Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Modifications of diagnosis-related group payment for hospital care: a systematic review

Luisa Sophie Schmidt

luisasophie.schmidt@med.uni-goettingen.de

University Medical Center Göttingen https://orcid.org/0009-0008-6001-9199

Regina Kampo

University Medical Center Göttingen

Christian Röver

University Medical Center Göttingen

Tim Mathes

University Medical Center Göttingen

Systematic Review

Keywords: Diagnosis-related groups, Inpatient payment systems, Payment reform, DRG refinement, Bundled payments, Quality of care, Health care costs

Posted Date: October 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11477233/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Payment systems based on diagnosis-related groups (DRGs) are increasingly the main system for reimbursing inpatient care. Since their introduction they have been continuously refined to mitigate unintended incentives and fit country-specific contexts, but the impact of these modifications remains unclear. This review assesses the effects of DRG design modifications on patient outcomes, quality of care, utilization, costs, provider outcomes, and equity.

Methods

We searched PubMed, Web of Science, EconLit, and WHOLIS up to August 2024 for studies evaluating modifications of DRG-based payment systems or their combination with other payment methods compared with the previous DRG system or classification. Study characteristics were extracted by one reviewer and intervention effects by a second, verified by the first. Risk of bias was assessed with ROBINS-I and certainty of evidence with GRADE. Because of heterogeneity in intervention design and clinical scope, findings were synthesized narratively and visualized in forest plots.

Results

We included 31 quasi-experimental studies, predominantly from the United States, most of which evaluated bundled payment reforms. Bundled payments were associated with reductions in episode spending and resource use, most consistently through lower use of institutional post-acute care, largely without statistically significant adverse effects on mortality, readmissions, emergency department visits, or complications. Evidence on equity was limited. Where assessed, no statistically significant changes in case mix or patient selection were observed. Patient-reported outcomes, available only from Swedish studies, showed no statistically significant changes. Beyond bundled payments, one study evaluating the introduction of a severity-adjusted DRG classification (Medicare Severity DRGs) found a statistically significant increase in upcoding.

Conclusion

Evidence on modifications to DRG-based payment beyond bundled payments and outside the United States Medicare context remains largely absent. Bundled payment models were associated with lower episode spending, driven primarily by reduced use of institutional post-acute care, largely without adverse effects on patient-important outcomes. These findings suggest bundled payments may improve efficiency without compromising quality for selected episodes, but effects may vary by context and

require careful implementation and risk adjustment. Certainty of evidence was very low, and the apparent consistency of findings partly reflects overlapping data.

Systematic review registration:

PROSPERO CRD42024565484.

Background

Diagnosis-Related Groups (DRGs) are a central mechanism for reimbursing hospital services worldwide [1]. Originally developed in the United States in the 1970s to define and measure hospital activity [2, 3], today they are mainly used for billing purposes. Hospital cases are classified by diagnosis into economically homogeneous groups that receive a fixed payment based on average resource use. Because payment is fixed per case irrespective of actual expenditure, financial risk shifts to hospitals, which retain surpluses when costs fall below DRG revenue and incur losses otherwise, creating incentives for cost containment and efficiency [3–5].

While DRGs increase transparency and may improve efficiency in inpatient care, concerns about unintended effects, such as case volume expansion, premature discharge, cost shifting in other sectors of care, and insufficient attention to quality (particularly the coordination of care beyond hospital settings), have led to ongoing country-specific reforms [3, 6–9]. Earlier refinements adjusted the grouping logic (e.g., by recalibration, relative weights, risk adjustment, outlier payments) to better reflect severity and resource consumption [10]. Other modifications combined DRGs with provider budgets or activity caps to constrain volume growth [8], or reimbursed selected high-cost services outside the DRG framework to prevent cost shifting and cherry-picking [11].

More recently, bundled (episode-based) payment models, which provide a single payment for an entire episode of care spanning multiple providers and settings, have been introduced [12]. They extend financial risk from the hospital stay to the full episode of care, aiming to prevent discontinuities across provider boundaries and cost shifting between settings, and they incorporate quality-related mechanisms [13]. Their underlying principle remains a case-based lump sum, with the unit of payment expanded from the inpatient stay to a defined episode of care. Episodes are typically delineated using DRG categories, and target prices are derived from historical DRG-based spending. In the United States, initiatives such as Bundled Payments for Care Improvement (BPCI) and Comprehensive Care for Joint Replacement (CJR) exemplify these efforts [14, 15].

Despite the ongoing development of DRG-based payment methods, evidence on modifications to these systems remains fragmented, focusing mainly on the initial introduction of DRGs or on isolated adjustments. To our knowledge, no systematic review has yet assessed the effects of these modifications.

Objective

This review aimed to assess the effects of modifications to DRG-based payment systems (e.g., changes to DRG design components or combinations with other payment methods) on patient outcomes, quality of care, utilization, costs, provider outcomes, and equity. Because these are complex health-system interventions, we aimed not only to determine whether they work in general but, where evidence allowed, to identify contextual conditions under which they may be more or less effective.

Methods

Our review was reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline [16]. The review was prospectively registered in PROSPERO (registration number CRD42024565484), and the protocol was published on the Open Science Framework (OSF) [17].

Eligibility criteria

Eligibility criteria followed the Population, Intervention, Comparison, Outcome, and Study design (PICOS) framework. Eligible studies evaluated payment systems applied to hospitals or provider groups including hospitals. Interventions comprised modifications of existing DRG-based payment systems, including refinements of DRG classification (e.g., case-mix classification, reimbursement weights, risk adjustment, outlier payments), combinations of DRGs with other mechanisms (e.g., global budgets or fee-for-service [FFS] components), or broader reforms derived from DRG systems such as bundled payments. Temporary COVID-19-related modifications were excluded. Comparators were the original DRG system or alternative DRG designs. Outcomes included patient outcomes (e.g., mortality, complications), quality of care (e.g., readmissions), utilization (e.g., length of stay, post-acute care [PAC] use), costs, provider-related outcomes, and equity. Because randomized controlled trials (RCTs) are rarely feasible for payment reforms, we focused on quasi-experimental designs, incorporating controlled before-after (CBA) studies and interrupted time series (ITS) with at least three measurements before and after implementation, and considered cluster-randomized trials (cRCTs) where available. We included only studies published in English or German that analyzed data from more than one hospital.

Search strategy and study selection

We searched PubMed, Web of Science, EconLit, and the WHO Library Information System (WHOLIS) up to 31 August 2024, combining terms for DRG-based payment systems and hospitals with filters for quasi-experimental designs, and screened reference lists of included studies and relevant reviews (see Additional file 1, Tables S1-S4). Records were deduplicated in R [18] and screened in Rayyan [19] by two reviewers at title/abstract and full-text stages, with disagreements resolved by a third reviewer.

Data extraction

Data were extracted using piloted, standardized forms: one reviewer extracted study characteristics and a second extracted intervention effects, which the first verified independently. Of the reviewers involved, one was a statistician and the other had expertise in healthcare research. Discrepancies were resolved by discussion.

Risk of bias assessment

Risk of bias was assessed independently by two reviewers using ROBINS-I [20], informed by the Cochrane Effective Practice and Organisation of Care (EPOC) guidance [21]. Results were visualized with robvis [22]. RoB 2 had been planned for cRCTs [23] but none were eligible. Studies using random allocation implemented by the payer as part of the program, rather than by the investigators, were also classified as natural experiments and assessed with ROBINS-I.

Data synthesis

Studies were grouped by type of DRG modification and outcome. For each study, we tabulated the follow-up period, clinical indication, and effect estimate with its direction. Meta-analysis was planned but not performed because too few comparable studies were available within any intervention and condition. Findings were synthesized narratively and, where appropriate, visualized in forest plots without pooling effect estimates across studies. Estimates were displayed only if they were adjusted, referred to 90-day outcomes, were reported on the same effect measure as the respective figure, and provided a confidence interval or an exact p-value from which one could be calculated. When only subgroup estimates (e.g., by ethnicity) were reported, these were pooled into an overall estimate using inverse-variance weights. Certainty of evidence was assessed with GRADE, applying guidance for non-randomized studies evaluated with ROBINS-I [24]. Summary of Findings (SoF) tables were prepared for outcomes prioritized with patients and stakeholders through workshops [25] that rated outcome importance on a 1–9 GRADE scale (1 = not important; 9 = critical) [26]. Plain language statements were developed in accordance with the Supporting the Use of Research Evidence (SURE) framework to facilitate interpretation by policymakers [27, 28].

Patient and public involvement

Patient representatives and stakeholders participated in workshops to prioritize outcomes for the GRADE assessment and SoF tables to improve the relevance of the review for those affected [25, 26]. They were not involved in the search, screening, or data extraction.

Results

Study selection

The database search identified 7,532 unique records, and reference screening identified 38 more. After title and abstract screening, 96 full texts were assessed and 31 studies met the inclusion criteria (Fig. 1).

Study characteristics

All 31 studies were published between 2016 and 2024 and used quasi-experimental designs, mostly difference-in-differences (DiD) and, less often, ITS. No RCTs were identified. Most were conducted in the United States (n = 29), primarily within Medicare, with two from Sweden. Bundled payment initiatives constituted the majority of reforms (n = 30), while one study evaluated the introduction of Medicare Severity DRGs (MS-DRGs). Individual study characteristics are provided in Additional file 1 (Tables S5-S35).

Payment reform designs

Reform designs, grouped by intervention type, are summarized in Additional file 1 (Tables S36-S38). Most reforms were episode-based bundled payment programs (Table S36). Episode boundaries varied across programs. They ranged from the inpatient stay alone (Medicare Acute Care Episode [ACE] Demonstration) to a 33-day perioperative period (Carrum Health) and the index admission plus up to 90 days after discharge (BPCI Model 2, CJR, BPCI-A). Clinical scope also varied. CJR covered only lower extremity joint replacement (LEJR), whereas BPCI Model 2 and BPCI-A covered 48 and 34 clinical episode categories, respectively. BPCI-A also included outpatient and multi-setting episodes. The programs aimed to reduce spending and improve care coordination across the episode. Most reconciled fee-for-service payments retrospectively against a predefined target price and placed providers at two-sided financial risk. ACE instead used a prospective global payment, and the Carrum Health program used prospectively negotiated contracts. Participation was voluntary, except under CJR, which was mandatory in selected metropolitan statistical areas. Another included reform refined the DRG classification itself (Table S37). MS-DRGs added severity-based subgroups to improve payment accuracy while retaining the inpatient stay as the unit of payment. The Swedish patient choice reform combined bundled payment with free patient choice and free provider entry (Table S38), aiming to improve quality and reduce waiting times through competition.

Effects of interventions

Detailed study-specific estimates are provided in Additional file 1 (Tables S39-S46).

Across bundled payment programs, mortality, readmissions, and emergency department (ED) visits showed no consistent differential change, whereas total episode spending was more frequently reduced. Individual study estimates for BPCI, CJR, and BPCI-A, grouped by program, are shown for mortality (Fig. 2), readmissions (Fig. 3), ED visits (Fig. 4), and total episode spending (Fig. 5). Estimates from the ACE Demonstration (30-day outcomes) and the Carrum Health program (33-day perioperative episode) are reported narratively because their observation windows differ from 90 days. The Swedish reform is

likewise reported narratively because it assessed different endpoints (a composite adverse-event [AE] measure, EQ-5D, and patient satisfaction).

BPCI (Model 2)

Quality outcomes. BPCI participation was mostly not associated with statistically significant changes in mortality, readmissions, ED visits, complications, or measured case mix (Figures 2-4). Mortality was largely unchanged at 30 and 90 days [29-35], although one study reported small differential decreases for medical and surgical episodes (DiD −1.1 and −0.7 percentage points) [36], and a heart-failure bundle was associated with lower in-hospital mortality (OR 0.67), while 30- and 90-day all-cause mortality did not change differentially [37]. Readmission rates showed little to no change in most studies [29-38]. One study reported an increased 90-day readmission rate for lumbar fusion procedures (percentage change, risk-bearing BPCI vs. nonparticipant hospitals: +2.7% vs. −10.7%; p=0.043), which are characterized by higher clinical complexity [39], while another found fewer readmissions among medical episodes (DiD −0.8 percentage points) [36]. Complication rates and ED visits were largely unchanged across studies [29-31, 34, 35, 38], except for a larger increase in repeat spine surgery after lumbar fusion (risk-bearing BPCI vs. nonparticipant hospitals: +30.6% vs. +7.1%; p=0.043) [39] and opposing changes in ED visits after PCI and CABG (90-day DiD +7.2 and −12.1 percentage points) [32]. Most studies detected no meaningful shift in case mix [30, 32, 40].

Utilization and costs. BPCI was most consistently associated with reduced institutional PAC use for LEJR and selected medical conditions [29, 36, 38, 41, 42], reflected in lower total episode spending [29, 33-36, 38, 41-43] (Figure 5), driven mainly by fewer skilled-nursing and inpatient-rehabilitation services. In contrast, coronary artery bypass grafting (CABG) and percutaneous coronary intervention (PCI) procedures showed no statistically significant differences in episode or PAC spending [32, 43]. One study reported a small increase in home-health days (+0.4 days), suggesting substitution between PAC settings [36]. Most studies reported a small reduction or little to no difference in hospital length of stay (LOS) across procedures [29, 30, 32, 37, 38].

CJR

Quality outcomes. Most quality indicators showed no statistically significant change. Mortality [44-46], complication rates [44-47], and ED visits [44, 46, 47] were largely unchanged (Figures 2, 4). Findings on readmissions varied. One study found a small reduction (DiD −0.6 percentage points) [44] (Figure 3), while others found none [45, 47]. Within one study, reductions were confined to Black patients and not seen among White or Hispanic patients [46]. Indicators of patient selection were largely unchanged [44, 47, 48], although one study found a small decline in the share of beneficiaries enrolled through disability (DiD −0.6 percentage points) [44], whereas another reported slightly higher comorbidity among patients at participating hospitals [48].

Utilization and costs. CJR was most consistently associated with reduced institutional PAC use, particularly fewer skilled-nursing and inpatient-rehabilitation discharges [44, 46-49], reflected in lower episode spending driven by PAC savings [44, 46, 49] (Figure 5). Among patients still discharged to institutional care after hip arthroplasty, one study found that fewer returned to the community by the end of the episode (OR 0.91) [50]. Outside institutional care, one study found fewer home-health physical therapy visits per episode after hip and knee arthroplasty. After knee arthroplasty, more outpatient visits partly offset this reduction, but this increase was not robust in sensitivity analyses [51]. Hospital LOS was mostly unchanged, with only one study reporting a small but statistically significant reduction [44-46, 48, 49].

BPCI-A

Quality outcomes. Quality indicators consistently showed no statistically significant change in both the all-condition [52] and cardiovascular-focused [53] evaluations (Figures 2 and 3). ED visits and complications were not evaluated. Evidence on patient selection indicated no meaningful change in case mix or in the risk profile of treated patients [52, 53]. Another study reported changes in quality, healthy days at home, and access to care only as differences by race [54]. The sole statistically significant difference concerned healthy days at home, which increased slightly more among Black than among White patients (difference by race in slope change, −0.07 days per quarter; 95% CI −0.12 to −0.01) [54].

Utilization and costs. The all-condition study reported reduced spending on institutional PAC without a rise in home-health spending [52], whereas the cardiovascular study found no difference in PAC spending, skilled-nursing LOS, or healthy days at home [53]. At the level of total episode spending, effects were also small. The all-condition study found that quarterly payments declined more steeply among participating hospitals, with the trend change differing by $52 per quarter from non-participating hospitals (95% CI 34 to 70) [52]. The cardiovascular study found no statistically significant difference for either cardiac medical ($16; 95% CI −228 to 261) or cardiac procedural bundles ($171; 95% CI −429 to 772) (Figure 5) [53].

ACE Demonstration

Quality outcomes. Evidence was limited to a single study of cardiac and orthopedic episodes [55]. ACE participation was not associated with statistically significant changes in 30-day mortality or complications for either group. Readmissions significantly decreased for orthopedic procedures (DiD −2.22 percentage points) but not for cardiac surgery (DiD 0.21 percentage points). ED visits and patient selection were not evaluated.

Utilization and costs. Total episode spending did not change significantly for either orthopedic (DiD −$358) or cardiac episodes (DiD $514). PAC spending was significantly reduced in both groups

(orthopedic DiD −$591, 95% CI −$1,161 to −$22; cardiac DiD −$718, 95% CI −$1,431 to −$6) [55]. A second study likewise found no change in LEJR episode payments during the ACE period [42].

Carrum Health Employer-Provider Direct Payment Program

Quality outcomes. The single study of selected surgical procedures did not report most quality measures. Readmission rates at 30 days were reported only descriptively, with lower unadjusted rates than national averages but no risk-adjusted or causal estimates [56].

Utilization and costs. Participation was associated with a statistically significant reduction in total episode spending (−$4,229 per episode; −10.7%) and lower patient cost sharing (−$498 per episode) [56].

Swedish patient choice reform

Quality outcomes. One study found a small but statistically significant reduction in 90-day AEs (incl. death, further surgery, and complications) [57]. Neither study found statistically significant changes in patient-reported quality of life (EQ-5D) [58] or satisfaction [57, 58].

Utilization and costs. Evidence was limited to one study. Length of surgical admission declined in both groups, but less in the intervention group, indicating a relative increase in resource use (DiD 0.669 days, SE 0.258; $p<0.01$) [57].

Introduction of MS-DRGs

Provider behavior (upcoding). One study evaluating the introduction of MS-DRGs found a statistically significant increase in upcoding after the reform ($p<0.001$) [59].

Risk of bias and certainty of evidence

All studies were at moderate or serious risk of bias under ROBINS-I, mainly from potential residual confounding in non-randomized designs. Certainty of evidence was very low for all outcomes, chiefly because of risk of bias (see Additional file 1, Supplements 5-8).

Discussion

Summary of main findings

This review examined modifications to DRG-based payment systems, including bundled payment programs, DRG refinements, and mixed models combining DRGs with other reimbursement mechanisms. Most included studies considered bundled payments, particularly the Medicare initiatives

BPCI (Model 2), CJR, and BPCI-A, targeting mainly LEJR and selected surgical or medical conditions. Across these programs, effects on patient-important outcomes were generally small. Most studies found no statistically significant change in mortality, readmissions, complications, or ED visits, and no systematic risk selection. In contrast, many studies of BPCI and CJR found lower total episode spending, particularly for uncomplicated LEJR procedures and selected medical conditions in BPCI. These reductions were driven mainly by lower institutional PAC use and shorter PAC stays, while in-hospital LOS was largely unchanged. This pattern suggests that savings depended less on the payment mechanism itself than on the clinical composition of the episode and the room it left to reduce discretionary PAC. Consistent with this, BPCI-A, which covered a broader and more complex case mix, showed smaller or less consistent spending reductions.

Studies of the Swedish reform reported small effects, each based on limited evidence. Adverse events showed a small reduction, consistent with the two-year complication guarantee that providers bear, while patient-reported outcomes showed little to no change. The relative rise in LOS is likewise compatible with that guarantee, which discourages premature discharge. Because the reform combined a bundled-payment redesign with patient-choice mechanisms and measures to facilitate provider entry and competition, the effects attributable specifically to the payment redesign cannot be disentangled.

Evidence on DRG refinements was limited to one study, which found increased upcoding after the introduction of severity-adjusted DRG categories.

We found no eligible evaluations of mixed models combining DRGs with additional mechanisms, despite their use in several European healthcare systems [11].

Interpretation in relation to previous evidence

Our findings, drawn from quasi-experimental studies that can support causal inference when adequately conducted, are consistent with those of an earlier review of bundled payment initiatives in the United States [60]. Like our review, it found no association with quality of care (readmissions, ED visits, mortality), while spending reductions were more frequent, mainly through PAC utilization. This suggests that extending financial responsibility across an episode primarily reshapes service use and lowers spending rather than improving clinical outcomes, despite the premise that bundled payments enhance coordination of care and reduce fragmentation of services. Unlike that review, we examined a broader range of DRG modifications. Nonetheless, most included studies likewise evaluated bundled payments, reflecting the concentration of evidence in the United States context.

Policy implications

The findings of this review can provide valuable information for decision makers considering the future development and refinements of DRG-based payment systems.

First, bundled payments mainly influence resource use and spending, with little to no change in patient-important outcomes. Savings arose most consistently from reduced institutional PAC use, suggesting efficiency gains without compromising measured quality.

Second, effects varied by clinical condition, being most consistent for elective uncomplicated LEJR and mixed for other procedures, so careful selection of clinical episodes is critical when implementing or expanding these models, as previously highlighted in evaluations of United States programs [61].

Third, the financial incentives in bundled payment models may have implications for access to care. Although we did not find evidence of significant risk selection, case-based reimbursement creates incentives to avoid more complex and costly patients. Because bundled payments extend accountability across the episode of care, risk adjustment should account for complexity beyond the inpatient stay, particularly expected PAC needs, to ensure performance-based reimbursement [62, 63].

Fourth, the applicability of the study results to other health systems must be carefully considered. Almost all evidence comes from the United States and is based on Medicare claims, which mainly cover people aged 65 years and older. The populations studied may therefore be older, frailer, and less representative of those covered by other health systems. In addition, voluntary participation may introduce selection bias, as participating providers may have greater implementation capacity and more to gain [64]. Beyond these study-level limitations, the transferability of findings may also depend on the broader health-system context. Payment reforms may be implemented alongside other changes in health-system organization, making it difficult to isolate the effects attributable to the payment component itself, as illustrated by the Swedish reform. In addition, financing arrangements, baseline payment structures, provider organization, and regulatory environments differ across health systems and may directly influence the implementation and effects of payment reforms. Policymakers should therefore be cautious when transferring bundled payment models across settings and consider context-specific adaptation.

Fifth, evaluation is indispensable if health systems are to learn from payment reforms, yet it has rarely been conducted. Across more than three decades of DRG modifications, we identified only a few eligible evaluations, and none of mixed payment approaches that combine DRGs with other reimbursement mechanisms such as global budgets or FFS components. Because payment reforms are complex system interventions embedded in continuously changing contexts, their true effects are unpredictable and can be observed only after implementation. Continuous scientific monitoring of these effects, with modifications if results suggest unintended effects, should therefore be established as a permanent component of the payment system.

Limitations

This review has several limitations. First, certainty of evidence was very low, mainly because of non-randomized designs. Many evaluations used DiD, which relies on parallel-trends assumptions that may

be violated by concurrent policies or secular trends, biasing estimates [65]. Second, many United States studies drew on overlapping Medicare data and recurring research groups, so apparent consistency may reflect shared data and analytical choices rather than independent replication. We could not determine the extent of this overlap because the studies did not report the identity of the underlying beneficiary samples in sufficient detail.

Conclusion

Almost all included studies focused on bundled payment programs in the United States Medicare context. Although many countries have used DRGs for decades and have introduced modifications to address observed unintended effects, evidence from reliable quasi-experimental studies evaluating these modifications is largely absent. The studied bundled payment programs were frequently associated with lower total spending per case, particularly for uncomplicated LEJR and selected medical conditions. Savings per episode were driven mainly by reduced institutional PAC use. Broader programs such as BPCI-A and studies of some cardiac procedures showed no comparable savings. Patient-important outcomes, including mortality, readmissions, ED visits, and complications, showed little to no change. Our findings indicate that bundled payments for well-defined care pathways, such as those in surgical care, may reduce total episode spending without compromising measured quality.

Evidence on DRG refinements, mixed models, patient-reported outcomes, and equity was scarce or absent, and evidence outside the United States was limited to one combined Swedish reform. While some patterns emerged across models and indications, robust conclusions about the conditions under which these models are effective and the transferability of effects across settings are not yet possible. As new payment reforms are introduced, including mandatory episode-based models (e.g., the US TEAM model) and capacity-based payments largely decoupled from case volume (e.g., the German Vorhaltevergütung), the findings of this review may inform their design and underscore the need for rigorous evaluation.

Abbreviations

ACE Acute Care Episode

AE Adverse Events

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

CI Confidence interval

CJR Comprehensive Care for Joint Replacement

cRCT Cluster-Randomized Controlled Trials

DiD Difference-in-Differences

DRG Diagnosis Related Group

ED Emergency Department

EPOC Effective Practice and Organisation of Care

EQ-5D EuroQol 5-Dimension

FFS Fee-for-Service

GRADE Grading of Recommendations, Assessment, Development and Evaluations

ITS Interrupted time series

LEJR Lower Extremity Joint Replacement

LOS Length of stay

MS-DRG Medicare Severity-Diagnosis Related Group

OR Odds Ratio

OSF Open Science Framework

PAC Post-Acute Care

PCI Percutaneous coronary Intervention

PICOS Population, Intervention, Comparator, Outcome, Study design

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCT Randomized controlled trial

ROBINS-I Risk Of Bias In Non-randomized Studies of Interventions

SE Standard error

SoF Summary of Findings

SURE Supporting the Use of Research Evidence

TEAM Transforming Episode Accountability Model

WHOLIS WHO Library Information System

Declarations

Ethics approval and consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: All data supporting the findings of this study are available within the paper and its supplementary information file.

Competing interests: The authors declare that they have no competing interests.

Funding: This work was supported by the German Federal Ministry of Research, Technology and Space (BMFTR) (grant number 01KG2311).

Authors' contributions: TM conceived the review. TM and LS designed the review methodology. LS developed and ran the searches and managed the retrieved records. LS and RK screened records, extracted data, and assessed risk of bias; TM acted as third reviewer and resolved disagreements. LS and CR analyzed and synthesized the data. LS and CR prepared the figures. LS managed the data, coordinated the project, and drafted the manuscript. TM, RK and CR critically revised the manuscript. TM supervised the review and obtained funding. All authors approved the final version.

Acknowledgements: We thank the patient representatives and stakeholders who participated in our workshops and helped prioritize the outcomes for this review.

References

1. Liu Y, Wang G, Qin TG et al (2024) Comparison of diagnosis-related groups (DRG)-based hospital payment system design and implementation strategies in different countries: the case of ischemic stroke. Biosci Trends 18(1):1–10
2. Thompson JD, Averill RF, Fetter RB (1979) Planning, budgeting, and controlling—one look at the future: case-mix cost accounting. Health Serv Res 14(2):111–125
3. Busse R, Geissler A, Aaviksoo A et al (2013) Diagnosis related groups in Europe: moving towards transparency, efficiency, and quality in hospitals? BMJ 346:f3197
4. Scheller-Kreinsen D, Quentin W, Busse R (2011) DRG-based hospital payment systems and technological innovation in 12 European countries. Value Health 14(8):1166–1172
5. Messerle R, Schreyögg J (2022) System-wide effects of hospital payment scheme reforms: the German introduction of diagnosis-related groups. HCHE Research Paper No. 26. Hamburg: Hamburg Center for Health Economics. https://hdl.handle.net/10419/251509. Accessed 17 Jul 2024

6. Barber SL, Lorenzoni L, Ong P (2019) Price setting and price regulation in health care: lessons for advancing universal health coverage. Geneva: World Health Organization, Organisation for Economic Co-operation and Development. 10.1787/ed3c16ff-en
7. Baicker K, Chernew ME (2017) Alternative alternative payment models. JAMA Intern Med 177(2):222–223
8. Quentin W, Scheller-Kreinsen D, Blümel M et al (2013) Hospital payment based on diagnosis-related groups differs in Europe and holds lessons for the United States. Health Aff (Millwood) 32(4):713–723
9. Charlesworth A, Davies A, Dixon J (2012) Reforming payment for health care in Europe to achieve better value. Research report. Nuffield Trust
10. Edwards N, Honemann D, Burley D et al (1994) Refinement of the Medicare diagnosis-related groups to incorporate a measure of severity. Health Care Financ Rev 16(2):45
11. Quentin W, Stephani V, Berenson RA et al (2022) How Denmark, England, Estonia, France, Germany, and the USA pay for variable, specialized and low volume care: a cross-country comparison of in-patient payment systems. Int J Health Policy Manag 11(12):2940–2950
12. Press MJ, Rajkumar R, Conway PH (2016) Medicare's new bundled payments: design, strategy, and evolution. JAMA 315(2):131–132
13. Glickman A, Dinh C, Navathe AS (2018) The current state of evidence on bundled payments. LDI Issue brief 22(3):1–5
14. Centers for Medicare & Medicaid Services (2025) Bundled Payments for Care Improvement (BPCI) Initiative: General Information. https://www.cms.gov/priorities/innovation/innovation-models/bundled-payments. Accessed 5 Aug
15. Centers for Medicare & Medicaid Services (2025) Comprehensive Care for Joint Replacement Model. https://www.cms.gov/priorities/innovation/innovation-models/cjr. Accessed 5 Aug
16. Page MJ, McKenzie JE, Bossuyt PM et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372:n71
17. Schmidt LS, Kampo R, Mathes T (2025) Modifications of DRGs for reimbursing health care services in hospitals: protocol for a systematic review. Open Sci Framew. https://doi.org/10.17605/OSF.IO/G7D3Y
18. R Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing (2024) https://www.R-project.org/. Accessed 30 Sep 2024
19. Ouzzani M, Hammady H, Fedorowicz Z et al (2016) Rayyan—a web and mobile app for systematic reviews. Syst Rev 5(1):210
20. Sterne JAC, Hernán MA, McAleenan A et al (2019) Assessing risk of bias in a non-randomized study. In: Higgins JPT, Thomas J, Chandler J et al (eds) Cochrane handbook for systematic reviews of interventions, 2nd edn. Wiley, Chichester (UK), pp 621–641. doi:10.1002/9781119536604.ch25.

21. Cochrane Effective Practice and Organisation of Care (EPOC). Suggested risk of bias criteria for EPOC reviews (2017) http://epoc.cochrane.org/sites/epoc.cochrane.org/files/uploads/Resources-for-authors2017/suggested_risk_of_bias_criteria_for_epoc_reviews.pdf. Accessed 2 Dec 2025
22. McGuinness LA, Higgins JPT (2021) Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 12(1):55–61
23. Sterne JAC, Savović J, Page MJ et al (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366:l4898
24. Schünemann HJ, Cuello C, Akl EA et al (2019) GRADE guidelines: 18. How ROBINS-I and other tools to assess risk of bias in nonrandomized studies should be used to rate the certainty of a body of evidence. J Clin Epidemiol 111:105–114
25. Brütt AL, Meister R, Bernges T et al (2017) Patient involvement in a systematic review: development and pilot evaluation of a patient workshop. Z Evid Fortbild Qual Gesundhwes 127–128:56–61
26. Wiercioch W, Nieuwlaat R, Zhang Y et al (2022) New methods facilitated the process of prioritizing questions and health outcomes in guideline development. J Clin Epidemiol 143:91–104
27. Cochrane Effective Practice and Organisation of Care (EPOC). EPOC resources for review authors (2017) https://epoc.cochrane.org/resources/epoc-resources-review-authors. Accessed 12 Nov 2025
28. Santesso N, Glenton C, Dahm P et al (2020) GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. J Clin Epidemiol 119:126–135
29. Dummit LA, Kahvecioglu D, Marrufo G et al (2016) Association between hospital participation in a Medicare bundled payment initiative and payments and quality outcomes for lower extremity joint replacement episodes. JAMA 316(12):1267–1278
30. Joynt Maddox KE, Orav EJ, Zheng J et al (2018) Evaluation of Medicare's bundled payments initiative for medical conditions. N Engl J Med 379(3):260–269
31. Maughan BC, Kahvecioglu DC, Marrufo G et al (2019) Medicare's Bundled Payments for Care Improvement Initiative maintained quality of care for vulnerable patients. Health Aff (Millwood) 38(4):561–568
32. McNeely C, Orav EJ, Zheng J et al (2020) Impact of Medicare's bundled payments initiative on patient selection, payments, and outcomes for percutaneous coronary intervention and coronary artery bypass grafting. Circ Cardiovasc Qual Outcomes 13(9):e006171
33. Murphy WS, Siddiqi A, Cheng T et al (2019) 2018 John Charnley Award: analysis of US hip replacement bundled payments: physician-initiated episodes outperform hospital-initiated episodes. Clin Orthop Relat Res 477(2):271–280
34. Navathe AS, Emanuel EJ, Venkataramani AS et al (2020) Spending and quality after three years of Medicare's voluntary bundled payment for joint replacement surgery. Health Aff (Millwood) 39(1):58–66
35. Rolnick JA, Liao JM, Emanuel EJ et al (2020) Spending and quality after three years of Medicare's bundled payments for medical conditions: quasi-experimental difference-in-differences study. BMJ 369:m1780

36. Liao JM, Huang Q, Wang E et al (2022) Performance of physician groups and hospitals participating in bundled payments among Medicare beneficiaries. JAMA Health Forum 3(12):e224889
37. Oddleifson DA, Holmes DN, Alhanti B et al (2024) Bundled Payments for Care Improvement and quality of care and outcomes in heart failure. JAMA Cardiol 9(3):222–232
38. Navathe AS, Liao JM, Linn KA et al (2021) Spillover effects of Medicare's voluntary bundled payments for joint replacement surgery to patients insured by commercial health plans. Ann Intern Med 174(2):200–208
39. Martin BI, Lurie JD, Farrokhi FR et al (2018) Early effects of Medicare's Bundled Payment for Care Improvement program for lumbar fusion. Spine (Phila Pa 1976) 43(10):705–711
40. Navathe AS, Liao JM, Dykstra SE et al (2018) Association of hospital participation in a Medicare bundled payment program with volume and case mix of lower extremity joint replacement episodes. JAMA 320(9):901–910
41. Joynt Maddox KE, Orav EJ, Zheng J et al (2019) Post-acute care after joint replacement in Medicare's Bundled Payments for Care Improvement Initiative. J Am Geriatr Soc 67(5):1027–1035
42. Navathe AS, Troxel AB, Liao JM et al (2017) Cost of joint replacement using bundled payment models. JAMA Intern Med 177(2):214–222
43. Chopra Z, Gulseren B, Chhabra KR et al (2023) Bundled Payments for Care Improvement efficacy across 3 common operations. Ann Surg 277(1):e16–e23
44. Barnett ML, Wilcock A, McWilliams JM et al (2019) Two-year evaluation of mandatory bundled payments for joint replacement. N Engl J Med 380(3):252–262
45. Haas DA, Zhang X, Kaplan RS et al (2019) Evaluation of economic and clinical outcomes under Centers for Medicare & Medicaid Services mandatory bundled payments for joint replacements. JAMA Intern Med 179(7):924–931
46. Kim H, Meath THA, Tran FW et al (2020) Association of Medicare mandatory bundled payment system for hip and knee joint replacement with racial/ethnic difference in joint replacement care. JAMA Netw Open 3(9):e2014475
47. Finkelstein A, Ji Y, Mahoney N et al (2018) Mandatory Medicare bundled payment program for lower extremity joint replacement and discharge to institutional postacute care: interim analysis of the first year of a 5-year randomized trial. JAMA 320(9):892–900
48. Kiani SN, Maron SZ, Rosenzweig S et al (2024) Did payment reform lead to patient selection in hip and knee arthroplasties? An observational study using New York State data. HSS J 20(2):261–267
49. Meyers DJ, Kosar CM, Rahman M et al (2019) Association of mandatory bundled payments for joint replacement with use of postacute care among Medicare Advantage enrollees. JAMA Netw Open 2(12):e1918535
50. Kallies K, Dillingham TR, Edelstein A et al (2022) The effect of CMS' Comprehensive Care for Joint Replacement (CJR) bundled payment model on trajectories of post-acute rehabilitation care after total hip arthroplasty. Arch Phys Med Rehabil 103(12):2398–2403

51. Stern BZ, Banashefski B, Rozanski G et al (2024) Impact of the Comprehensive Care for Joint Replacement bundled payment model on postoperative utilization of home health and outpatient physical therapy services. Arch Phys Med Rehabil 105(9):1682–1690
52. Joynt Maddox KE, Orav EJ, Zheng J et al (2021) Year 1 of the Bundled Payments for Care Improvement-Advanced Model. N Engl J Med 385(7):618–627
53. Shashikumar SA, Zheng J, Orav EJ et al (2023) Changes in cardiovascular spending, care utilization, and clinical outcomes associated with participation in Bundled Payments for Care Improvement-Advanced. Circulation 148(14):1074–1083
54. Hammond G, Orav EJ, Zheng J et al (2022) Changes in racial equity associated with participation in the Bundled Payments for Care Improvement Advanced Program. JAMA Netw Open 5(12):e2244959
55. Chen LM, Ryan AM, Shih T et al (2018) Medicare's Acute Care Episode Demonstration: effects of bundled payments on costs and quality of surgical care. Health Serv Res 53(2):632–648
56. Whaley CM, Dankert C, Richards M et al (2021) An employer-provider direct payment program is associated with lower episode costs. Health Aff (Millwood) 40(3):445–452
57. Goude F, Garellick G, Kittelsen S et al (2022) Effects of competition and bundled payment on the performance of hip replacement surgery in Stockholm, Sweden: results from a quasi-experimental study. BMJ Open 12(7):e061077
58. Goude F, Kittelsen SAC, Malchau H et al (2021) The effects of competition and bundled payment on patient reported outcome measures after hip replacement surgery. BMC Health Serv Res 21(1):387
59. Cook A, Averett S (2020) Do hospitals respond to changing incentive structures? Evidence from Medicare's 2007 DRG restructuring. J Health Econ 73:102319
60. Agarwal R, Liao JM, Gupta A et al (2020) The impact of bundled payment on health care spending, utilization, and quality: a systematic review. Health Aff (Millwood) 39(1):50–57
61. Centers for Medicare & Medicaid Services. Bundled payments (2025) https://www.cms.gov/priorities/innovation/key-concepts/bundled-payments. Accessed 13 Mar 2026
62. Hendriks CMR, Koster F, Cattel D et al (2025) How do bundled payment initiatives account for differences in patient risk profiles? A systematic review. Value Health 28(4):652–669
63. de Brantes F, Delbanco SF (2017) The payment reform landscape: is the debate over retrospective versus prospective bundled payments a distraction? Health Aff Forefront. 23. 10.1377/forefront.20170623.060736
64. Wadhera RK, Yeh RW, Joynt Maddox KE (2018) The rise and fall of mandatory cardiac bundled payments. JAMA 319(4):335–336
65. Wing C, Simon K, Bello-Gomez RA (2018) Designing difference in difference studies: best practices for public health policy research. Annu Rev Public Health 39:453–469

Figures

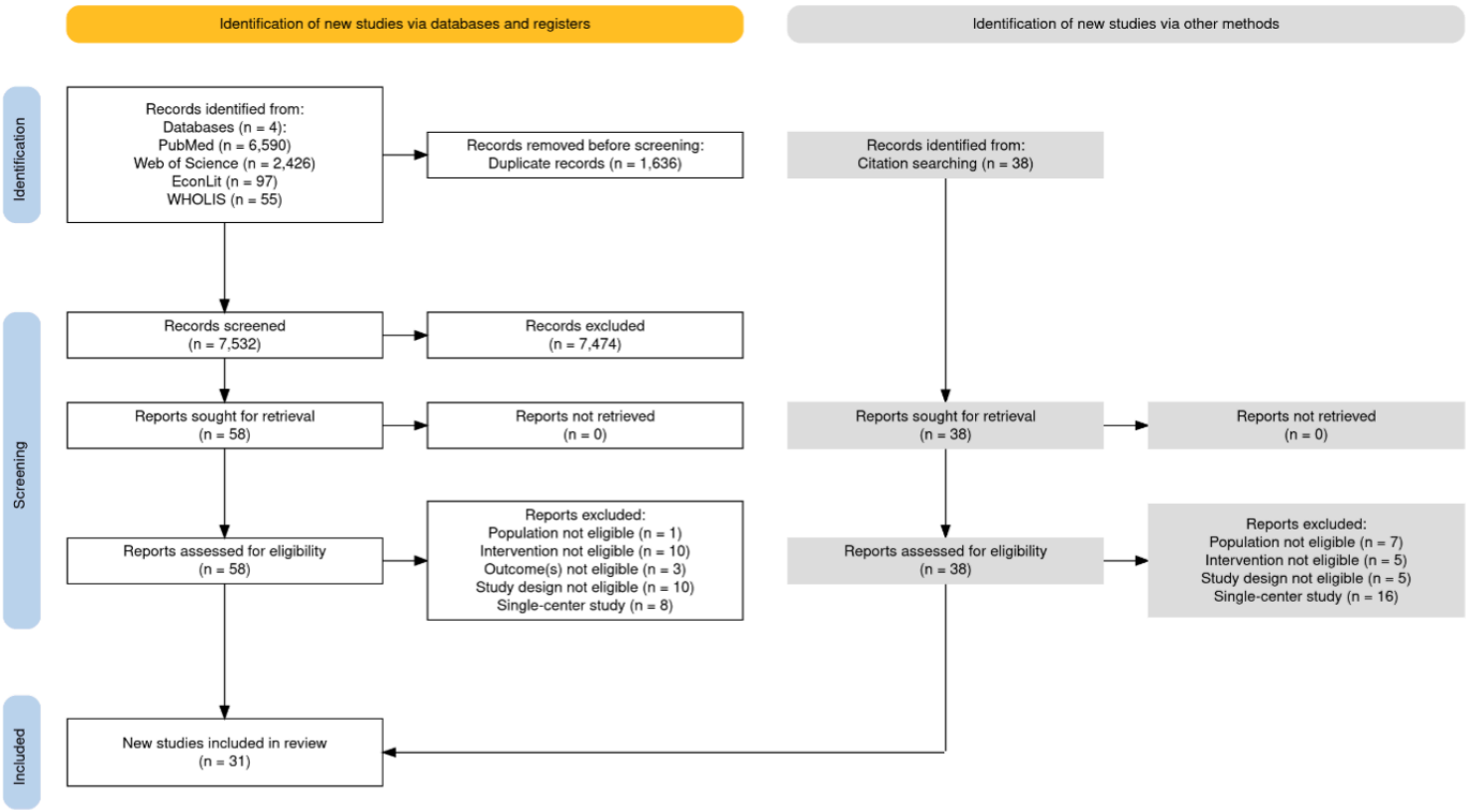


Figure 1

PRISMA 2020 flow diagram

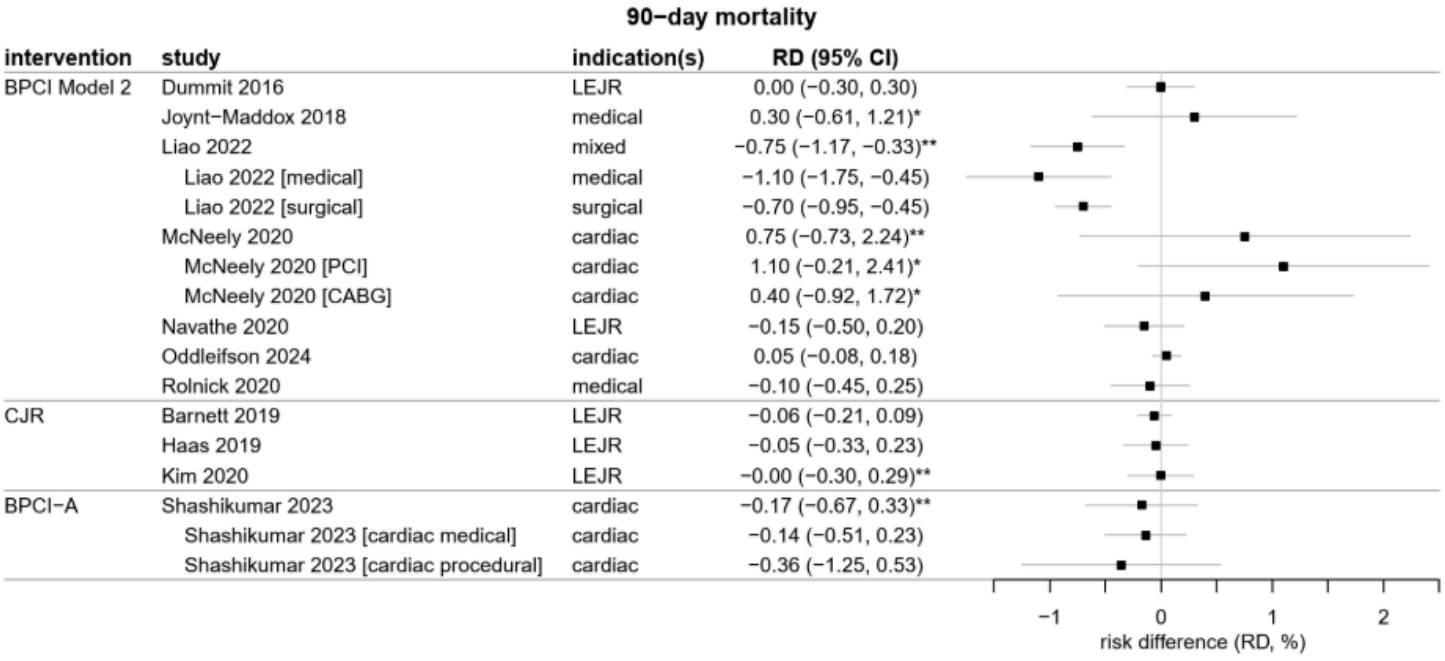


Figure 2

Mortality at 90 days across bundled payment programs

Rows show individual study estimates. The intervention column indicates the initiative (BPCI, CJR, BPCI-A) and the indication(s) column the patient's procedure or condition. No pooled estimate across studies is presented because of clinical heterogeneity. Estimates to the left favor the intervention. Effect measure: risk difference (RD, %).

* No confidence interval was reported. The 95% CI was reconstructed from the point estimate and the reported two-sided p-value assuming a normal sampling distribution.

** Pooled across the study's subgroups.

Abbreviations: RD, risk difference; CI, confidence interval; BPCI, Bundled Payments for Care Improvement; CJR, Comprehensive Care for Joint Replacement; BPCI-A, BPCI-Advanced; LEJR, lower-extremity joint replacement; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting.

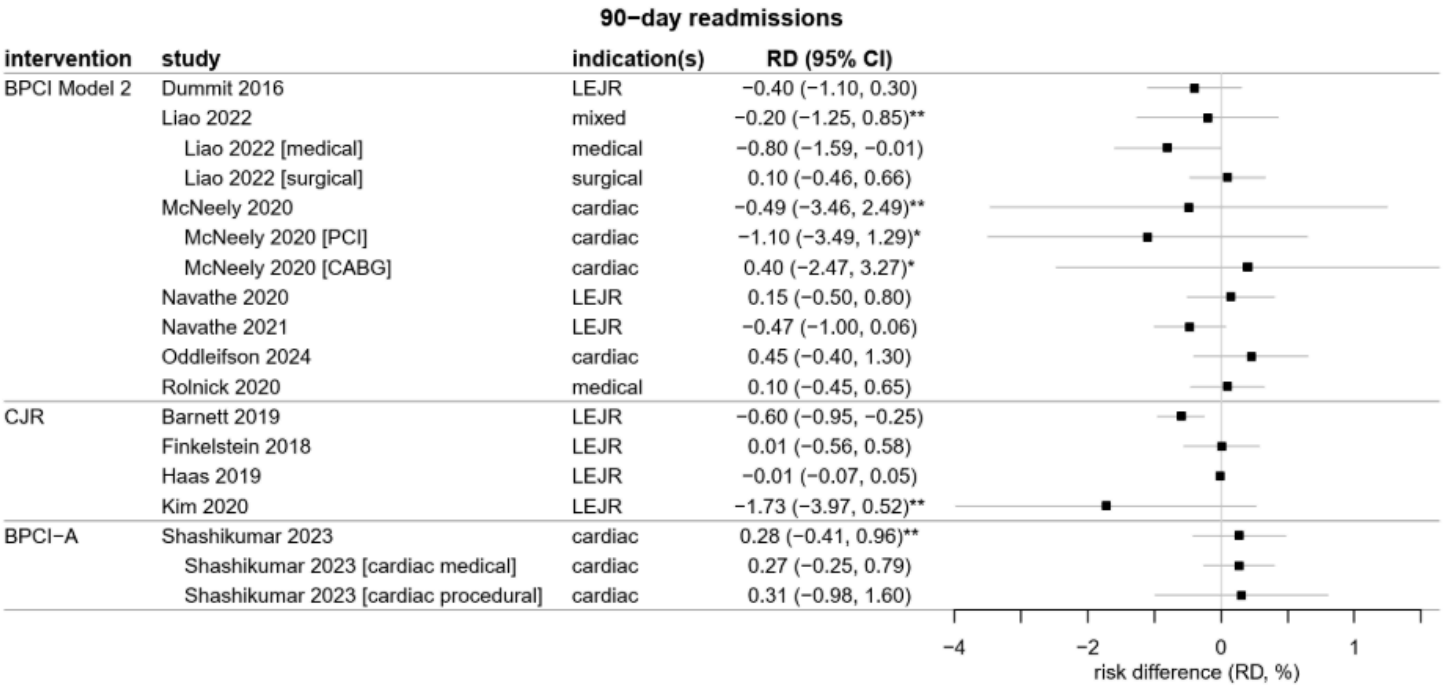


Figure 3

Readmissions within 90 days across bundled payment programs

Rows show individual study estimates. The intervention column indicates the initiative (BPCI, CJR, BPCI-A) and the indication(s) column the patient's procedure or condition. No pooled estimate across studies is presented because of clinical heterogeneity. Estimates to the left favor the intervention. Effect measure: risk difference (RD, %).

* No confidence interval was reported. The 95% CI was reconstructed from the point estimate and the reported two-sided p-value assuming a normal sampling distribution.

** Pooled across the study's subgroups.

Abbreviations: RD, risk difference; CI, confidence interval; BPCI, Bundled Payments for Care Improvement; CJR, Comprehensive Care for Joint Replacement; BPCI-A, BPCI-Advanced; LEJR, lower-extremity joint replacement; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting.

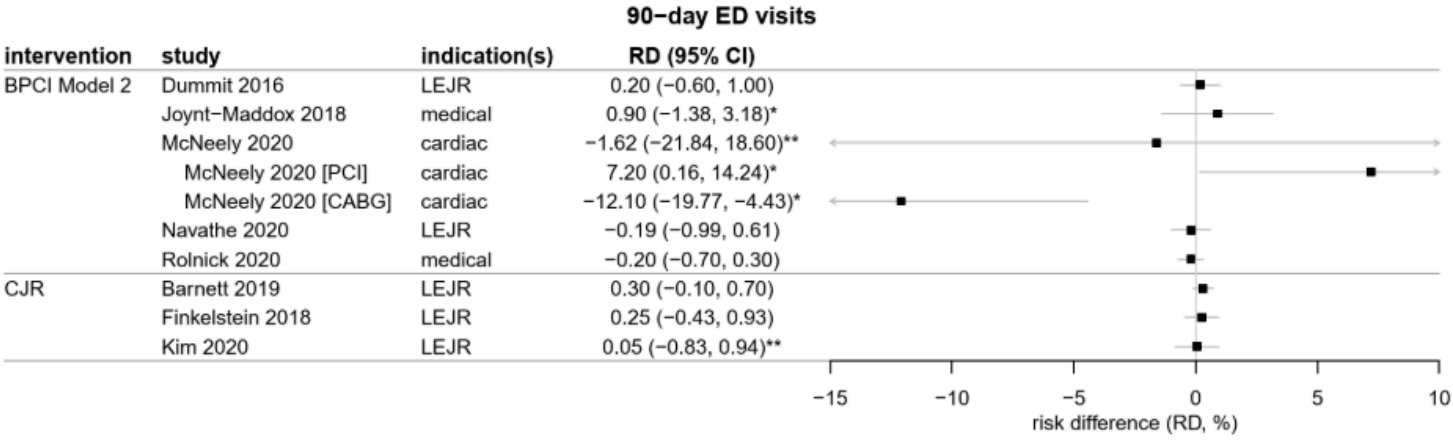


Figure 4

ED visits within 90 days across bundled payment programs

Rows show individual study estimates. The intervention column indicates the initiative (BPCI, CJR) and the indication(s) column the patient's procedure or condition. No pooled estimate across studies is presented because of clinical heterogeneity. Estimates to the left favor the intervention. Effect measure: risk difference (RD, %).

* No confidence interval was reported. The 95% CI was reconstructed from the point estimate and the reported two-sided p-value assuming a normal sampling distribution.

** Pooled across the study's subgroups.

Abbreviations: ED, emergency department; RD, risk difference; CI, confidence interval; BPCI, Bundled Payments for Care Improvement; CJR, Comprehensive Care for Joint Replacement; LEJR, lower-extremity joint replacement; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting.

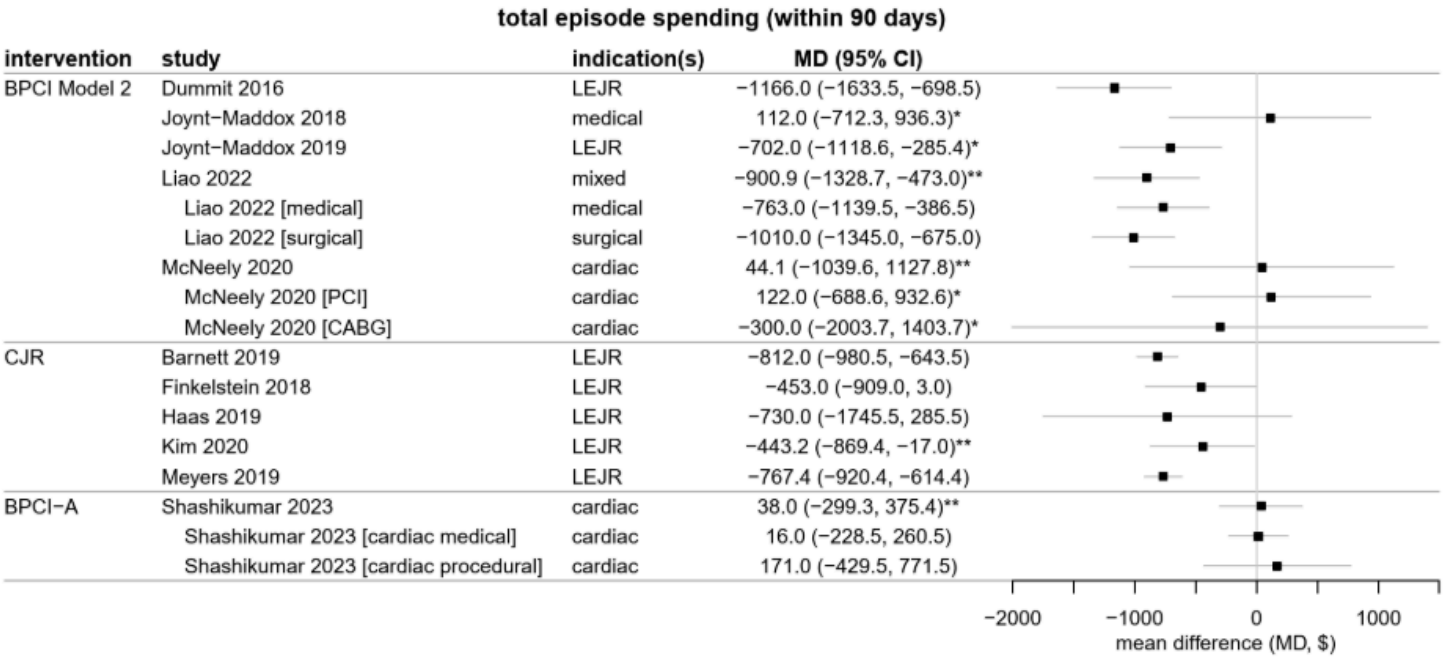


Figure 5

Episode-based spending across bundled payment programs (90-day episodes)

Rows show individual study estimates. The intervention column indicates the initiative (BPCI, CJR, BPCI-A) and the indication(s) column the patient's procedure or condition. No pooled estimate across studies is presented because of clinical heterogeneity. Estimates to the left favor the intervention. For Barnett 2019 and Joynt Maddox 2019, estimates refer to institutional (Medicare Part A) payments only. Effect measure: mean difference (MD, $).

* No confidence interval was reported. The 95% CI was reconstructed from the point estimate and the reported two-sided p-value assuming a normal sampling distribution.

** Pooled across the study's subgroups.

Abbreviations: MD, mean difference; CI, confidence interval; BPCI, Bundled Payments for Care Improvement; CJR, Comprehensive Care for Joint Replacement; BPCI-A, BPCI-Advanced; LEJR, lower-extremity joint replacement; PCI, percutaneous coronary intervention; CABG, coronary artery bypass grafting.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- DRGmodifyReviewSupplementalAppendix.pdf