.4)	↓	281/562
PAC spending (SNF)	Navathe (2017)	MJRLE	30d	BPCI year 2 vs. baseline: $1,875 vs. $2,476 (GEE: p=0.24; trend test: p=0.02)	↔	5 BHS hospitals
PAC spending (IRF)	Navathe (2017)	MJRLE	30d	BPCI year 2 vs. baseline: $1184.99 vs. $2600.87 (p=0.01)	↓	5 BHS hospitals
PAC spending (LTCF)	Navathe (2017)	MJRLE	30d	BPCI year 2 vs. baseline: $7 vs. $135	CI/p-value NR	5 BHS hospitals
PAC spending (HHA)	Navathe (2017)	MJRLE	30d	BPCI year 2 vs. baseline: $2,234 vs. $2,046 (p= 0.05)	↑	5 BHS hospitals
PAC spending (institutional)	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD –5.2% (95% CI –8.5 to –1.7)	↓	226/700
PAC use (institutional PAC discharge)	Dummit (2016)	LEJR	90d	aDiD −5.7 pp (95% CI −9.2 to −2.2)	↓	176/841
PAC use (institutional PAC days)	Dummit (2016)	LEJR	90d	aDiD −0.4 d (95% CI −1.1 to 0.3)	↔	176/841
PAC use (Initial discharge destination)	Joynt Maddox (2019)	LEJR	90d	Home: aDiD 2.4 pp (p=0.057) SNF: aDiD –0.6 pp (p=0.590) IRF: aDiD –1.6 pp (p=0.013) LTCH: aDiD 0.0 pp (p=0.400) HHA: aDiD 0.3 pp (p=0.802)	↔ (↓) IRF	264/439
PAC use (Duration of PAC stay, users only)	Joynt Maddox (2019)	LEJR	90d	SNF: aDiD –1.9 d (p<0.001) IRF: aDiD 0.1 d (p=0.484)	↓	264/439

				LTCH: aDiD –4.1 d (p=0.208) HHA: aDiD 0.0 visits/episode (p=0.987) **Any institutional PAC: aDiD –1.5 d (p=<0.001)**		
PAC use (Dispersion of PAC)	Joynt Maddox (2019)	LEJR	90d	SNF: aDiD 0.2 pp (p=0.844) IRF: aDiD –2.8 pp (p=0.343) LTCH: aDiD 8.0 pp (p=0.377) HHA: aDiD 0.7 pp (p=0.611)	↔	264/439
PAC use (institutional PAC discharge)	Liao (2022)	Medical episodes	90d	aDiD −1.1 pp (p<0.05)	↓	379/ 1,441[19]
PAC use (institutional PAC discharge)	Liao (2022)	Surgical episodes	90d	aDiD –4.5 pp (p<0.05)	↓	379/ 1,441[19]
PAC use (HHA discharge)	Liao (2022)	Medical episodes	90d	aDiD +2.1 pp (<0.05)	↑	379/ 1,441[19]
PAC use (HHA discharge)	Liao (2022)	Surgical episodes	90d	aDiD +0.6 pp (not significant)	↔	379/ 1,441[19]
PAC use (HHA days)	Liao (2022)	Medical episodes	90d	aDiD +0.4 d (p<0.05)	↑	379/ 1,441[19]
PAC use (HHA days)	Liao (2022)	Surgical episodes	90d	aDiD +0.2 d (not significant)	↔	379/ 1,441[19]
PAC use (SNF LOS)	Liao (2022)	Medical episodes	90d	aDiD −1.0 d (p<0.05)	↓	379/ 1,441[19]
PAC use (SNF LOS)	Liao (2022)	Surgical episodes	90d	aDiD –1.2 d (p<0.05)	↓	379/ 1,441[19]
PAC use (Discharge to SNF or IRF)	Navathe (2021)	LEJR	90d	aDiD −3.7 pp (95% CI –5.3 to –2.1)	↓	281/562
PAC use (decision to discharge patients to home with home health services)	Navathe (2021)	LEJR	90d	aDiD +1.3 pp (upper 95% CI bound +3.7 pp; lower bound NR); no significant difference	↔	281/562
PAC use (SNF discharge)	Oddleifson (2024)	Heart failure	90d	aOR 1.00 (95% CI 0.82 to 1.22)	↔	23/224
PAC use (Discharge to hospice/ hospice home)	Oddleifson (2024)	Heart failure	90d	aOR 1.13 (95% CI 0.89 to 1.43)	↔	23/224
PAC use (Referral to	Oddleifson (2024)	Heart failure	90d	aOR 1.94 (95% CI 0.35 to 10.83)	↔	23/224

[19] The extracted effect refers only to the comparison between BPCI hospitals and non-BPCI providers, as this reflects our primary review question. Accordingly, the number of participating providers refers to BPCI hospitals and non-BPCI providers only; BPCI physician group practices analyzed in the study were not considered.

cardiac rehabilitation among eligible patients)						
PAC use (any institutional use)	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD 0.1 pp (95% CI –0.4 to 0.7)	↔	226/700
PAC use (SNF discharge)	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD 0.1 pp (95% CI –0.5 to 0.7)	↔	226/700
PAC use (discharge to home health service)	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD 0.02 pp (95% CI –0.5 to 0.5)	↔	226/700
PAC use (SNF days)	Rolnick (2020)	AMI, CHF, COPD, or pneumonia	90d	aDiD −6.2% (95% CI −9.8 to −2.6)	↓	226/700

Table S40. CJR compared to standard traditional Medicare payment

Outcome domain	**Study**	**Condition/ procedure**	**Follow-up**	**Effect estimate**	**Direction**	**n hospitals (i/c)**
Mortality	Barnett (2019)	LEJR	90d	aDiD −0.06 pp (95% CI −0.2 to 0.1)	↔	803/962
Mortality	Haas (2019)	LEJR	30d	aDiD 0.01 pp (95% CI −0.30 to 0.30)	↔	815/955
Mortality	Haas (2019)	LEJR	90d	aDiD −0.05 pp (95% CI −0.33 to 0.24)	↔	815/955
Mortality	Kim (2020)	LEJR	90d	White: aDDD 0.0 pp (95% CI –0.1 to 0.2) Black: aDDD –0.3 pp (95% CI –0.9 to 0.4) Hispanic: aDDD 0.2 pp (95% CI –0.4 to 0.9)	↔	NR (included MSAs: 67/103)
Readmissions	Barnett (2019)	LEJR	90d	aDiD −0.6 pp (95% CI −0.9 to −0.2)	↓	803/962
Readmissions	Finkelstein (2018)	LEJR	90d	MD 0.01 pp (95% CI –0.55 to 0.58)	↔	1,633
Readmissions	Haas (2019)	LEJR	90d	aDiD –0.01 pp (95% CI –0.08 to 0.03)	↔	815/955
Readmissions	Kim (2020)	LEJR	90d	Black: aDDD −3.1 pp (95% CI −5.9 to −0.4) White: aDDD –1.4 pp (95% CI –2.9 to 0.0) Hispanic: aDDD –1.6 pp (95% CI –3.7 to 0.6)	↔ (↓) Black only	NR (included MSAs: 67/103)
ED visits	Barnett (2019)	LEJR	90d	aDiD 0.3 pp (95% CI −0.1 to 0.7)	↔	803/962
ED visits	Finkelstein (2018)	LEJR	90d	MD 0.25 pp (95% CI –0.44 to 0.93)	↔	1,633

ED visits	Kim (2020)	LEJR	90d	White: aDDD 0.1 pp (95% CI –0.4 to 0.6) Black: aDDD –0.7 pp (95% CI –2.7 to 1.2) Hispanic: aDDD 0.1 pp (95% CI –1.6 to 1.9)	↔	NR (included MSAs: 67/103)
Complications (composite rate)	Barnett (2019)	LEJR	90d	aDiD –0.04 pp (95% CI –0.2 to 0.2)	↔	803/962
Complications	Finkelstein (2018)	LEJR	90d	MD −0.002 pp (95% CI –0.21 to 0.21)	↔	1,633
Complications	Haas (2019)	LEJR	90d	aDiD 0.26 pp (95% CI –0.04 to 0.56)	↔	815/955
Complications	Kim (2020)	LEJR	90d	White: aDDD 0.1 pp (95% CI –0.2 to 0.3) Black: aDDD 0.0 pp (95% CI –0.8 to 0.8) Hispanic: aDDD 0.7 pp (95% CI –0.3 to 1.7)	↔	NR (included MSAs: 67/103)
LOS	Barnett (2019)	LEJR	in-hospital	aDiD −0.02 d (95% CI −0.04 to −0.004)	↓	803/962
LOS	Haas (2019)	LEJR	in-hospital	aDiD 0.01 d (95% CI –0.03 to 0.04)	↔	815/955
LOS	Kiani (2024)	LEJR	in-hospital	prolonged LOS (>3 days): aDiD −0.007 (change in probability; 95% CI –0.020 to 0.006)	↔	49/95
LOS	Kim (2020)	LEJR	in-hospital	White: aDDD 0.0 d (95% CI –0.1 to 0.0) Black: aDDD 0.0 d (95% CI –0.1 to 0.2) Hispanic: aDDD 0.0 d (95% CI –0.1 to 0.1)	↔ all groups	NR (included MSAs: 67/103)
LOS	Meyers (2019)	LEJR	in-hospital	TM patients: aDiD –0.02 d (95% CI –0.03 to 0.02) MA patients: aDiD 0.01 d (95% CI –0.01 to 0.03)	↔	NR (included MSAs: 75/121)
Risk selection (change in top risk-score quartile)	Barnett (2019)	LEJR	index admission	aDiD –0.1 pp (95% CI –0.5 to 0.4)	↔	803/962
Risk selection (disability enrollment)	Barnett (2019)	LEJR	index admission	aDiD −0.6 pp (95% CI −1.0 to −0.2)	↓	803/962
Risk selection (change in Elixhauser	Finkelstein (2018)	LEJR	index admission	MD –0.01 (95% CI –0.07 to 0.05)	↔	1,633

Comorbidity Index)						
Risk selection (changes in case mix)	Kiani (2024)	LEJR	index admission	Older (aDiD +0.16% [95% CI –0.33 to 0.64]) and more comorbid (aDiD CCI=0: −0.015 [change in probability; 95% CI –0.026 to –0.003]) at CJR hospitals	↔	49/95
Total episode spending (institutional payments)	Barnett (2019)	LEJR	90d	aDiD −$812 (95% CI −981 to −644)	↓	803/962
Total episode spending (incl. professional fees)	Barnett (2019)	LEJR	90d	aDiD −$1,084 (95% CI −1,409 to −760)[20]	↓	803/962
Total episode spending	Finkelstein (2018)	LEJR	90d	MD –$453 (95% CI –909 to 3)	↔	1,633
Total episode spending	Haas (2019)	LEJR	90d	aDiD –$730 (95% CI –1,745 to 286)	↔	815/955
Total episode spending	Kim (2020)	LEJR	90d	White: aDDD −$439 (95% CI −718 to −161) Black: aDDD −$500 (95% CI −$1247 to $247) Hispanic: aDDD −$419 (95% CI −$1129 to $291)	↔ Hispanic & Black (↓) White only	NR (included MSAs: 67/103)
Total episode spending	Meyers (2019)	LEJR	90d	TM: aDiD −$767.4 (95% CI −920.5 to −614.4) MA: aDiD −$348 (95% CI −560.6 to −135.5)	↓	NR (included MSAs: 75/121)
PAC spending (SNF)	Barnett (2019)	LEJR	90d	aDiD −$527 (95% CI −611 to −443)	↓	803/962
PAC spending (IRF)	Barnett (2019)	LEJR	90d	aDiD −$227 (95% CI −274 to −180)	↓	803/962
PAC spending (LTCH)	Barnett (2019)	LEJR	90d	aDiD $9 (95% CI –13 to 30)	↔	803/962
PAC spending (HHA)	Barnett (2019)	LEJR	90d	aDiD –$4 (95% CI –22 to 14)	↔	803/962
PAC spending (institutional)	Finkelstein (2018)	LEJR	90d	MD –$307 (95% CI –587 to –27)	↓	1,633
PAC spending	Haas (2019)	LEJR	90d	aDiD −$585 (95% CI −874 to −296)	↓	815/955
PAC spending (SNF)	Haas (2019)	LEJR	90d	aDiD −$243 (95% CI −460 to −26)	↓	815/955
PAC spending (IRF)	Haas (2019)	LEJR	90d	aDiD −$337 (95% CI −497 to −176)	↓	815/955

[20] This spending estimate is based on a 20% random sample of hip- or knee-replacement episodes because data on spending for professional services were available only for 20% of the sample of Medicare beneficiaries.

PAC spending (HHA)	Haas (2019)	LEJR	90d	aDiD $35 (95% CI −47 to 118)	↔	815/955
PAC spending (LTCH)	Haas (2019)	LEJR	90d	aDiD –$3 (95% CI −44 to 39)	↔	815/955
PAC spending (institutional PAC)	Kim (2020)	LEJR	90d	White: aDDD –$450 (95% CI –792 to –107) Black: aDDD –$1,162 (95% CI –1,873 to –452) Hispanic: aDDD –$687 (95% CI –1297 to –76)	↓ all groups	NR (included MSAs: 67/103)
PAC spending (LTCH)	Kim (2020)	LEJR	90d	White: aDDD $4 (95% CI –30 to 38) Black: aDDD –$9 (95% CI –154 to 137) Hispanic: aDDD –$63 (95% CI –175 to 49)	↔ all groups	NR (included MSAs: 67/103)
PAC spending (inpatient rehabilitation)	Kim (2020)	LEJR	90d	White: aDDD –$82 (95% CI –381 to 216) Black: aDDD –$481 (95% CI –1,117 to 154) Hispanic: aDDD –$133 (95% CI –523 to 257)	↔ all groups	NR (included MSAs: 67/103)
PAC spending (SNF)	Kim (2020)	LEJR	90d	White: aDDD –$347 (95% CI –580 to –113) Black: aDDD –$583 (95% CI –1,147 to –20) Hispanic: aDDD –$541 (95% CI –1,072 to –9)	↓ all groups	NR (included MSAs: 67/103)
PAC spending (Home health)	Kim (2020)	LEJR	90d	White: aDDD –$3 (95% CI –149 to 143) Black: aDDD $83 (95% CI –148 to 314) Hispanic: aDDD $143 (95% CI –86 to 371)	↔ all groups	NR (included MSAs: 67/103)
PAC use (Discharge to PAC facility)	Barnett (2019)	LEJR	90d	aDiD −2.5 pp (95% CI −3.0 to −2.1)	↓	803/962
PAC use (Discharge to HHA)	Barnett (2019)	LEJR	90d	aDiD 2.5 pp (95% CI 2.0 to 2.9)	↑	803/962
PAC use (mean LOS in PAC facility)	Barnett (2019)	LEJR	90d	aDiD –1.7 d (95% CI –2.0 to –1.4)	↓	803/962

PAC use (Discharge to institutional PAC)	Finkelstein (2018)	LEJR	90d	MD –2.9 pp (95% CI –4.95 to –0.90)	↓	1,633
PAC use (Discharge to HHA)	Finkelstein (2018)	LEJR	90d	MD 1.3 pp (95% CI –1.73 to 4.34)	↔	1,633
PAC use (Discharge to home)	Finkelstein (2018)	LEJR	90d	MD 2.6 pp (95% CI –0.79 to 5.90)	↔	1,633
PAC use (LOS in institutional PAC)	Finkelstein (2018)	LEJR	90d	MD –0.48 d (95% CI –1.06 to 0.11)	↔	1,633
PAC use (escalation to institutional care)	Kallies (2022)	THA	90d	aOR 1.16 (95% CI 0.98 to 1.36)	↔	NR
PAC use (community living post-discharge)	Kallies (2022)	THA	90d	aOR 0.91 (95% CI 0.84 to 0.98)	↓	NR
PAC use (Discharge to institutional PAC)	Kiani (2024)	LEJR	90d	aDiD –0.006 (change in probability; 95% CI –0.009 to –0.003)	↓	49/95
PAC use (Discharge to institutional PAC)	Kim (2020)	LEJR	90d	White: aDDD −2.5 pp (95% CI −4.7 to −0.4) Black: aDDD −6.0 pp (95% CI −9.8 to −2.2) Hispanic: aDDD −4.3 pp (95% CI −7.6 to −1.0)	↓ all groups	NR (included MSAs: 67/103)
PAC use (Discharge to home health)	Kim (2020)	LEJR	90d	White: aDDD 2.4 pp (95% CI −1.9 to 6.6) Black: aDDD 6.2 pp (95% CI 0.8 to 11.5) Hispanic: aDDD 6.0 pp (95% CI 1.2 to 10.9)	↑ Black & Hispanic (↔) White only	NR (included MSAs: 67/103)
PAC use (days in institutional PAC)	Kim (2020)	LEJR	90d	White: aDDD –1.0 d (95% CI −1.5 to –0.4) Black: aDDD –1.6 d (95% CI –2.8 to –0.5) Hispanic: aDDD –1.7 d (95% CI –2.7 to –0.7)	↓ all groups	NR (included MSAs: 67/103)
PAC use (any PAC use)	Meyers (2019)	LEJR	90d	TM: aDiD −2.6 pp (95% CI −2.9 to −2.2) MA: aDiD −1.5 pp (95% CI −2.0 to −1.0)	↓	NR (included MSAs: 75/121)

PAC use (Discharge to SNF)	Meyers (2019)	LEJR	90d	TM: aDiD −0.6 pp (95% CI −1.0 to −0.3) MA: aDiD −0.8 pp (95% CI −1.3 to −0.3)	↓	NR (included MSAs: 75/121)
PAC use (Discharge to IRF)	Meyers (2019)	LEJR	90d	TM: aDiD −1.9 pp (95% CI −2.1 to −1.8) MA: aDiD –0.7 pp (95% CI –0.8 to –0.6)	↓	NR (included MSAs: 75/121)
PAC use (Home health care)	Meyers (2019)	LEJR	90d	TM: aDiD 2.1 pp (95% CI 1.6 to 2.6) MA: aDiD −0.5 pp (95% CI −1.2 to 0.2)	↑ TM (↔) MA only	NR (included MSAs: 75/121)
PAC use (days in any institutional PAC)	Meyers (2019)	LEJR	90d	TM: aDiD –0.8 d (95% CI –0.9 to –0.7) MA: aDiD −0.3 d (95% CI −0.5 to –0.2)	↓	NR (included MSAs: 75/121)
PAC use (days in SNF)	Meyers (2019)	LEJR	90d	TM: aDiD –0.6 d (95% CI –0.7 to 0.4) MA: aDiD −0.3 d (95% CI −0.4 to –0.1)	↔ TM (↓) MA only	NR (included MSAs: 75/121)
PAC use (days in IRF)	Meyers (2019)	LEJR	90d	TM: aDiD –0.2 d (95% CI –0.2 to –0.2) MA: aDiD −0.1 d (95% CI −0.1 to –0.1)	↓	NR (included MSAs: 75/121)
PAC use (days in home health care)	Meyers (2019)	LEJR	90d	TM: aDiD –0.4 d (95% CI –0.6 to –0.2) MA: aDiD −0.4 d (95% CI −0.6 to –0.1)	↓	NR (included MSAs: 75/121)
PAC use (home health PT use)	Stern (2024)	THA	90d	Greater increase in home health PT use (+8.0 pp; 95% CI 3.5 to 12.6) than outpatient PT use (+2.5 pp; 95% CI –3.4 to 8.5)	↑ Home health PT use ↔ outpatient PT use	NR (included MSAs: 34/42)
PAC use (home health PT use)	Stern (2024)	TKA	90d	Greater increase in home health PT use (+2.7 pp; 95% CI –0.4 to 5.8) than outpatient PT use (+1.1 pp; 95% CI –2.3 to 4.4)	↔	NR (included MSAs: 34/42)
PAC use (PT visits)	Stern (2024)	THA	90d	Composite No. of PT visits: –1.1 (95% CI –1.2 to –0.2) with greater decrease in home health PT visits	↓	NR (included MSAs: 34/42)

				(–1.1; 95% CI –1.6 to –0.6) than in outpatient PT visits (–0.4; 95% CI –1.3 to 0.6)		
PAC use (PT visits)	Stern (2024)	TKA	90d	Composite No. of PT visits: –0.5 (95% CI –1.2 to 0.2) with greater decrease in home health PT visits (–1.1; 95% CI –1.4 to –0.7) than in outpatient PT visits (+0.8; 95% CI 0.1 to 1.4)	↔ (↓) Home health PT visits only	NR (included MSAs: 34/42)

Table S41. BPCI-A compared to standard traditional Medicare payment

Outcome domain	**Study**	**Condition/ procedure**	**Follow-up**	**Effect estimate**	**Direction**	**n hospitals (i/c)**
Mortality	Joynt Maddox (2021)	all 29 inpatient conditions	30d	ratio of ratios of relative changes, non-joining: 0.998 (95% CI 0.994 to 1.002)	↔	826/2016
Mortality	Joynt Maddox (2021)	all 29 inpatient conditions	30d	ratio of ratios of relative changes, late-joining: 0.995 (95% CI 0.987 to 1.003)	↔	826/334
Mortality	Joynt Maddox (2021)	all 29 inpatient conditions	90d	ratio of ratios of relative changes, non-joining: 1.000 (95% CI 0.997 to 1.003)	↔	826/2016
Mortality	Joynt Maddox (2021)	all 29 inpatient conditions	90d	ratio of ratios of relative changes, late-joining: 0.999 (95% CI 0.993 to 1.005)	↔	826/334
Mortality	Shashikumar (2023)	Cardiac medical bundles	90d	aDiD −0.14 pp (95% CI −0.50 to 0.23)	↔	540/540
Mortality	Shashikumar (2023)	Cardiac procedural bundles	90d	aDiD −0.36 pp (95% CI −1.25 to 0.54)	↔	540/540
Readmissions	Joynt Maddox (2021)	all 29 inpatient conditions	30d	ratio of ratios of relative changes, non-joining: 1.001 (95% CI 0.999 to 1.003)	↔	826/2016
Readmissions	Joynt Maddox (2021)	all 29 inpatient conditions	30d	ratio of ratios of relative changes, late-joining: 1.003 (95% CI 0.999 to 1.008)	↔	826/334
Readmissions	Joynt Maddox (2021)	all 29 inpatient conditions	90d	ratio of ratios of relative changes, non-joining:	↔	826/2016

				1.001 (95% CI 1.000 to 1.003)		
Readmissions	Joynt Maddox (2021)	all 29 inpatient conditions	90d	ratio of ratios of relative changes, late-joining: 1.004 (95% CI 1.000 to 1.008)	↔	826/334
Readmissions	Shashikumar (2023)	Cardiac medical bundles	90d	aDiD 0.27 pp (95% CI −0.25 to 0.80)	↔	540/540
Readmissions	Shashikumar (2023)	Cardiac procedural bundles	90d	aDiD 0.31 pp (95% CI −0.98 to 1.60)	↔	540/540
Risk selection (Case mix and patient selection)	Joynt Maddox (2021)	all 29 inpatient conditions	index admission	BPCI-A vs non-joining hospitals: no differential changes in age >80 years (−0.03 pp/quarter), female sex (0.01 pp/quarter), Black race (0.01 pp/quarter), comorbidity burden (−0.002 CCWs/quarter), DRG complexity (0.01 pp/quarter), or outlier payments (0.01 pp/quarter)	↔	826/2016
Risk selection (frailty, multimorbidity or dual enrollment)	Shashikumar (2023)	Cardiac medical bundles	index admission	No evidence of risk selection based on frailty (aDiD −0.4 pp; 95% CI −0.9 to 0.1), multimorbidity (aDiD 0.0 pp; 95% CI −0.5 to 0.5), or dual enrollment (aDiD 0.4 pp; 95% CI −0.1 to 0.9)	↔	540/540
Risk selection (frailty, multimorbidity or dual enrollment)	Shashikumar (2023)	Cardiac procedural bundles	index admission	No evidence of risk selection based on frailty (aDiD 0.2 pp; 95% CI −0.6 to 1.1), multimorbidity (aDiD 0.6 pp; 95% CI −0.8 to 1.9), or dual enrollment (aDiD 0.2 pp; 95% CI −0.8 to 1.2)	↔	540/540
Total episode spending	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, non-joining minus	↓	826/2016

				BPCI-A: $52 (95% CI 34 to 70)		
Total episode spending	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, late-joining minus BPCI-A: $82 (95% CI 41 to 122)	↓	826/334
Total episode spending	Shashikumar (2023)	Cardiac medical bundles	90d	aDiD $16 (95% CI −228 to 261)	↔	540/540
Total episode spending	Shashikumar (2023)	Cardiac procedural bundles	90d	aDiD $171 (95% CI −429 to 772)	↔	540/540
PAC spending (SNF)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, non-joining minus BPCI-A: $24 (95% CI 14 to 33)	↓	826/2016
PAC spending (SNF)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, late-joining minus BPCI-A: $54 (95% CI 33 to 74)	↓	826/334
PAC spending (IRF)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, non-joining minus BPCI-A: $7 (95% CI 3 to 11)	↓	826/2016
PAC spending (IRF)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, late-joining minus BPCI-A: $3 (95% CI –6 to 13)	↔	826/334
PAC spending (LTCH)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, non-joining minus BPCI-A: –$2 (95% CI –6 to 3)	↔	826/2016
PAC spending (LTCH)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, late-joining minus BPCI-A: –$2 (95% CI –13 to 10)	↔	826/334
PAC spending (HHA)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, non-joining minus BPCI-A: –$0 (95% CI –2 to 2)	↔	826/2016
PAC spending (HHA)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend change, late-	↔	826/334

				joining minus BPCI-A: –$2 (95% CI –5 to 2)		
PAC spending (SNF)	Shashikumar (2023)	Cardiac medical bundles	90d	aDiD −$82 (95% CI −184 to 21)	↔	540/540
PAC spending (SNF)	Shashikumar (2023)	Cardiac procedural bundles	90d	aDiD −$18 (95% CI −271 to 235)	↔	540/540
PAC use (Healthy days at home; days, out of 90)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend, non-joining: −0.04 days (95% CI −0.06 to −0.01)	↓	826/2016
PAC use (Healthy days at home; days, out of 90)	Joynt Maddox (2021)	all 29 inpatient conditions	90d	difference in quarterly trend, late-joining: −0.08 days (95% CI −0.13 to −0.03)	↓	826/334
PAC use (SNF LOS)	Shashikumar (2023)	Cardiac medical bundles	90d	aDiD –0.11 d (95% CI −0.28 to 0.07)	↔	540/540
PAC use (SNF LOS)	Shashikumar (2023)	Cardiac procedural bundles	90d	aDiD −0.07 d (95% CI −0.50 to 0.36)	↔	540/540
PAC use (healthy days at home)	Shashikumar (2023)	Cardiac medical bundles	90d	aDiD 0.14 d (95% CI −0.14 to 0.43)	↔	540/540
PAC use (healthy days at home)	Shashikumar (2023)	Cardiac procedural bundles	90d	aDiD 0.27 d (95% CI –0.43 to 0.96)	↔	540/540
Equity (Black-White differences in readmissions, mortality, healthy days at home, and access to care)	Hammond (2022)	all 29 inpatient conditions	90d	Black–White gap (difference in slope change, White minus Black): no change in readmissions (0.05 pp/quarter; 95% CI −0.04 to 0.15) or mortality (0.02 pp/quarter; 95% CI −0.05 to 0.08); greater gain in healthy days at home among Black patients (−0.07 days/ quarter; 95% CI −0.12 to −0.01; differed from controls, p = .01). Access: no change in the proportion of Black patients vs controls (−0.02 pp/quarter; 95% CI −0.05 to 0.01).	↔	826/2016

Table S42. ACE model compared to standard Medicare DRG-based payment

Outcome domain	Study	Condition/ procedure	Follow-up	Effect estimate	Direction	n hospitals (i/c)
Mortality	Chen (2018)	Hip or knee replacement or revision	30d	aDiD −0.10 pp (95% CI −0.50 to 0.31)	↔	5/22
Mortality	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD −0.27 pp (95% CI −1.25 to 0.72)	↔	4/9
Readmissions	Chen (2018)	Hip or knee replacement or revision	30d	aDiD −2.22 pp (95% CI −3.25 to −1.19)	↓	5/22
Readmissions	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD 0.21 pp (95% CI –1.67 to 2.09)	↔	4/9
Complications	Chen (2018)	Hip or knee replacement or revision	30d	aDiD −1.14 pp (95% CI −2.33 to 0.04)	↔	5/22
Complications	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD 0.81 pp (95% CI −1.80 to 3.42)	↔	4/9
Total episode spending	Chen (2018)	Hip or knee replacement or revision	30d	aDiD −$358 (95% CI −894 to 178)	↔	5/22
Total episode spending	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD $514 (95% CI −1,517 to 2,545)	↔	4/9
Total episode spending	Navathe (2017)	MJRLE without complications	30d	ACE period vs. baseline: no change (p= 0.62)	↔	5 BHS hospitals
Total episode spending	Navathe (2017)	MJRLE with complications	30d	ACE period vs. baseline: no change (p= 0.47)	↔	5 BHS hospitals
PAC spending (total PAC)	Chen (2018)	Hip or knee replacement or revision	30d	aDiD –$591 (95% CI –1,161 to –22)	↓	5/22
PAC spending (total PAC)	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD –$718 (95% CI –1,431 to –6)	↓	4/9
PAC spending (SNF)	Chen (2018)	Hip or knee replacement or revision	30d	aDiD –$204 (95% CI –737 to 330)	↔	5/22
PAC spending (SNF)	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD $1,013 (95% CI –154 to 2,180)	↔	4/9
PAC spending (IRF)	Chen (2018)	Hip or knee replacement or revision	30d	aDiD –$410 (95% CI –1,656 to 837)	↔	5/22
PAC spending (IRF)	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD $5,234 (95% CI –3,868 to 14,336)	↔	4/9
PAC spending (HHA)	Chen (2018)	Hip or knee replacement or revision	30d	aDiD –$6 (95% CI –288 to 300)	↔	5/22
PAC spending (HHA)	Chen (2018)	Cardiac valve procedures or CABG	30d	aDiD $16 (95% CI –196 to 229)	↔	4/9

PAC spending (SNF)	Navathe (2017)	MJRLE	30d	ACE year 3 vs. baseline: $2,476 vs. $2,799	CIs/p-values NR	5 BHS hospitals
PAC spending (IRF)	Navathe (2017)	MJRLE	30d	ACE year 3 vs. baseline: $2,601 vs. $2,340	CIs/p-values NR	5 BHS hospitals
PAC spending (LTCF)	Navathe (2017)	MJRLE	30d	ACE year 3 vs. baseline: $135 vs. $203	CIs/p-values NR	5 BHS hospitals
PAC spending (HHA)	Navathe (2017)	MJRLE	30d	ACE year 3 vs. baseline: $2,046 vs. $2,139	CIs/p-values NR	5 BHS hospitals

Table S43. Bundled payment period (ACE + BPCI) compared to standard traditional Medicare payment

Outcome domain	**Study**	**Condition/ procedure**	**Follow-up**	**Effect estimate**	**Direction**	**n hospitals (i/c)**
Prolonged LOS	Navathe (2017)	MJRLE	in-hospital	Declined from 22.4% at baseline to 7.3% during the bundled payment period (p<0.001)	↓	5 BHS hospitals
Readmissions	Navathe (2017)	MJRLE	30d	Declined from 6.4% at baseline to 5.0% during the bundled payment period (p=0.16)	↔	5 BHS hospitals
ED visits	Navathe (2017)	MJRLE	30d	Declined from 7.4% at baseline to 6.5% during the bundled payment period (p=0.98)	↔	5 BHS hospitals

Table S44. Carrum Health Direct Payment Program compared to usual commercial payment arrangements

Outcome domain	**Study**	**Condition/ procedure**	**Follow-up**	**Effect estimate**	**Direction**	**n hospitals (i/c)**
Total episode spending	Whaley (2021)	Spinal fusion, major joint replacement, and bariatric surgery	30d	Adjusted differential change −$4,229 (95% CI −$7,315 to −$1,142); relative change: −10.72% (95% CI −17.78 to −3.05)	↓	15/NR
Equity (patient cost sharing)	Whaley (2021)	Spinal fusion, major joint replacement, and bariatric surgery	30d	Adjusted differential change −$498 per episode (95% CI −$883 to −$113); relative change −27.7% (95% CI −43.7 to −7.3)	↓	15/NR

Table S45. Patient choice and bundled payment reform compared to standard Swedish DRG-based payment

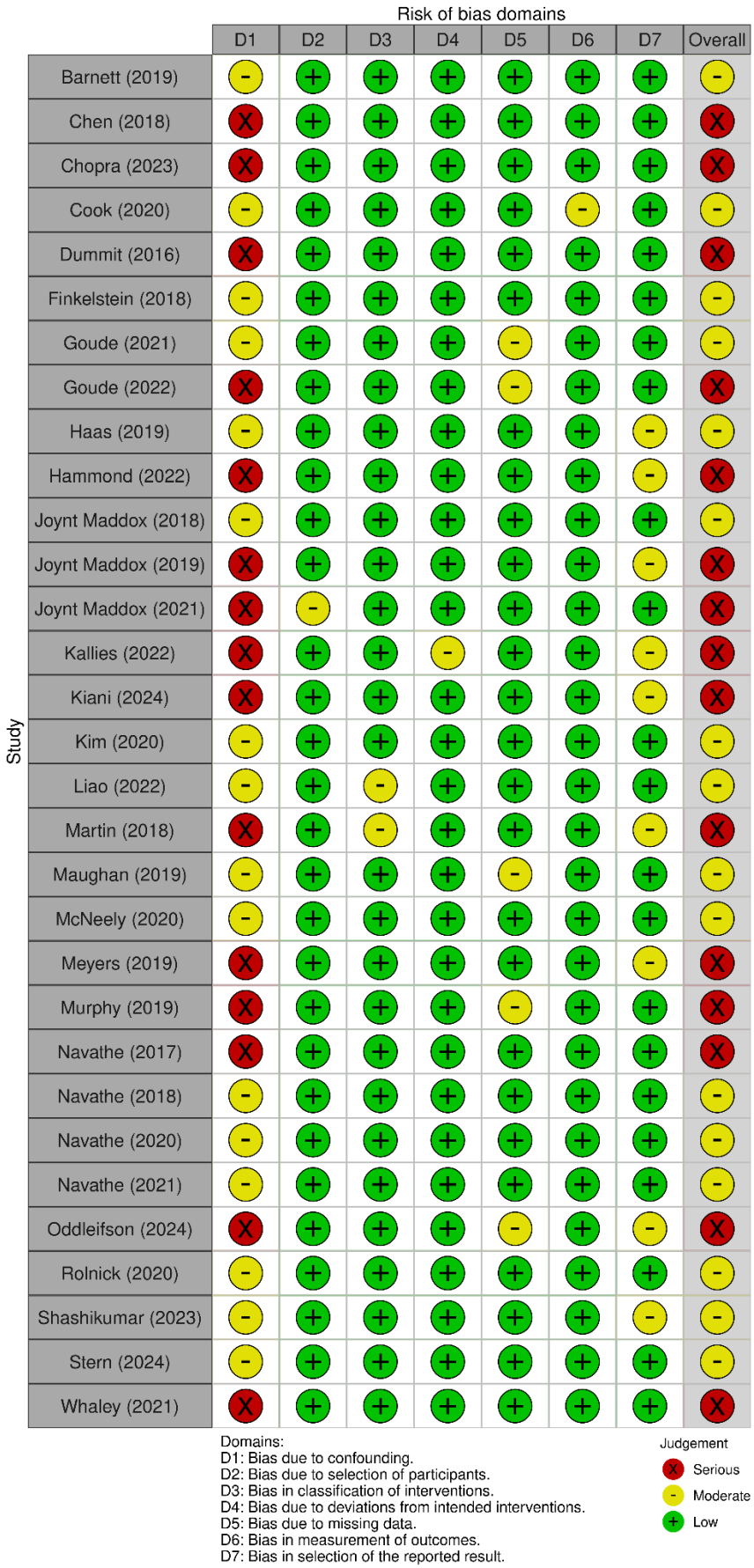
Outcome domain	Study	Condition/ procedure	Follow-up	Effect estimate	Direction	n hospitals (i/c)
Complications (AE incl. death, further surgery, and complications)	Goude (2022)	THA and TKA	90d	aDiD −1.6 pp SE=0.7 (p<0.05)	↓	NR
HRQoL (EQ-5D index; –0.594 to 1)	Goude (2021)	THA	1 year	aDiD –0.003, SE = 0.011 (not significant)	↔	NR
HRQoL (EQ-5D index; –0.594 to 1)	Goude (2021)	THA	6 years	aDiD 0.012, SE = 0.010 (not significant)	↔	NR
Pain reduction (VAS 0–100)	Goude (2021)	THA	1 year	aDiD 0.073, SE = 0.991 (not significant)	↔	NR
Pain reduction (VAS 0–100)	Goude (2021)	THA	6 years	aDiD 0.650, SE = 0.974 (not significant)	↔	NR
Patient satisfaction (VAS 0–100)	Goude (2021)	THA	1 year	aDiD 0.310, SE = 1.366 (not significant)	↔	NR
Patient satisfaction (VAS 0–100)	Goude (2021)	THA	6 years	aDiD 0.431, SE = 1.159 (not significant)	↔	NR
Patient satisfaction (satisfied: VAS ≤ 40 of 100)	Goude (2022)	THA and TKA	1 year	aDiD −1.0 pp SE=1.7 (not significant)	↔	NR
LOS (Length of surgical admission)	Goude (2022)	THA and TKA	in-hospital	aDiD 0.669 d, SE = 0.258 (p<0.01)[21]	↑	NR

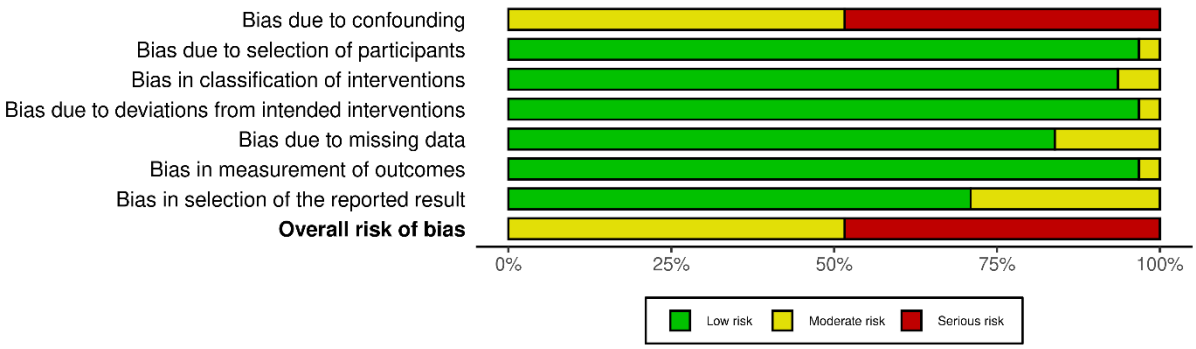
Table S46. Introduction of MS-DRGs compared to previous DRG-classification

Outcome domain	Study	Condition/ procedure	Follow-up	Effect estimate	Direction	n hospitals (i/c)
Upcoding	Cook (2020)	Overall	NR	aDiD +0.0437 DRG weight (SE=0.00242; p<0.001)	↑	≈1,000 hospitals total

[21] LOS decreased in both groups, but less in the intervention group than in the control group; the positive DiD estimate therefore indicates a relative increase in resource use.

Supplement 5. Risk of bias





Supplement 6. Assessment of the relative importance of outcomes

The relative importance of each outcome was assessed using the EPOC worksheet for preparing a Summary of Findings table with GRADE. Outcomes were rated on a 9-point scale ranging from 1, not important, to 9, critical for decision-making. Patient-important outcomes were rated by patient representatives. Quality-of-care outcomes were rated by patient representatives and stakeholders. Economic outcomes were rated by the review authors with a health economics background. Outcomes rated 1–3 were considered of limited importance, outcomes rated 4–6 important, and outcomes rated 7–9 critical for decision-making. Important and critical outcomes were included in the certainty assessment (Supplement 7) and the Summary of Findings tables (Supplement 8) if they were reported in the included studies.

Outcome		Initials of people assessing the relative importance of the outcomes			**Consensus**
		Patient representative	Stakeholder	Review team	
		Relative importance (1-9)			
Patient outcomes	a) Mortality	6			Important
	b) Morbidity	8			Critical
	c) HRQL	9			Critical
	d) Side effects/complications	9			Critical
	e) Patient satisfaction	8			Critical
Quality of Care	f) Complications	9	9		Critical
	g) Readmissions	5.6	8		Critical
	h) Length of stay (LOS)	6.4	4		Important
	i) Access to service	8.4	5		Critical
	j) Communication and time	9	5		Critical
	k) Continuity of PAC	9	6		Critical
	l) Waiting time	7.6	4		Important
	m) Adherence to clinical guidelines	7.0	6		Important

Outcome		Initials of people assessing the relative importance of the outcomes			Consensus
		Patient representative	Stakeholder	Review team	
		Relative importance (1-9)			
economic outcomes	n) Spending			9	Critical
	o) Resource use			8	Critical
	p) Shift in service volume between providers			6	Important
Healthcare provider outcomes	q) Workload				Not available
	r) Work morale				Not available
	s) Stress				Not available
	t) Sick leave				Not available
u) Equity outcomes				9	Critical

Note: Several outcomes rated important or critical by patient representatives and stakeholders, such as continuity of PAC, guideline adherence, and waiting time, were not reported in the included studies. Therefore, effect estimates and certainty assessments were not available for these outcomes.

Supplement 7. Certainty assessment of evidence

Certainty of evidence was assessed with GRADE using the EPOC worksheets. Because all included studies were non-randomized, certainty started at low for every outcome and was downgraded by one level for risk of bias, resulting in very low certainty. Further concerns about inconsistency, indirectness, or imprecision are shown in the tables but could not lower the rating further.

BPCI (Model 2) compared to standard traditional Medicare payment

Assessing the certainty of evidence across studies for an outcome

Intervention: BPCI Initiative (Model 2)
Control: Non-participating hospitals under traditional Medicare FFS payment[22]
Country: USA

Table S47. Certainty assessment of evidence for each outcome: BPCI (Model 2)

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Certainty (overall score)
Outcome: Mortality (follow-up: range from in-hospital to 90-d post-discharge)							
9	CBA	Serious	Not serious	Not serious	Serious	None	⨁◯◯◯ Very low

[22] Under traditional Medicare FFS, acute inpatient stays are generally reimbursed per case under the IPPS (MS-DRGs), while physician and PAC services are paid separately under their respective Medicare payment systems.

Outcome: Readmissions (follow-up: range from 30-d to 90-d post-discharge)							
11	CBA	Serious	Not serious	Not serious	Serious	None	⊕◯◯◯ Very low
Outcome: Emergency Department (ED) visits (follow-up: range from 30-d to 90-d post-discharge)							
6	CBA	Serious	Not serious	Not serious	Serious	None	⊕◯◯◯ Very low
Outcome: Complications (follow-up: 90-d post-discharge for one study; the other 2 NR)							
3	CBA	Serious	Not serious	Not serious	Serious[23]	None	⊕◯◯◯ Very low
Outcome: Length of stay (LOS) (follow-up: in-hospital)							
5	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Risk selection (follow-up: in-hospital)							
3	CBA	Serious	Not serious	Serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Total Episode Spending (follow-up: range from 30-d to 90-d post-discharge)							
12	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Post-acute Care (PAC[24]) spending (follow-up: range from 30-d to 90-d post-discharge)							
9	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Resource use (PAC utilization and intensity[25]) (follow-up: range from 30-d to 90-d post-discharge)							
6	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low

Participation in the BPCI model 2 compared to standard traditional Medicare payment

4 ⊕⊕⊕⊕ **High** = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different** is low.

3 ⊕⊕⊕◯ **Moderate** = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different** is moderate.

2 ⊕⊕◯◯ **Low** = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different** is high.

1 ⊕◯◯◯ **Very low** = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different** is very high.

CJR compared to standard traditional Medicare payment

Assessing the certainty of evidence across studies for an outcome

Intervention: CJR Model
Control: Hospitals in non-selected MSAs under traditional Medicare FFS payment[26]
Country: USA

Table S48. Certainty assessment of evidence for each outcome: CJR

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Certainty (overall score)
Outcome: Mortality (follow-up: range from 30d to 90d post-discharge)							
3	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯

[23] Two studies reported complication outcomes narratively without providing numerical estimates
[24] PAC includes SNF, IRF, LTCH, and HHA.
[25] Component-specific effects were not graded separately.
[26] Under traditional Medicare FFS, acute inpatient stays are generally reimbursed per case under the IPPS (MS-DRGs), while physician and PAC services are paid separately under their respective Medicare payment systems.

							Very low
Outcome: Readmissions (follow-up: range from 30d to 90d post-discharge)							
4	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Emergency Department (ED) visits (follow-up: 90d post-discharge)							
3	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Complications (follow-up: 90d post-discharge)							
4	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Length of stay (LOS) (follow-up: in-hospital)							
5	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Risk selection (follow-up: in-hospital)							
3	CBA	Serious	Not serious	Serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Total Episode Spending (follow-up: 30-d to 90-d post-discharge)							
5	CBA	Serious	Serious	Not serious	Serious	None	⊕◯◯◯ Very low
Outcome: Post-acute Care (PAC[27]) spending (follow-up: 30-d to 90-d post-discharge)							
4	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low
Outcome: Resource use (PAC utilization and intensity[28]) (follow-up: 30-d to 90-d post-discharge)							
7	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low

Participation in the CJR model compared to standard traditional Medicare payment

4 ⊕⊕⊕⊕ **High** = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different** is low.

3 ⊕⊕⊕◯ **Moderate** = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different** is moderate.

2 ⊕⊕◯◯ **Low** = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different** is high.

1 ⊕◯◯◯ **Very low** = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different** is very high.

BPCI-A compared to standard traditional Medicare payment

Assessing the certainty of evidence across studies for an outcome

Intervention: BPCI-A
Control: Non-participating hospitals and PGPs under traditional Medicare FFS payment[29]
Country: USA

Table S49. Certainty assessment of evidence for each outcome: BPCI-A

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Certainty (overall score)
Outcome: Mortality (follow-up: 90-d post-discharge)							
2	CBA	Serious	Not serious	Not serious	Not serious	None	⊕◯◯◯ Very low

[27] PAC includes SNF, IRF, LTCH, and HHA.
[28] Component-specific effects were not graded separately.
[29] Under traditional Medicare FFS, acute inpatient stays are generally reimbursed per case under the IPPS (MS-DRGs), while physician and PAC services are paid separately under their respective Medicare payment systems.

Outcome: Readmissions (follow-up: 90-d post-discharge)							
2	CBA	Serious	Not serious	Not serious	Serious	None	⨁◯◯◯ Very low
Outcome: Risk selection (follow-up: in-hospital)							
2	CBA	Serious	Not serious	Serious	Serious	None	⨁◯◯◯ Very low
Outcome: Total Episode Spending (follow-up: 90-d post-discharge)							
2	CBA	Serious	Not serious	Not serious	Serious	None	⨁◯◯◯ Very low
Outcome: Post-acute Care (PAC[30]) spending (follow-up: 90-d post-discharge)							
2	CBA	Serious	Not serious	Not serious	Serious	None	⨁◯◯◯ Very low
Outcome: Resource use (PAC utilization[31] and Healthy days at home) (follow-up: 90-d post-discharge)							
2	CBA	Serious	Not serious	Serious	Serious	None	⨁◯◯◯ Very low
Outcome: Equity (follow-up: in-hospital)							
1	CBA	Serious	Not serious	Serious	Not serious	None	⨁◯◯◯ Very low

Participation in the BPCI-A model compared to standard traditional Medicare payment

4 ⨁⨁⨁⨁ **High** = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different** is low.

3 ⨁⨁⨁◯ **Moderate** = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different** is moderate.

2 ⨁⨁◯◯ **Low** = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different** is high.

1 ⨁◯◯◯ **Very low** = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different** is very high.

ACE Demonstration compared to standard Medicare DRG-based payment

Assessing the certainty of evidence across studies for an outcome

Intervention: ACE Demonstration
Control: Non-participating hospitals under traditional Medicare, with inpatient hospital stays reimbursed under IPPS and physician services paid separately under the Medicare Physician Fee Schedule
Country: USA

Table S50. Certainty assessment of evidence for each outcome: ACE Demonstration

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Certainty (overall score)
Outcome: Mortality (follow-up: 30-d post-discharge)							
1	CBA	Serious	Not applicable	Not serious	Serious	None	⨁◯◯◯ Very low
Outcome: Readmissions (follow-up: 30-d post-discharge)							
1	CBA	Serious	Not applicable	Not serious	Serious	None	⨁◯◯◯ Very low
Outcome: Complications (follow-up: 30-day post-discharge)							
1	CBA	Serious	Not applicable	Not serious	Serious	None	⨁◯◯◯

[30] PAC includes SNF, IRF, LTCH, and HHA.
[31] Component-specific effects were not graded separately.

							Very low
Outcome: Total Episode Spending (follow-up: 30-d post-discharge)							
2	CBA, ITS	Serious	Serious	Not serious	Not serious	None	⨁◯◯◯ Very low
Outcome: Post-acute Care (PAC[32]) spending (follow-up: 30-d post-discharge)							
1	CBA	Serious	Not applicable	Not serious	Not serious	None	⨁◯◯◯ Very low

Participation in the ACE Demonstration compared to standard Medicare DRG-based payment

4 ⨁⨁⨁⨁ **High** = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different** is low.

3 ⨁⨁⨁◯ **Moderate** = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different** is moderate.

2 ⨁⨁◯◯ **Low** = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different** is high.

1 ⨁◯◯◯ **Very low** = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different** is very high.

Carrum Health Direct Payment program compared to usual payment arrangements

Assessing the certainty of evidence across studies for an outcome

Intervention: Carrum Health Direct Payment (an Employer-Provider Direct Payment Program)
Control: Non-participating providers under the employers' standard commercial health plan, with facility and physician services reimbursed separately through existing negotiated commercial insurance arrangements, without a direct bundled payment for the episode
Country: USA

Table S51. Certainty assessment of evidence for each outcome: Carrum Health direct payment program

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Certainty (overall score)
Outcome: Total episode spending (follow-up: 30-d post-discharge)							
1	CBA	Serious	Not applicable	Not serious	Not serious	None	⨁◯◯◯ Very low
Outcome: Equity (patient cost sharing) (follow-up: 30-d post-discharge)							
1	CBA	Serious	Not applicable	Not serious	Not serious	None	⨁◯◯◯ Very low

The Carrum Health employer-provider direct payment program compared to usual payment arrangements

4 ⨁⨁⨁⨁ **High** = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different** is low.

3 ⨁⨁⨁◯ **Moderate** = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different** is moderate.

2 ⨁⨁◯◯ **Low** = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different** is high.

1 ⨁◯◯◯ **Very low** = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different** is very high.

[32] PAC includes SNF, IRF, LTCH, and HHA.

Patient choice reform compared to usual DRG-based hospital payment

Assessing the certainty of evidence across studies for an outcome

Intervention: Patient choice reform (consists of three central elements: free choice of provider for patients, free entry of new providers and bundled payment where the providers receive a package price for the entire cycle of care)
Control: Providers in other Swedish regions not exposed to the patient choice reform, with orthopaedic care reimbursed through DRG-based activity-based funding
Country: Sweden

Table S52. Certainty assessment of evidence for each outcome: Patient choice reform

No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other	Certainty (overall score)
Outcome: Adverse events (follow-up: 90-d post-discharge)							
1	CBA	Serious	Not applicable	Not serious	Not serious	None	⨁◯◯◯ Very low
Outcome: Quality of life (EQ-5D) (follow-up: ranged from 1 year to 6 years)							
1	CBA	Serious	Not applicable	Not serious	Not serious	None	⨁◯◯◯ Very low
Outcome: Patient satisfaction (VAS) (follow-up: ranged from 1 year to 6 years)							
2	CBA	Serious	Not serious	Not serious	Not serious	None	⨁◯◯◯ Very low
Outcome: LOS (follow-up: in-hospital)							
1	CBA	Serious	Not applicable	Not serious	Not serious	None	⨁◯◯◯ Very low

Participation in the patient choice reform compared to usual DRG-based hospital payment

4 ⨁⨁⨁⨁ **High** = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different** is low.
3 ⨁⨁⨁◯ **Moderate** = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different** is moderate.
2 ⨁⨁◯◯ **Low** = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different** is high.
1 ⨁◯◯◯ **Very low** = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different** is very high.

Supplement 8. Summary of Findings (SoF) using plain language summaries

Statements follow the GRADE guidance for informative statements (Santesso et al., J Clin Epidemiol 2020;119:126–135). Because certainty was very low for all outcomes, each statement indicates the direction of the observed effects and states that the evidence is very uncertain. Certainty started at low because all studies were non-randomized and was downgraded for risk of bias (Supplement 7).

Table S53. Summary of Findings with plain language summary statements for the BPCI (Model 2) initiative

BPCI (Model 2) compared to standard traditional Medicare payment			
Patients or population: Hospitals **Settings:** USA **Intervention:** BPCI Initiative (Model 2) **Comparison:** Standard traditional Medicare payment			
Outcomes	**Impact**	**Number of hospitals (Studies)**	**Certainty of the evidence (GRADE)**
Mortality	BPCI Model 2 may result in little to no difference in all-cause mortality, but the evidence is very uncertain. Seven of nine studies found no statistically significant difference. One study reported small reductions for medical and surgical episodes, and one found lower in-hospital mortality for heart failure, with no difference at 30 or 90 days.	15,832[33, 34] (9 observational studies)	⨁⊖⊖⊖ Very low [a, b]
Readmissions	BPCI Model 2 may result in little to no difference in readmissions, but the evidence is very uncertain. Most of the 11 studies found no statistically significant change. One study found a small reduction for medical episodes, and another found an increase after lumbar fusion.	17,980[33, 34] (11 observational studies)	⨁⊖⊖⊖ Very low [a, b]
ED visits	BPCI Model 2 may result in little to no difference in ED visits, but the evidence is very uncertain. Five of six studies found no statistically significant change. One study found more ED visits after PCI and fewer after CABG.	13,487[34] (6 observational studies)	⨁⊖⊖⊖ Very low [a, b]
Complications	BPCI Model 2 may result in little to no difference in complications, but the evidence is very uncertain. Two of three studies found no statistically significant difference in LEJR complication rates. One study found a larger increase in repeat surgery after lumbar fusion in BPCI hospitals, while other complications did not differ.	2,710 (3 observational studies)	⨁⊖⊖⊖ Very low [a, b]

[33] One study contributed effect estimates but did not report the number of hospitals included in the control group.
[34] One study reported the numbers of participating and matched control units at the level of clinical departments rather than distinct hospitals. Because information on whether departments across clinical episodes belonged to the same hospital was not available, these numbers represent department-level counts and may overestimate the number of unique hospitals.

LOS	BPCI Model 2 may result in little to no difference in hospital LOS, but the evidence is very uncertain. Four of five studies found no statistically significant change, and one found a small reduction.	3,821 (5 observational studies)	⨁⊖⊖⊖ Very low [a]
Risk selection (Change in Case Mix)	BPCI Model 2 may result in little to no difference in patient case mix, but the evidence is very uncertain. None of the three studies found a meaningful overall shift in case mix.	3,376 (3 observational studies)	⨁⊖⊖⊖ Very low [a, c]
Total episode spending	BPCI Model 2 may reduce total episode spending, but the evidence is very uncertain. Most of the 12 studies found lower spending, mainly for orthopedic episodes (LEJR) and selected medical conditions. Studies of cardiac procedures (PCI, CABG) found no statistically significant difference, and one study found a relative increase in spending for lumbar fusion.	11,407 (12 observational studies)	⨁⊖⊖⊖ Very low [a]
PAC[35] spending	BPCI Model 2 may reduce PAC spending, but the evidence is very uncertain. Across nine studies, reductions occurred mainly in institutional PAC (SNF, IRF), particularly for LEJR. Home health spending and PAC spending after PCI or CABG mostly showed no statistically significant change.	9,318 (9 observational studies)	⨁⊖⊖⊖ Very low [a]
HRQL	No studies on this outcome identified.		
Patient satisfaction	No studies on this outcome identified.		
Resource use (PAC use[36])	BPCI Model 2 may reduce institutional PAC use, but the evidence is very uncertain. Most of the six studies found fewer discharges to institutional PAC or shorter SNF stays, while effects in other settings were mixed. One study found more home health use after medical episodes.	5,556 (6 observational studies)	⨁⊖⊖⊖ Very low [a]
Equity	No studies on this outcome identified.		

* GRADE Working Group grades of evidence

High = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different† is low.
Moderate = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different† is moderate.
Low = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different† is high.
Very low = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different† is very high.

† Substantially different = a large enough difference that it might affect a decision

[a] Downgraded 1 level for risk of bias
[b] Downgraded 1 level for imprecision
[c] Downgraded 1 level for indirectness

[35] PAC includes SNF, IRF, LTCH, and HHA.
[36] Component-specific effects were not graded separately.

Table S54. Summary of Findings with plain language summary statements for the CJR model

CJR compared to standard traditional Medicare payment			
Patients or population: Hospitals **Settings:** USA **Intervention:** CJR Model **Comparison:** Standard traditional Medicare payment			
Outcomes	**Impact**	**Number of participants (Studies)**	**Certainty of the evidence (GRADE)**
Mortality	CJR may result in little to no difference in all-cause mortality, but the evidence is very uncertain. None of the three studies found a statistically significant difference.	3,535[37] (3 observational studies)	⨁⨀⨀⨀ Very low [a]
Readmissions	CJR may result in little to no difference in readmissions, but the evidence is very uncertain. Three of four studies found no statistically significant overall change, and one found a small reduction. One of the three found a reduction among Black patients only.	5,168[37] (4 observational studies)	⨁⨀⨀⨀ Very low [a]
ED visits	CJR may result in little to no difference in ED visits, but the evidence is very uncertain. None of the three studies found a statistically significant difference.	3,398[37] (3 observational studies)	⨁⨀⨀⨀ Very low [a]
Complications	CJR may result in little to no difference in complications, but the evidence is very uncertain. None of the four studies found a statistically significant difference.	5,168[37] (4 observational studies)	⨁⨀⨀⨀ Very low [a]
LOS	CJR may result in little to no difference in hospital LOS, but the evidence is very uncertain. Four of five studies found no statistically significant change, and one found a small reduction.	3,679[38] (5 observational studies)	⨁⨀⨀⨀ Very low [a]
Risk selection (Change in Case Mix)	CJR may result in little to no difference in patient case mix, but the evidence is very uncertain. Across three studies, most risk indicators did not change. One study found a small decline in beneficiaries enrolled through disability, and another found slightly more comorbid patients at CJR hospitals, so no consistent pattern of selection emerged.	3,542 (3 observational studies)	⨁⨀⨀⨀ Very low [a, b]
Total episode spending	CJR may reduce total episode spending, but the evidence is very uncertain. Two of five studies found statistically significant reductions, and one found a reduction among White patients only. The other two found no statistically significant change.	5,168[38] (5 observational studies)	⨁⨀⨀⨀ Very low [a, c, d]

[37] One study reported effect estimates but did not provide the number of included hospitals; only the number of participating MSAs was reported.
[38] Two studies reported effect estimates but did not provide the number of included hospitals; only the number of participating MSAs was reported.

PAC[39] spending	CJR may reduce PAC spending, but the evidence is very uncertain. All four studies found lower institutional PAC spending, mainly for SNF and IRF care. Home health and LTCH spending did not change significantly.	5,168[40] (4 observational studies)	⨁⊖⊖⊖ Very low [a]
HRQL	No studies on this outcome identified.		
Patient satisfaction	No studies on this outcome identified.		
Resource use (PAC use[41])	CJR may reduce institutional PAC use, but the evidence is very uncertain. Five of seven studies found fewer discharges to institutional PAC, while discharges to home health were unchanged or increased. One study found that, among patients discharged to institutional care after hip arthroplasty, fewer returned to the community.	3,542[42] (7 observational studies)	⨁⊖⊖⊖ Very low [a]
Equity	No studies on this outcome identified.		

* GRADE Working Group grades of evidence

High = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different[†] is low.
Moderate = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[†] is moderate.
Low = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[†] is high.
Very low = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[†] is very high.

[†] Substantially different = a large enough difference that it might affect a decision

[a] Downgraded 1 level for risk of bias
[b] Downgraded 1 level for indirectness
[c] Downgraded 1 level for inconsistency
[d] Downgraded 1 level for imprecision

[39] PAC includes SNF, IRF, LTCH, and HHA.
[40] One study reported effect estimates but did not provide the number of included hospitals; only the number of participating MSAs was reported.
[41] Component-specific effects were not graded separately.
[42] Three studies reported effect estimates but did not provide the number of included hospitals; only the number of participating MSAs was reported in two studies.

Table S55. Summary of Findings with plain language summary statements for the BPCI-A model

BPCI-A compared to standard traditional Medicare payment			
Patients or population: Hospitals **Setting:** USA **Intervention:** BPCI-A **Comparison:** Standard traditional Medicare payment			
Outcomes	**Impact**	**Number of participants (Studies)**	**Certainty of the evidence (GRADE)**
Mortality	BPCI-A may result in little to no difference in all-cause mortality, but the evidence is very uncertain. Neither of the two studies found a statistically significant difference.	4,256 (2 observational studies)	⨁⊖⊖⊖ Very low [a]
Readmissions	BPCI-A may result in little to no difference in readmissions, but the evidence is very uncertain. Neither of the two studies found a statistically significant difference.	4,256 (2 observational studies)	⨁⊖⊖⊖ Very low [a, b]
ED visits	No studies on this outcome identified.		
Complications	No studies on this outcome identified.		
LOS	No studies on this outcome identified.		
Risk selection (Change in Case Mix or patient complexity)	BPCI-A may result in little to no difference in patient case mix, but the evidence is very uncertain. Neither of the two studies found meaningful changes in case mix or patient selection.	4,256 (2 observational studies)	⨁⊖⊖⊖ Very low [a, b, c]
Total episode spending	BPCI-A may slightly reduce total episode spending, but the evidence is very uncertain. One study found a small reduction across all inpatient conditions. The other found no statistically significant difference for cardiac medical or procedural bundles.	4,256 (2 observational studies)	⨁⊖⊖⊖ Very low [a, b]
PAC[43] spending	BPCI-A may slightly reduce PAC spending, but the evidence is very uncertain. One study found lower SNF and IRF spending across all inpatient conditions, with no change in LTCH or home health spending. The other found no change in SNF spending for cardiac bundles.	4,256 (2 observational studies)	⨁⊖⊖⊖ Very low [a, b]
HRQL	No studies on this outcome identified.		
Patient satisfaction	No studies on this outcome identified.		

[43] PAC includes SNF, IRF, LTCH, and HHC (Home Health Care).

Resource use (PAC use[44] and healthy days at home)	BPCI-A may result in little to no difference in PAC use, but the evidence is very uncertain. One study found a very small relative decline in healthy days at home across all inpatient conditions. The other found no change in SNF LOS or healthy days at home for cardiac bundles.	4,256 (2 observational studies)	⨁⊖⊖⊖ Very low [a, b, c]
Equity	BPCI-A may result in little to no difference in racial inequities in mortality, readmissions, or access to care, but the evidence is very uncertain. The single study found a slightly larger increase in healthy days at home among Black than among White patients.	2,842 (1 observational study)	⨁⊖⊖⊖ Very low [a, c]

* GRADE Working Group grades of evidence

High = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different[†] is low.
Moderate = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[†] is moderate.
Low = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[†] is high.
Very low = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[†] is very high.

[†] Substantially different = a large enough difference that it might affect a decision

[a] Downgraded 1 level for risk of bias
[b] Downgraded 1 level for imprecision
[c] Downgraded 1 level for indirectness

Table S56. Summary of Findings with plain language summary statements for the ACE Demonstration

ACE Demonstration compared to standard Medicare DRG-based payment			
Patients or population: Hospitals **Setting:** USA **Intervention:** ACE Demonstration **Comparison:** Standard Medicare DRG-based payment			
Outcomes	**Impact**	**Number of participants (Studies)**	**Certainty of the evidence (GRADE)**
Mortality	The ACE Demonstration may result in little to no difference in 30-day mortality, but the evidence is very uncertain. The single study found no statistically significant difference for orthopedic or cardiac procedures.	40[45] (1 observational study)	⨁⊖⊖⊖ Very low [a, b]
Readmissions	The ACE Demonstration may reduce 30-day readmissions after orthopedic surgery, but the evidence is very uncertain. The single study found no change after cardiac surgery.	40[45] (1 observational study)	⨁⊖⊖⊖ Very low [a, b]

[44] Component-specific effects were not graded separately.
[45] One study reported ACE participation and matched control hospitals separately for orthopedic (5 ACE; 22 control) and cardiac (4 ACE; 9 control) service units. Because overlap of hospitals across service lines was not reported, totals reflect the sum across service lines and may overestimate the number of unique hospitals.

ED visits	No studies on this outcome identified.		
Complications	The ACE Demonstration may result in little to no difference in 30-day complications, but the evidence is very uncertain. The single study found no statistically significant difference after orthopedic or cardiac procedures.	40[45] (1 observational study)	⨁⊖⊖⊖ Very low [a, b]
LOS	No studies on this outcome identified.		
Risk selection	No studies on this outcome identified.		
Total episode spending	The ACE Demonstration may result in little to no difference in total episode spending, but the evidence is very uncertain. Neither of the two studies found a statistically significant change for orthopedic or cardiac episodes.	45[45] (2 observational studies)	⨁⊖⊖⊖ Very low [a, c]
PAC[46] spending	The ACE Demonstration may reduce total PAC spending, but the evidence is very uncertain. The single study found lower total PAC spending for orthopedic and cardiac procedures, while individual PAC settings showed no statistically significant change.	40[45] (1 observational study)	⨁⊖⊖⊖ Very low [a]
HRQL	No studies on this outcome identified.		
Patient satisfaction	No studies on this outcome identified.		
Resource use (PAC use)	No studies on this outcome identified.		
Equity	No studies on this outcome identified.		

* GRADE Working Group grades of evidence

High = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different† is low.
Moderate = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different† is moderate.
Low = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different† is high.
Very low = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different† is very high.

† Substantially different = a large enough difference that it might affect a decision

[a] Downgraded 1 level for risk of bias
[b] Downgraded 1 level for imprecision
[c] Downgraded 1 level for inconsistency

[46] PAC includes SNF, IRF, and other PAC

Table S57. Summary of Findings with plain language summary statements for the Carrum Health direct payment program

The Carrum Health employer-provider direct payment program compared to usual payment arrangements, including DRG-based inpatient payment			
Patients or population: Hospitals **Setting:** USA **Intervention:** Employer-Provider Direct Payment Program **Comparison:** Usual payment arrangements			
Outcomes	**Impact**	**Number of participants (Studies)**	**Certainty of the evidence (GRADE)**
Mortality	No studies on this outcome identified.		
Readmissions	No studies on this outcome identified.		
ED visits	No studies on this outcome identified.		
Complications	No studies on this outcome identified.		
LOS	No studies on this outcome identified.		
Risk selection	No studies on this outcome identified.		
Total episode spending	The Carrum Health direct payment program may reduce total episode spending, but the evidence is very uncertain. The single study found about 11% lower spending per episode.	15[47] (1 observational study)	⨁⊖⊖⊖ Very low [a]
PAC spending	No studies on this outcome identified.		
HRQL	No studies on this outcome identified.		
Patient satisfaction	No studies on this outcome identified.		
Resource use	No studies on this outcome identified.		

[47] The intervention group included 15 participating hospitals (Centers of Excellence). The control group comprised non–direct-payment episodes across multiple non-participating hospitals; the number of control hospitals was not reported.

Equity (patient cost sharing)	The Carrum Health direct payment program may reduce patient cost sharing, but the evidence is very uncertain. The single study found lower out-of-pocket costs per episode.	15[47] (1 observational study)	⨁⊖⊖⊖ Very low [a]
* GRADE Working Group grades of evidence **High** = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different[†] is low. **Moderate** = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[†] is moderate. **Low** = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[†] is high. **Very low** = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[†] is very high. [†] Substantially different = a large enough difference that it might affect a decision			

[a] Downgraded 1 level for risk of bias

Table S58. Summary of Findings with plain language summary statements for the patient choice reform

Patient choice reform compared to usual payment arrangements, including DRG-based inpatient payment			
Patients or population: Hospitals **Setting:** Sweden **Intervention:** Patient choice reform (consists of three central elements: free choice of provider for patients, free entry of new providers and bundled payment where the providers receive a package price for the entire cycle of care) **Comparison:** Usual payment arrangements			
Outcomes	**Impact**	**Number of participants (Studies)**	**Certainty of the evidence (GRADE)**
Mortality	No studies on this outcome identified.		
Readmissions	No studies on this outcome identified.		
ED visits	No studies on this outcome identified.		
Complications (adverse events[48])	Sweden's patient choice and bundled payment reform may reduce adverse events within 90 days after hip and knee replacement surgery, but the evidence is very uncertain. This finding comes from a single study.	NR[49] (1 observational study)	⨁⊖⊖⊖ Very low [a]

[48] Adverse events (AE) incl. death, further surgery, and complications

[49] Data were drawn from nationwide registers covering hip replacement surgery in Sweden; the number of contributing hospitals was not explicitly reported.

LOS	Sweden's patient choice and bundled payment reform may increase the length of surgical admission relative to control regions, but the evidence is very uncertain. In the single study of hip and knee replacement surgery, LOS declined in both groups, but less under the reform.	NR[49] (1 observational study)	⨁⊖⊖⊖ Very low [a]
Risk selection	No studies on this outcome identified.		
Total episode spending	No studies on this outcome identified.		
PAC spending	No studies on this outcome identified.		
HRQL (EQ-5D)	Sweden's patient choice and bundled payment reform may result in little to no difference in quality of life after hip replacement surgery, measured by EQ-5D index gain, but the evidence is very uncertain.	NR[49] (1 observational study)	⨁⊖⊖⊖ Very low [a]
Patient satisfaction	Sweden's patient choice and bundled payment reform may result in little to no difference in patient satisfaction after hip or knee replacement surgery, but the evidence is very uncertain. Neither of the two studies found a statistically significant difference.	NR[49] (2 observational studies)	⨁⊖⊖⊖ Very low [a]
Resource use	No studies on this outcome identified.		
Equity	No studies on this outcome identified.		

* GRADE Working Group grades of evidence

High = This research provides a very good indication of the likely effect. The likelihood that the effect will be substantially different[†] is low.
Moderate = This research provides a good indication of the likely effect. The likelihood that the effect will be substantially different[†] is moderate.
Low = This research provides some indication of the likely effect. However, the likelihood that it will be substantially different[†] is high.
Very low = This research does not provide a reliable indication of the likely effect. The likelihood that the effect will be substantially different[†] is very high.

[†] Substantially different = a large enough difference that it might affect a decision

[a] Downgraded 1 level for risk of bias