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.

Impact of Value-Based Payment Models on Short-Term Costs and Clinical Outcomes in Surgical Oncology: A Systematic Review

Julian Musa

Julian.Musa@chiru.med.uni-giessen.de

Department of General, Visceral, Thoracic, and Transplant Surgery, University Hospital of Giessen and Marburg, Giessen site, Giessen, Germany

Tim Goedert

Department of General, Visceral, Thoracic, and Transplant Surgery, University Hospital of Giessen and Marburg, Giessen site, Giessen, Germany

Toby A. Gordon

Johns Hopkins Carey Business School, Baltimore, MD, USA

Rainer Sibbel

Frankfurt School of Finance and Management, Frankfurt, Germany

Martin Schneider

Department of General, Visceral, Thoracic, and Transplant Surgery, University Hospital of Giessen and Marburg, Giessen site, Giessen, Germany

Afschin Gandjour

Frankfurt School of Finance and Management, Frankfurt, Germany

Systematic Review

Keywords: Value-based healthcare, value-based payment, value-based reimbursement, VBHC, VBP, surgical oncology, cancer surgery, fee-for-service, bundled payments, accountable care organizations, ACO, capitation

Posted Date: March 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8989147/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.

Impact of Value-Based Payment Models on Short-Term Costs and Clinical Outcomes in Surgical Oncology: A Systematic Review

Julian Musa, MD[1-4], Tim Goedert[1], Toby A. Gordon, ScD[5], Rainer Sibbel, PhD, MBA[2], Martin Schneider, MD, MBA[1], Afschin Gandjour, MD, PhD, MBA[2§]

Author affiliations:
1 Department of General, Visceral, Thoracic, and Transplant Surgery, University Hospital of Giessen and Marburg, Giessen site, Giessen, Germany
2 Frankfurt School of Finance and Management, Frankfurt, Germany
3 Division of Translational Pediatric Sarcoma Research (B410), German Cancer Research Center (DKFZ) Heidelberg, Germany
4 Hopp-Children's Cancer Center (KiTZ), Heidelberg, Germany
5 Johns Hopkins Carey Business School, Baltimore, MD, USA

§ Address for correspondence
Julian Musa, MD
Department of General, Visceral, Thoracic, and Transplant Surgery
University Hospital of Giessen and Marburg
Rudolf Buchheim Str. 7, 35392 Giessen, Germany
Phone: +49-641-985-44701
Email: Julian.Musa@chiru.med.uni-giessen.de

Word count: 5,453 (excluding Acknowledgments, References, and Figure Legends)
Display items: 2 Figures, 3 Tables, 6 Supplementary Tables
Key words: Value-based healthcare, value-based payment, value-based reimbursement, VBHC, VBP, surgical oncology, cancer surgery, fee-for-service, bundled payments, accountable care organizations, ACO, capitation.
Running title: Value-based payment in surgical oncology.
Conflict of interests: We declare no conflicts of interest.

ABSTRACT

Background:

Value-based payment (VBP) models aim to align financial incentives with quality and efficiency of care, yet its impact on surgical oncology, a field defined by clinical complexity and high utilization of resources, remains poorly understood. Therefore, this systematic review aims to synthesize evidence in the field regarding the impact of VBP on costs and short-term surgical outcomes.

Methods:

We conducted a systematic review following PRISMA guidelines (PROSPERO CRD420261296628). MEDLINE, EMBASE, Web of Science, and CENTRAL were searched from January 2006 through December 2025 for studies containing quantitative data evaluating the impact of VBP models on costs and clinical outcomes in surgical oncology. Two reviewers independently screened, extracted data, and assessed risk of bias using ROBINS-I. Due to substantial conceptual and methodological heterogeneity across VBP models and outcome reporting, a qualitative synthesis was performed. Primary outcomes included short-term expenditures (≤90 days), mortality, complications, readmissions, and length of stay measures.

Results:

Seven studies reflecting six distinct cohorts were included, predominantly retrospective quasi-experimental analyses using difference-in-difference (DID) approaches. Across studies, participation in VBP models was not associated with statistically significant changes in short-term treatment expenditures, including total costs, index hospitalization costs, readmission-related spending, or skilled nursing facility (SNF) expenditures. Clinical outcomes were likewise largely unaffected: no consistent effect on 30- or 90-day mortality, postoperative complications, readmissions, or length of stay measures were observed. One study reported a

modest reduction in 30-day readmissions under a capitation-based model and another study found improved oncological outcomes in a pay-for-performance (P4P) program; however, the latter demonstrated serious risk of bias. Overall, the available evidence did not demonstrate consistent financial or clinical advantages of VBP implementation in surgical oncology.

Conclusion:

Current evidence does not show consistent short-term financial or clinical benefits of VBP in surgical oncology. However, methodological and conceptual constrains limit definitive conclusions. Oncology-specific payment models and rigorous evaluations incorporating additional long-term, patient-centered outcomes are needed to clarify the role of VBP in cancer surgery.

INTRODUCTION

Healthcare systems worldwide are under increasing pressure to improve patient outcomes while controlling rising costs[1–3]. Traditional fee-for-service (FFS) reimbursement models have been criticized for incentivizing quantity rather than quality of care, potentially leading to unnecessary procedures, fragmentation of care, and non-optimal outcomes[4–6]. In response, value-based payment (VBP) models have been proposed in 2006 by Porter & Teisberg as a strategic shift to align financial incentives with patient outcomes and cost-effectiveness, defining value as outcomes relative to cost[4–6].

VBP encompasses heterogeneous reimbursement models that can be broadly grouped into financial risk-structuring arrangements and performance/accountability-based incentives, which differ in how directly they operationalize value. Bundled or episode-based payments and capitation primarily redefine the unit of payment and allocate financial risk to providers, thereby incentivizing cost control, care coordination, and, depending on design, complication and readmission prevention across an episode or population[7–16]. They are not inherently value-based but can support value creation when embedded within an appropriate performance framework including outcome measurement, performance bench-marks, and/or accountability mechanisms[7–16]. In contrast, pay-for-performance (P4P) models explicitly tie reimbursement to achievement on defined quality or outcome metrics[17,18], while accountable care organizations (ACOs) and shared savings arrangements hold provider organizations jointly accountable for the quality and total cost of care across a defined population[19,20]. Together, these models aim to align financial incentives with patient outcomes and resource efficiency, either by restructuring payment units, embedding performance benchmarks, or combining both within layered reimbursement frameworks[21–25].

In surgical oncology, where care is highly complex and resource-intensive, aligning reimbursement with value is particularly relevant yet challenging due to high complication risk, ethical concerns around underuse of care, and potential consequences for long-term oncologic

outcomes[26–28]. While short-term clinical endpoints such as readmission rates, mortality, and complication rates represent clinically meaningful and widely accepted outcome measures, surgical outcomes are influenced by a complex interplay of factors, including tumor entity and stage, procedural complexity, perioperative therapies, and patient characteristics[26,29–31]. At the same time, outcome variability and ethical considerations regarding what constitutes meaningful benefit further complicate the classical definition of value as outcomes relative to cost[26,29–31]. Given this background of diverse VBP models and their application in surgical oncology, the definition of value in this field remains the subject of ongoing clinical, ethical, and health policy debate. Evidence from recent systematic reviews and the increasing use of patient-reported measures supports a multidimensional framework of value in surgical oncology, encompassing oncological outcomes, surgical outcomes, patient-reported outcomes (PROMs) and experiences (PREMs), and utilization measures such as length of hospital stay[26,30–37]. These domains are intrinsically interconnected, as perioperative complications can adversely affect survival, quality of life, and downstream healthcare use, thereby amplifying their combined impact on overall value. Conceptually, treatment value can be understood as the balance between beneficial and adverse outcomes across these domains in relation to costs; however, not all outcomes contribute equally. Available evidence indicates that patients and clinicians place greatest weight on survival and functional well-being, with complication avoidance also highly prioritized, while system-level efficiency measures are generally considered supportive rather than primary determinants of value[30].

Within this multifaceted context, whether VBP meaningfully improves care in surgical oncology remains an open and highly relevant question. Although VBP models are conceptually appealing, discrepancies between how quality and value are defined and how they are actually measured, heterogeneous outcome reporting, and the frequent inclusion of surgical oncology within broader, non-specific study populations have resulted in an incomplete and fragmented evidence base regarding their effects on clinical outcomes and costs.

By addressing this gap, this systematic review aims to move beyond theoretical appeal toward an evidence-informed understanding of how value is currently measured -and how it might be more appropriately defined- in surgical oncology, and whether VBP meaningfully improves patient outcomes and treatment efficiency to inform health policy.

METHODS

This systematic review was conducted according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines[38]. The study was registered and the protocol was published at the PROSPERO database under the accession number CRD420261296628.

Literature Search, Study Selection, and Data Extraction

MEDLINE (via PubMed), EMBASE, Web of Science, and the Cochrane Central Register of Controlled Trials (CENTRAL) were systematically searched for relevant studies until 13th of December 2025 by two independent reviewers (J.M. and T.G.). Detailed search terms for each database are available in the PROSPERO record (CRD420261296628). The search strategy was limited to English-language articles published from January 1, 2006 onward, as VBP models were first introduced in 2006[4]. For evaluating the effect of VBP-models on outcome parameters, only quantitative data has been used for systematic review. Data were extracted independently by two reviewers (J.M. and T.G.) using a standardized extraction form. Extracted variables included study characteristics (design, cohort sizes, data sources, country, study period), clinical characteristics (patient and hospital characteristics, comorbidity burden, included cancer entities), methodological information (analytic approaches and adjustment strategies), and clinical, economic, and utilization outcome measures (mortality, complication rates, readmissions, total episode costs, index hospitalization costs, post-acute care expenditures, length of stay measures). For cost outcomes, adjusted DID estimates with

corresponding 95% confidence intervals were extracted when available. Disagreements between both reviewers were resolved by discussion and consensus.

Study Design

This systematic review synthesized quantitative evidence on the impact of value-based payment (VBP) models on economic, clinical, and utilization outcomes in surgical oncology. Most included studies applied quasi-experimental DID designs, comparing pre- and post-implementation changes in outcomes between VBP-participating providers (intervention group) and contemporaneous non-VBP providers (control group) to estimate the association between VBP participation and outcome trends over time.

Given the limited number of studies specifically evaluating VBP models in surgical oncology and the substantial heterogeneity across interventions, a qualitative systematic synthesis was undertaken without formal meta-analysis. The included interventions comprised structurally distinct payment models, including P4P programs, bundled payments, ACOs, and capitation-based arrangements. Although these models share the overarching objective of linking reimbursement to quality and efficiency, which justifies their consideration as a class of value-based interventions, they differ fundamentally in incentive design, financial risk allocation, scope of accountability (episode-based vs. population-based), and implementation context. Consequently, they cannot be considered a single, uniform “treatment” exposure for the purpose of quantitative pooling. Beyond conceptual heterogeneity, substantial methodological variability, including non-uniform effect measures and follow-up periods as well as insufficient validation of key DID assumptions such as parallel trends, did not permit harmonization of point estimates and variance parameters, which further precluded quantitative formal meta-analysis. Pooling such estimates would have risked compounding bias rather than improving inferential precision.

Accordingly, no formal meta-analysis was performed. Instead, findings were synthesized qualitatively across studies and VBP models, with emphasis on the direction, magnitude, and statistical significance of reported associations, while carefully considering differences in intervention design, analytic approach, and risk of bias.

Outcome Measures

The review evaluated treatment expenditures, including overall episode-based costs and specific components such as index hospitalization, readmission-related care, and skilled nursing facility (SNF) expenditures, as well as surgical outcomes (mortality, complications, and readmissions), oncological survival outcomes, utilization measures (e.g., length of stay, post-anaesthesia care unit (PACU) utilization, intensive care unit (ICU)), and patient-reported outcomes and experiences (PROMs/PREMs). These domains were screened across all included studies, and if available, relevant data were extracted. Data were synthesized qualitatively and reported only when more than two included studies provided outcomes in a sufficiently consistent and methodologically compatible format.

Risk of Bias Assessment

The risk of bias was assessed using the Cochrane risk of bias tool ROBINS-I (v2)[39] by two independent reviewers (J.M. and T.G.). Disagreements were resolved by discussion and consensus. The assessment was based on the following domains: bias arising from confounding, bias due to classification of interventions, bias due to selection of participants, bias due to missing data, bias arising from measurement of the outcome, bias in selection of reported results. Based on these domains, an overall risk of bias was determined for each included study.

For studies using a DID design, a structured checklist was applied to assess key methodological assumptions and analytical rigor. This included evaluation of treatment timing and handling of staggered adoption, justification and assessment of the parallel trends

assumption (including event-study plots and pre-trend diagnostics), sensitivity analyses for potential violations, and appropriate clustering of standard errors.

RESULTS

Study characteristics

A total of 4,967 references were identified by the literature search after duplicate removal. 21 references were full-text reviewed and 7 studies were included for final analysis (PRISMA flow-chart, **Figure 1**). The design of all seven studies included are displayed in **Table 1** with a corresponding DID checklist[40] displayed in **Supplementary Table 1**, whereas basic characteristics from included patient and hospital cohorts are displayed in **Supplementary Tables 2,3**[41–47].

Across studies, data were derived from large population-based administrative and registry sources. Kuo et al. used linked Taiwanese cancer registry and national insurance claims data[46], whereas all other studies relied on US Medicare claims, often supplemented with hospital-level datasets (e.g., American Hospital Association (AHA) or Centers for Medicare & Medicaid Services (CMS) files) to characterize facilities and identify VBP participation[41–45,47]; Offodile et al. additionally incorporated state-level data within Maryland's Global Budget Revenue (GBR) program[41]. All analyses were retrospective, evaluating VBP implementation versus conventional payment models with respect to financial, clinical, and/or utilization outcomes (**Table 1**). Except for Kuo et al., studies applied quasi-experimental DID designs comparing pre- and post-implementation changes between VBP-participating and non-participating providers. In terms of model types, one study assessed a P4P program (Breast Cancer - P4P; BC-P4P) in Taiwan[46], one evaluated bundled payments (Bundled Payments for Care Improvement - Advanced; BPCI-A)[42] in the US, and the remaining five studies (four cohorts) examined ACO participation in the US[43–45,47] (**Table 1**). Two of the ACO studies conducted by Worku et al. were based on the same patient cohort and were therefore treated as

a single study[43,44]. Four of the seven studies (representing three of six cohorts) applied propensity score matching (PSM) to hospital characteristics prior to analysis[41–44], whereas the remaining studies used multivariable regression adjustment, most commonly within DID frameworks, to control for confounders[45–47] (**Table 1**). The studies include a broad range of cancer entities, with gastrointestinal and genitourinary cancers most frequently represented (**Table 1**).

Across studies, patient baseline characteristics were largely comparable between the intervention (VBP) and comparison (non-VBP) groups (**Supplementary Table 2**). Age distribution was similar in most cohorts, although Kuo et al. reported statistically significant differences in mean age ($P = 0.022$) and Katragadda et al. observed significant differences in age category distributions ($P < 0.001$). Robbins et al., Worku et al., and Offodile et al. reported no significant age differences. Gender distributions were not significantly different in most studies, except for a small but statistically significant higher proportion of female patients in the VBP-group in Katragadda et al. ($P < 0.001$). Also, Kuo et al. included exclusively female patients due to the focus on breast cancer. Comorbidity burden was comparable across groups in most cohorts, although Kuo et al. ($P = 0.004$) and Katragadda et al. ($P < 0.001$) reported statistically significant differences in comorbidity measures; absolute differences were modest. Overall, although some statistically significant baseline differences were observed, these were small in absolute terms and unlikely to be clinically relevant, such that intervention (VBP) and comparison (non-VBP) groups were broadly comparable across studies.

Hospital characteristics were reported in varying formats across studies but showed more pronounced differences between VBP and non-VBP groups in studies that did not apply hospital level PSM (**Supplementary Table 2**). Where profit status/ownership was reported, studies without PSM (Herrel et al., $P < 0.001$; and Katragadda et al., $P < 0.001$) showed that VBP hospitals were more often non-for-profit, while non-VBP hospitals included higher proportions of for-profit and public institutions. In contrast, Robbins et al., which applied PSM,

showed no such differences. Hospital size also differed across several studies. In Robbins et al., Herrel et al. ($P = 0.02$), and Katragadda et al. ($P < 0.001$) (two of which did not use PSM), VBP hospitals were more frequently medium to large institutions, whereas non-VBP hospitals were more frequently small. By comparison, the PSM-adjusted cohorts analyzed by Worku et al. and Offodile et al. did not demonstrate significant size differences between groups. Hospital treatment volume was inconsistently reported and generally similar between groups, although Herrel et al. noted higher median oncologic surgical volumes among VBP hospitals, this difference did not reach statistical significance ($P = 0.06$). Teaching status showed differences in several cohorts, again primarily in studies without PSM. VBP-participating hospitals are more frequently teaching institutions in Robbins et al., Herrel et al. ($P = 0.002$), and Katragadda et al. ($P < 0.001$) (two of which did not use PSM), while other studies reported no meaningful differences. An exception was Kuo et al. where teaching hospitals were more frequent in the non-VBP group ($P < 0.001$). Overall, hospital characteristics differed between VBP and non-VBP groups, particularly in studies without hospital-level PSM, with VBP hospitals more often being larger, non-for-profit, and teaching institutions.

Risk of bias assessment using the Cochrane ROBINS-I tool[39] revealed that across included studies, the overall risk of bias was generally moderate (**Figure 2**). Most evaluations used quasi-experimental designs with DID analyses including extensive adjustment and, in some cases, PSM approaches, which mitigated but did not eliminate concerns related to residual confounding and selection bias (**Figure 2**). Across studies, the domains of classification of interventions, missing data, outcome measurement, and selection of reported results were consistently rated as low risk of bias due to the use of standardized administrative data (**Figure 2**). Kuo et al. demonstrated severe risk of bias, driven primarily by confounding and participant selection, whereas Offodile et al. demonstrated the lowest overall risk of bias, with only minor concerns related to participant selection (**Figure 2**). In addition, assessment of each DID study using a structured DID-specific checklist demonstrated that key identifying assumptions were

frequently insufficiently documented (**Supplementary Table 1**). In particular, most studies did not explicitly justify the parallel trends assumption, present event-study plots, conduct formal pre-trend diagnostics, or report sensitivity analyses assessing robustness to potential violations of identifying assumptions (**Supplementary Table 1**), thereby limiting confidence in the causal interpretation of reported DID estimates.

Treatment Expenditures

Information on treatment expenditures was available from five of seven studies (reflecting four of six cohorts)[41–44,47], whereby one study only included expenditures for SNF treatment[47]. Treatment expenditure outcomes across studies are summarized in **Table 2** as well as **Supplementary Tables 4,5** and were reported for short-term total spending (≤ 90 days, **Table 2**), index hospitalization and readmission costs, and SNF expenditures (**Supplementary Tables 4,5**). All analyses were based on DID estimates comparing pre- and postimplementation changes between VBP and non-VBP groups.

Across studies assessing short-term spending (≤ 90 days), no statistically significant differences were identified between VBP and non-VBP groups (**Table 2**). Robbins et al. reported a non-significant DID estimate of +$316 for 90-day episode spending (95% CI −2,076 to 2,709; $P > 0.05$, **Table 2**). Similarly, Worku et al. observed a non-significant reduction of −$246 in 30-day episode expenditures (95% CI −1,497 to 1,005; $P = 0.69$, **Table 2**). Offodile et al. reported a DID estimate of +$646 for 30-day episode costs (95% CI −208 to 1,542; $P = 0.14$, **Table 2**), which also did not reach statistical significance. Overall, VBP implementation was not associated with significant changes in short-term expenditures compared with conventional payment models (**Table 2**).

For index hospitalization costs, Robbins et al. reported a non-significant DID estimate of −$430 (95% CI −1,132 to 272; $P > 0.05$, **Supplementary Table 4**). Regarding readmission expenditures, Robbins et al. observed a non-significant increase of +$757 for 90-day

readmission costs (95% CI −274 to 1,787; $P > 0.05$, **Supplementary Table 4**). Worku et al., analyzing combined index and 30-day readmission expenditures, found a non-significant DID estimate of −$123 (95% CI −1,280 to 1,035; $P = 0.83$, **Supplementary Table 4**). Thus, no consistent or statistically significant effects of VBP participation were identified for index hospitalization or readmission-related spending (**Supplementary Table 4**).

For SNF-related expenditures, Robbins et al. reported a non-significant DID estimate of −$204 at 90 days (95% CI −1,210 to 802; $P > 0.05$, **Supplementary Table 5**). Worku et al. observed a non-significant reduction of −$447 at 30 days (95% CI −1,335 to 474; $P = 0.32$). Katragadda et al. similarly reported no significant differences in SNF expenditures across prostate, bladder, and kidney cancer surgeries, with DID estimates of $40.4 ($P = 0.12$), $42.56 ($P = 0.93$), and −$92 ($P = 0.48$), respectively, whereby **Supplementary Table 5** presents the results as ranges of SNF expenditures across entities. Overall, VBP participation was not associated with statistically significant changes in SNF-related expenditures (**Supplementary Table 5**).

In summary, across five studies, VBP participation was not associated with significant changes in short-term total, index, readmission, or SNF expenditures, with all DID estimates remaining non-significant.

Clinical outcomes

Oncological outcomes, such as overall survival (OS) and recurrence-free survival (RFS) were only reported by one of seven studies (reflecting one of six cohorts)[46]. Most studies reported surgical outcomes (six of seven studies, reflecting 5 of 6 cohorts)[41–45,47], including mortality rates (five of seven studies, four of six cohorts)[41–45], complication rates (four of seven studies, three of six cohorts)[43–45,47], and readmission rates (six of seven studies, five of six cohorts)[41–45,47]. Notably, no study reported PROMs or PREMs.

Oncological outcomes

Kuo et al. was the only study included reporting on oncological outcomes[46]. The study reported that enrollment in Taiwan's nationwide breast cancer pay-for-performance (BC-P4P) program was associated with improved 5-year OS (OR 0.167; 95% CI 0.064–0.432; $P < 0.001$) and reduced RFS (OR 0.370; 95% CI 0.200–0.685; $P = 0.002$) after adjustment for patient characteristics, quality of care, and propensity scores. However, this study was found to have the highest overall risk of bias, primarily due to confounding and selection bias (**Figure 2**). As a result, the validity of these findings is uncertain, and causal inferences regarding oncological outcomes should be made with caution.

Surgical outcomes

Surgical outcomes were reported across most included studies (**Table 3**)[41–45,47]. Outcomes included short-term mortality (30- and 90-day), postoperative complications (index, 30-day, and 90-day), and readmissions (30- and 90-day). All analyses were based on DID estimates comparing pre- and post-implementation changes between VBP and non-VBP groups.

Short-term mortality rates were evaluated in five studies (reflecting four cohorts)[41–45]. Robbins et al. reported a DID estimate of −1.22 percentage points for 90-day mortality (95% CI −3.60 to 1.17; $P > 0.05$), showing no statistically significant difference between VBP and non-VBP groups (**Table 3**). Herrel et al. likewise observed no significant association for 30-day mortality, with DID estimates reported separately for each cancer entity with all $P > 0.05$. In **Table 3**, DID-estimates and CIs were given as ranges across cancer entities. Worku et al., analyzing both 30- and 90-day mortality, reported relative risk ratios (RRR) close to unity (30-day RRR 0.96, 95% CI 0.77–1.21; $P = 0.745$; 90-day RRR 1.03, 95% CI 0.85–1.25; $P = 0.777$, **Table 3**), demonstrating no significant effect. Offodile et al. similarly found no statistically significant difference in 30-day mortality (DID −0.4; 95% CI −1.4 to 0.5; $P = 0.360$, **Table 3**).

Overall, none of the included studies demonstrated a statistically significant short-term mortality reduction attributable to VBP implementation.

Postoperative complication rates were reported in four studies (representing six cohorts)[43–45,47]. Herrel et al. reported no significant association for 30-day complications, with DID estimates presented separately for each cancer entity and all $P > 0.05$. In **Table 3**, DID estimates and corresponding CIs are shown as ranges across cancer entities. Worku et al., analyzing index admission, 30-day, and 90-day complications, reported RRR approximating 1.0 across all time points (index RRR 1.03, 95% CI 0.90–1.18; $P = 0.713$; 30-day RRR 0.96, 95% CI 0.72–1.27; $P = 0.774$; 90-day RRR 0.94, 95% CI 0.76–1.17; $P = 0.580$, **Table 3**), demonstrating no statistically significant effect. Katragadda et al. reported 30- and 90-day complication rates separately for each cancer entity, with DID estimates presented individually per entity and all $P > 0.05$. In **Table 3**, DID estimates and corresponding confidence intervals are shown as ranges across cancer entities. Overall, none of the included studies demonstrated a statistically significant reduction in postoperative complications attributable to VBP implementation.

Readmission outcomes were reported in six studies (reflecting 5 cohorts)[41–45,47]. Robbins et al. reported a non-significant DID estimate for 90-day readmissions (95% CI −1.12 to 8.76; $P > 0.05$, **Table 3**). Herrel et al. likewise observed no significant association for 30-day readmissions, with DID estimates reported separately for each cancer entity and all $P > 0.05$. In **Table 3**, DID estimates and CIs are presented as ranges across entities. Katragadda et al. likewise observed no significant association for 30- or 90-day readmissions, with DID estimates reported separately for each cancer entity and all $P > 0.05$. In **Table 3**, DID estimates and corresponding CIs are presented as ranges across cancer entities. Worku et al., analyzing both 30- and 90-day readmissions, reported RRRs approximating 1.0 (30-day RRR 1.07, 95% CI 0.92–1.25; $P = 0.390$; 90-day RRR 1.02, 95% CI 0.92–1.13; $P = 0.688$, **Table 3**), indicating no statistically significant effect. In contrast, Offodile et al. reported a modest but statistically

significant reduction in 30-day readmissions associated with participation in Maryland's GBR model (DID −2.2; 95% CI −4.3 to −0.1; $P = 0.040$, **Table 3**). This finding, however, was not replicated across other studies or payment models.

Overall, except for a single study demonstrating a modest reduction in 30-day readmissions, none of the included analyses showed a statistically significant change in readmission rates attributable to VBP implementation.

Utilization Outcomes

Utilization outcomes reported in more than one study were confined to different measures of length-of-stay (LOS), including overall or prolonged hospital stay and SNF stay[43–45,47].

LOS outcomes were reported in four studies (reflecting three cohorts)[43–45,47] (**Supplementary Table 6**). However, different measurement formats were used across analyses. Herrel et al. evaluated prolonged LOS as a percentage outcome, Worku et al. reported total hospital LOS in days using relative risk measures, and Katragadda et al. assessed mean length of stay in SNF as measured in days.

Herrel et al. reported no significant association for prolonged length of stay, with DID estimates presented separately for each cancer entity and all $P > 0.05$. In **Supplementary Table 6**, DID estimates and corresponding CIs are shown as ranges across cancer entities. Worku et al., reporting LOS in days, observed RRs of 0.95 for the intervention group and 0.92 for the control group, resulting in a RRR of 1.04 (95% CI 0.92–1.67; $P = 0.523$, **Supplementary Table 6**), consistent with no significant effect. Katragadda et al. likewise observed no significant association for mean LOS in SNFs, with DID estimates reported separately for each cancer entity and all $P > 0.05$. In **Supplementary Table 6**, DID estimates are presented as ranges across cancer entities. Overall, despite variation in outcome definitions and reporting formats,

none of the included studies demonstrated a statistically significant change in LOS outcomes attributable to VBP participation (**Supplementary Table 6**).

DISCUSSION

In this systematic review, we evaluated the impact of VBP models on treatment expenditures and clinical outcomes in surgical oncology. Across seven studies representing six distinct cohorts, we found no consistent evidence that participation in VBP models was associated with reductions in short-term episode-based spending or improvements in short-term surgical outcomes. Qualitative synthesis of treatment expenditure data, including total episode costs, index hospitalization costs, readmission-related spending, and skilled nursing facility expenditures, revealed predominantly small and statistically non-significant DID estimates. Similarly, clinical outcomes were largely unaffected. Across studies, no consistent associations were observed between VBP participation and short-term mortality, postoperative complications, or readmissions. While one statewide capitation-based reform demonstrated a modest reduction in 30-day readmissions, this finding was not replicated in other payment models or settings. Robust conclusions regarding oncological outcomes could not be drawn, as these were reported in only a single study with serious risk of bias. Utilization outcomes such as length of stay were reported heterogeneously and did not demonstrate consistent effects. PROMs and PREMs were not reported in any study. Taken together, current evidence does not demonstrate consistent financial or short-term clinical benefits of VBP implementation in surgical oncology.

Importantly, the absence of consistent positive effects in this systematic review should be interpreted cautiously considering the methodological and conceptual limitations of the included literature. With regard to methodological aspects, first, the outcomes predominantly assessed in the included studies did not fully reflect the dimensions most central to value from the perspective of patients and treating clinicians in surgical oncology, namely survival and

functional well-being[30]. Although no reductions in costs were observed, it remains possible that improvements occurred in meaningful outcomes that were not captured by the evaluated measures. Such improvements could theoretically increase overall value, defined as outcomes achieved relative to costs. However, the included studies focus predominantly on short-term outcome measures. These outcomes, such as perioperative complications, readmissions, and early mortality, are obviously clinically relevant and meaningful. However, value in cancer surgery often unfolds over longer time horizons through sustained survival, durable disease control, and long-term functional recovery. Focusing predominantly on short-term metrics therefore captures only part of the value equation. This temporal mismatch between how value is generated and how it is measured may obscure meaningful longer-term effects that are not detectable within limited follow-up periods. These findings underscore a misalignment between the conceptual definition of value-based care, emphasizing outcomes across the full cycle of care relative to costs, and the predominantly administrative, short-term metrics used in empirical evaluations. Without broader, standardized, and patient-centered outcome measurement, that incorporates both short- and long-term dimensions, VBP models risk incentivizing what is most readily measurable rather than what is most meaningful to patients, potentially placing disproportionate emphasis on administratively captured registry metrics rather than on comprehensive clinical outcomes across the full trajectory of care. Accordingly, given that the available evidence concerning the field of surgical oncology is characterized by incomplete and non-systematic assessment of the most highly weighted endpoints, no definitive conclusions regarding the true value of these models can be drawn. Second, all included studies were observational, and most relied on quasi-experimental DID designs in the context of voluntary program participation, introducing the potential for selection bias and residual confounding (**Figure 2**). Moreover, key identifying assumptions underlying DID analyses, particularly the parallel trends assumption, were frequently insufficiently documented (**Supplementary Table 1**). Event-study analyses, formal pre-trend diagnostics, and sensitivity

analyses assessing robustness to assumption violations were rarely reported, thereby reducing confidence in the causal interpretation of the findings (**Supplementary Table 1**). Third, the number of eligible studies was small, and substantial conceptual and methodological heterogeneity across VBP models, outcome definitions, and follow-up periods precluded quantitative pooling in a formal meta-analysis. Without such quantitative pooling, it is not possible to increase precision through aggregation or to formally assess consistency and derive pooled effect estimates, limiting the certainty of the conclusions. Finally, most studies combined multiple cancer entities within a single analytic framework[41–44], whereas others reported entity-specific estimates[45–47]. Given the substantial variation in treatment algorithms, surgical complexity, and postoperative care across tumor types, the effects of VBP may differ by cancer entity, and pooled estimates may obscure such heterogeneity.

Beyond these methodological considerations, several conceptual limitations warrant attention. First, the findings seem to contrast with the expectation that aligning financial incentives toward care coordination, complication prevention, and cost containment would be particularly impactful in high-cost, high-risk settings such as cancer surgery[4–6,31]. However, an important conceptual aspect is that in surgical oncology, volume and value are not inherently opposing constructs. A robust body of evidence demonstrates a volume-outcome relationship, whereby hospitals and surgeons performing higher volumes of specific oncologic procedures often achieve better outcomes[48–51]. In such settings, greater procedural volume may itself be a mechanism through which quality, and thus value, is generated. Consequently, efforts to merely move incentives away from volume may oversimplify the relationship between production and performance in complex cancer surgery, where selective concentration of care in distinguished high-volume centers rather than indiscriminate volume reduction may be more consistent with value creation[52–55]. Second, many VBP models were introduced alongside existing activity-based hospital payment systems and service-based physician reimbursement rather than replacing them, thereby creating overlapping and potentially conflicting financial

incentives[17,56,57]. Providers therefore continue to operate within predominantly activity-based or volume-driven reimbursement structures, while value-based elements such as shared savings arrangements, quality bonuses, or bundled components are layered on and are frequently modest in magnitude. At the same time, hospitals may participate in multiple programs, such as ACOs, bundled payments, and capitation-models, that overlap and use different rules and metrics[21–25]. This combination of layered incentive structures and overlapping VBP models can create mixed or unclear financial incentives. In complex settings like surgical oncology, such ambiguity may weaken the overall impact of payment reform and limit meaningful changes in provider behavior. Third, surgical oncology is characterized by substantial clinical heterogeneity[31,58]. Outcomes are influenced by tumor biology, disease stage, procedural complexity, perioperative therapies, institutional capacity, and patient-level factors such as frailty and comorbidity[31,58]. The multifactorial determinants complicate risk adjustment and the attribution of outcomes to payment models, especially when analyses depend on administrative data lacking granular clinical detail, as in the included studies.

When placing the findings of this review in a broader context, they are consistent with prior evaluations of VBP models in surgical populations (potentially including, but not limited to cancer surgery)[59–63]. Across major surgical cohorts, bundled payment programs have demonstrated heterogeneous and procedure-specific effects, with some studies reporting modest reductions in episode spending, primarily driven by decreased post-acute care utilization, but without consistent improvements in surgical outcomes[59,61,63]. Similarly, studies of ACO participation in surgical cohorts have generally shown no meaningful changes in episode spending, mortality, complications, or readmissions[60,62]. These surgical cohort analyses closely align with the findings of the present systematic review in surgical oncology, suggesting that VBP models have thus far demonstrated limited and inconsistent effects in complex, high-risk surgical care, while facing similar methodological and conceptual limitations as outlined above for the included studies.

From a policy perspective, the limited observed effects should be interpreted cautiously and may reflect either constraints in current model design or limitations in how these models have been evaluated to date. Rather than necessarily indicating a fundamental incompatibility between VBP and surgical oncology, the findings highlight the need for more oncology-specific payment structures that align incentives with condition-specific care pathways across the cancer care continuum, incorporate robust risk adjustment for disease severity and patient complexity, and extend accountability beyond short-term postoperative windows to better capture long-term oncologic value.

In summary, current evidence neither confirms meaningful benefit nor definitively contradicts the potential of VBP in surgical oncology but instead reveals a critical need to better align payment design, outcome measurement, and methodological rigor with the clinical realities of cancer surgery before its true impact can be determined.

ACKNOWLEDGMENTS

J.M. was supported by the Ministry of Education and Research (HEROES-AYA) and intramural funding of the University Hospital of Giessen and Marburg and the Justus-Liebig-University of Giessen.

REFERENCES

1. Page B, Irving D, Amalberti R, Vincent C. Health services under pressure: a scoping review and development of a taxonomy of adaptive strategies. *BMJ Qual Saf*. 2024;33(11):738-747. doi:10.1136/bmjqs-2023-016686

2. Pichon-Riviere A, Drummond M, Palacios A, Garcia-Marti S, Augustovski F. Determining the efficiency path to universal health coverage: cost-effectiveness thresholds for 174 countries based on growth in life expectancy and health expenditures. *The Lancet Global Health*. 2023;11(6):e833-e842. doi:10.1016/S2214-109X(23)00162-6

3. Kruk ME, Pate M. The Lancet Global Health Commission on High Quality Health Systems 1 year on: progress on a global imperative. *The Lancet Global Health*. 2020;8(1):e30-e32. doi:10.1016/S2214-109X(19)30485-1

4. Porter ME, Teisberg EO. *Redefining Health Care: Creating Value-Based Competition on Results*. Harvard Business Press; 2006.

5. Porter ME. What Is Value in Health Care? *New England Journal of Medicine*. 2010;363(26):2477-2481. doi:10.1056/NEJMp1011024

6. Porter ME, Lee TH. From Volume to Value in Health Care: The Work Begins. *JAMA*. 2016;316(10):1047-1048. doi:10.1001/jama.2016.11698

7. Agarwal R, Liao JM, Gupta A, Navathe AS. The Impact Of Bundled Payment On Health Care Spending, Utilization, And Quality: A Systematic Review. *Health Affairs*. 2020;39(1):50-57. doi:10.1377/hlthaff.2019.00784

8. Joynt Maddox KE, Orav EJ, Zheng J, Epstein AM. Evaluation of Medicare's Bundled Payments Initiative for Medical Conditions. *N Engl J Med*. 2018;379(3):260-269. doi:10.1056/NEJMsa1801569

9. Shashikumar SA, Gulseren B, Berlin NL, Hollingsworth JM, Joynt Maddox KE, Ryan AM. Association of Hospital Participation in Bundled Payments for Care Improvement Advanced With Medicare Spending and Hospital Incentive Payments. *JAMA*. 2022;328(16):1616-1623. doi:10.1001/jama.2022.18529

10. de Silva Etges APB, Liu HH, Jones P, Polanczyk CA. Value-based Reimbursement as a Mechanism to Achieve Social and Financial Impact in the Healthcare System. *J Health Econ Outcomes Res*. 2023;10(2):100-103. doi:10.36469/001c.89151

11. James BC, Poulsen GP. The Case for Capitation. *Harv Bus Rev*. 2016;94(7-8):102-111, 134.

12. Berwick DM. Quality of health care. Part 5: Payment by capitation and the quality of care. *N Engl J Med*. 1996;335(16):1227-1231. doi:10.1056/NEJM199610173351611

13. Navathe AS, Emanuel EJ, Bond A, et al. Association Between the Implementation of a Population-Based Primary Care Payment System and Achievement on Quality Measures in Hawaii. *JAMA*. 2019;322(1):57-68. doi:10.1001/jama.2019.8113

14. Wagenschieber E, Blunck D. Impact of reimbursement systems on patient care – a systematic review of systematic reviews. *Health Economics Review*. 2024;14(1):22. doi:10.1186/s13561-024-00487-6

15.Elliott WB, Strand HM, Meyers FH, et al. The Hospital Capitation Payment Project: new incentives and tools for cost containment. *Inquiry*. 1983;20(2):114-120.

16.Tummalapalli SL, Estrella MM, Jannat-Khah DP, Keyhani S, Ibrahim S. Capitated versus fee-for-service reimbursement and quality of care for chronic disease: a US cross-sectional analysis. *BMC Health Serv Res*. 2022;22(1):19. doi:10.1186/s12913-021-07313-3

17.Mathes T, Pieper D, Morche J, Polus S, Jaschinski T, Eikermann M. Pay for performance for hospitals. *Cochrane Database Syst Rev*. 2019;7(7):CD011156. doi:10.1002/14651858.CD011156.pub2

18.Mendelson A, Kondo K, Damberg C, et al. The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review. *Ann Intern Med*. 2017;166(5):341-353. doi:10.7326/M16-1881

19.McWilliams JM, Landon BE, Chernew ME. Changes in health care spending and quality for Medicare beneficiaries associated with a commercial ACO contract. *JAMA*. 2013;310(8):829-836. doi:10.1001/jama.2013.276302

20.McConnell KJ, Renfro S, Chan BKS, et al. Early Performance in Medicaid Accountable Care Organizations: A Comparison of Oregon and Colorado. *JAMA Intern Med*. 2017;177(4):538-545. doi:10.1001/jamainternmed.2016.9098

21.Cattel D, Eijkenaar F. Value-Based Provider Payment Initiatives Combining Global Payments With Explicit Quality Incentives: A Systematic Review. *Med Care Res Rev*. 2020;77(6):511-537. doi:10.1177/1077558719856775

22.Navathe AS, Dinh C, Dykstra SE, Werner RM, Liao JM. Overlap between Medicare's Voluntary Bundled Payment and Accountable Care Organization Programs. *Journal of Hospital Medicine*. 2020;15(6):356-359. doi:10.12788/jhm.3288

23.Navathe AS, Liao JM, Wang E, et al. Association of Patient Outcomes With Bundled Payments Among Hospitalized Patients Attributed to Accountable Care Organizations. *JAMA Health Forum*. 2021;2(8):e212131. doi:10.1001/jamahealthforum.2021.2131

24.Leao DLL, Cremers HP, van Veghel D, Pavlova M, Groot W. The Impact of Value-Based Payment Models for Networks of Care and Transmural Care: A Systematic Literature Review. *Appl Health Econ Health Policy*. 2023;21(3):441-466. doi:10.1007/s40258-023-00790-z

25.Tobey R, Maxwell J, Turer E, et al. Health Centers and Value-Based Payment: A Framework for Health Center Payment Reform and Early Experiences in Medicaid Value-Based Payment in Seven States. *Milbank Q*. 2022;100(3):879-917. doi:10.1111/1468-0009.12580

26.Abbott DE. Addressing value in surgical oncology: Why and how. *J Surg Oncol*. 2016;114(3):263-267. doi:10.1002/jso.24229

27.Borras JM, Corral J, Aggarwal A, et al. Innovation, value and reimbursement in radiation and complex surgical oncology: Time to rethink. *Radiother Oncol*. 2022;169:114-123. doi:10.1016/j.radonc.2021.08.002

28. Sullivan R, Alatise OI, Anderson BO, et al. Global cancer surgery: delivering safe, affordable, and timely cancer surgery. *The Lancet Oncology*. 2015;16(11):1193-1224. doi:10.1016/S1470-2045(15)00223-5

29. Bateni SB, Gingrich AA, Hoch JS, Canter RJ, Bold RJ. Defining Value for Pancreatic Surgery in Early-Stage Pancreatic Cancer. *JAMA Surg*. 2019;154(10):e193019. doi:10.1001/jamasurg.2019.3019

30. Allen CJ, Smith GL, Prakash L, et al. What is "Value"? Results of a Survey of Cancer Patients and Providers. *Ann Surg Oncol*. 2022;29(11):6537-6545. doi:10.1245/s10434-022-11534-z

31. Sabet CJ, Jain B, Fayanju OM. Implementing Value-Based Health Care in Surgical Oncology: A Pathway to Improved Outcomes and Efficiency. *JCO Oncol Pract*. 2025;21(12):1730-1733. doi:10.1200/OP-25-00125

32. Cocks K, Wells JR, Johnson C, et al. Content validity of the EORTC quality of life questionnaire QLQ-C30 for use in cancer. *Eur J Cancer*. 2023;178:128-138. doi:10.1016/j.ejca.2022.10.026

33. Howell D, Molloy S, Wilkinson K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. *Annals of Oncology*. 2015;26(9):1846-1858. doi:10.1093/annonc/mdv181

34. Balitsky AK, Rayner D, Britto J, et al. Patient-Reported Outcome Measures in Cancer Care: An Updated Systematic Review and Meta-Analysis. *JAMA Netw Open*. 2024;7(8):e2424793. doi:10.1001/jamanetworkopen.2024.24793

35. Darwish M, Nandy S, Willis S, Coulson J, Withers K, Bosanquet DC. Surgical patient-reported experience measures and qualitative experience studies: systematic review. *BJS Open*. 2025;9(1):zrae142. doi:10.1093/bjsopen/zrae142

36. McLeod M, Leung K, Pramesh CS, et al. Quality indicators in surgical oncology: systematic review of measures used to compare quality across hospitals. *BJS Open*. 2024;8(2):zrae009. doi:10.1093/bjsopen/zrae009

37. Allen CJ, Thaker NG, Prakash L, et al. Communicating Value: Use of a Novel Framework in the Assessment of an Enhanced Recovery Initiative. *Annals of Surgery*. 2021;273(1):e7. doi:10.1097/SLA.0000000000004050

38. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71

39. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919. doi:10.1136/bmj.i4919

40. Roth J, Sant'Anna PHC, Bilinski A, Poe J. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*. 2023;235(2):2218-2244. doi:10.1016/j.jeconom.2023.03.008

41. Offodile AC, Lin YL, Melamed A, Rauh-Hain JA, Kinzer D, Keating NL. Association of Maryland Global Budget Revenue With Spending and Outcomes Related to Surgical Care

for Medicare Beneficiaries With Cancer. *JAMA Surg*. 2022;157(6):e220135. doi:10.1001/jamasurg.2022.0135

42. Robbins KJ, Zheng J, Orav EJ, et al. Outcomes for Gastrointestinal Oncologic Surgery Episodes in the Bundled Payments for Care Improvement Advanced Program. *Ann Surg*. Published online August 21, 2025. doi:10.1097/SLA.0000000000006915

43. Worku E, Woldesenbet S, Khalil M, Pawlik TM. Impact of Medicare Accountable Care Organizations on surgical and postoperative expenditures among patients with gastrointestinal cancer. *Surgery*. 2025;182:109211. doi:10.1016/j.surg.2025.109211

44. Worku E, Khalil M, Macedo AB, Woldesenbet S, Pawlik TM. Evaluating Accountable Care Organizations impact on gastrointestinal cancer care: are they falling short on health outcomes? *J Gastrointest Surg*. 2025;29(5):102028. doi:10.1016/j.gassur.2025.102028

45. Herrel LA, Norton EC, Hawken SR, Ye Z, Hollenbeck BK, Miller DC. Early impact of Medicare accountable care organizations on cancer surgery outcomes. *Cancer*. 2016;122(17):2739-2746. doi:10.1002/cncr.30111

46. Kuo RNC, Chung KP, Lai MS. Effect of the Pay-for-Performance Program for Breast Cancer Care in Taiwan. *J Oncol Pract*. 2011;7(3 Suppl):e8s-e15s. doi:10.1200/JOP.2011.000314

47. Katragadda C, Fung C, Yousefi-Nooraie R, et al. Medicare Accountable Care Organizations: Post-acute Care Use and Post-surgical Outcomes in Urologic Cancer Surgery. *Urology*. 2022;167:102-108. doi:10.1016/j.urology.2022.06.018

48. Gandjour A, Bannenberg A, Lauterbach KW. Threshold volumes associated with higher survival in health care: a systematic review. *Med Care*. 2003;41(10):1129-1141. doi:10.1097/01.MLR.0000088301.06323.CA

49. Yun YH, Kim YA, Min YH, et al. The influence of hospital volume and surgical treatment delay on long-term survival after cancer surgery. *Annals of Oncology*. 2012;23(10):2731-2737. doi:10.1093/annonc/mds101

50. Ihse I. The Volume-Outcome Relationship in Cancer Surgery. *Ann Surg*. 2003;238(6):777-781. doi:10.1097/01.sla.0000098616.19622.af

51. Begg CB, Cramer LD, Hoskins WJ, Brennan MF. Impact of Hospital Volume on Operative Mortality for Major Cancer Surgery. *JAMA*. 1998;280(20):1747-1751. doi:10.1001/jama.280.20.1747

52. Sakowitz S, Bakhtiyar SS, Mallick S, Yanagawa J, Benharash P. Travel to High-Volume Centers and Survival After Esophagectomy for Cancer. *JAMA Surg*. 2025;160(1):19-27. doi:10.1001/jamasurg.2024.5009

53. Rompen IF, Menso JE, Ingwersen E, et al. Impact of Merging two high-volume Centers on Patient Outcome: 1000 Consecutive Pancreatoduodenectomies. *Ann Surg*. Published online September 30, 2025. doi:10.1097/SLA.0000000000006953

54. Simon EF, Boutros CS, Khan SZ, et al. Hospital operative volume impacts surgical outcomes for patients with T4 rectal cancer following neoadjuvant chemoradiation: a

national cancer database analysis. *Surg Endosc*. 2025;39(10):6903-6914. doi:10.1007/s00464-025-12064-x

55. Salobir J, Sever P, Birk M, et al. Gastric cancer surgery in high volume university medical centers influences long-term survival. *European Journal of Surgical Oncology*. 2026;52(2). doi:10.1016/j.ejso.2025.111312

56. Berenson RA, Shartzer A, Pham HH. Beyond demonstrations: implementing a primary care hybrid payment model in Medicare. *Health Aff Sch*. 2023;1(2):qxad024. doi:10.1093/haschl/qxad024

57. Conrad DA. The Theory of Value-Based Payment Incentives and Their Application to Health Care. *Health Serv Res*. 2015;50 Suppl 2(Suppl 2):2057-2089. doi:10.1111/1475-6773.12408

58. Nizam W, Yeo HL, Obeng-Gyasi S, Brock MV, Johnston FM. Disparities in Surgical Oncology: Management of Advanced Cancer. *Ann Surg Oncol*. 2021;28(13):8056-8073. doi:10.1245/s10434-021-10275-9

59. Mullens CL, Schwartzman D, Savitch SL, et al. Bundled Payment Programs and Changes in Practice Patterns and Episode Spending in Major Gastrointestinal Surgery. *Health Serv Res*. 2026;61(1):e70046. doi:10.1111/1475-6773.70046

60. Nathan H, Thumma JR, Ryan AM, Dimick JB. Early Impact of Medicare Accountable Care Organizations on Inpatient Surgical Spending. *Ann Surg*. 2019;269(2):191-196. doi:10.1097/SLA.0000000000002819

61. Goldstone RN, Zhang J, Stafford C, et al. Impact of Bundled Payment Care Improvement Initiative on Health Care Expenditure in Major Bowel Procedures. *Dis Colon Rectum*. 2022;65(6):851-859. doi:10.1097/DCR.0000000000002211

62. Cole AP, Krasnova A, Ramaswamy A, et al. Prostate cancer in the medicare shared savings program: are Accountable Care Organizations associated with reduced expenditures for men with prostate cancer? *Prostate Cancer Prostatic Dis*. 2019;22(4):593-599. doi:10.1038/s41391-019-0138-1

63. Chopra Z, Gulseren B, Chhabra KR, Dimick JB, Ryan AM. Bundled Payments for Care Improvement Efficacy Across 3 Common Operations. *Ann Surg*. 2023;277(1):e16-e23. doi:10.1097/SLA.0000000000004869

FIGURES AND TABLES

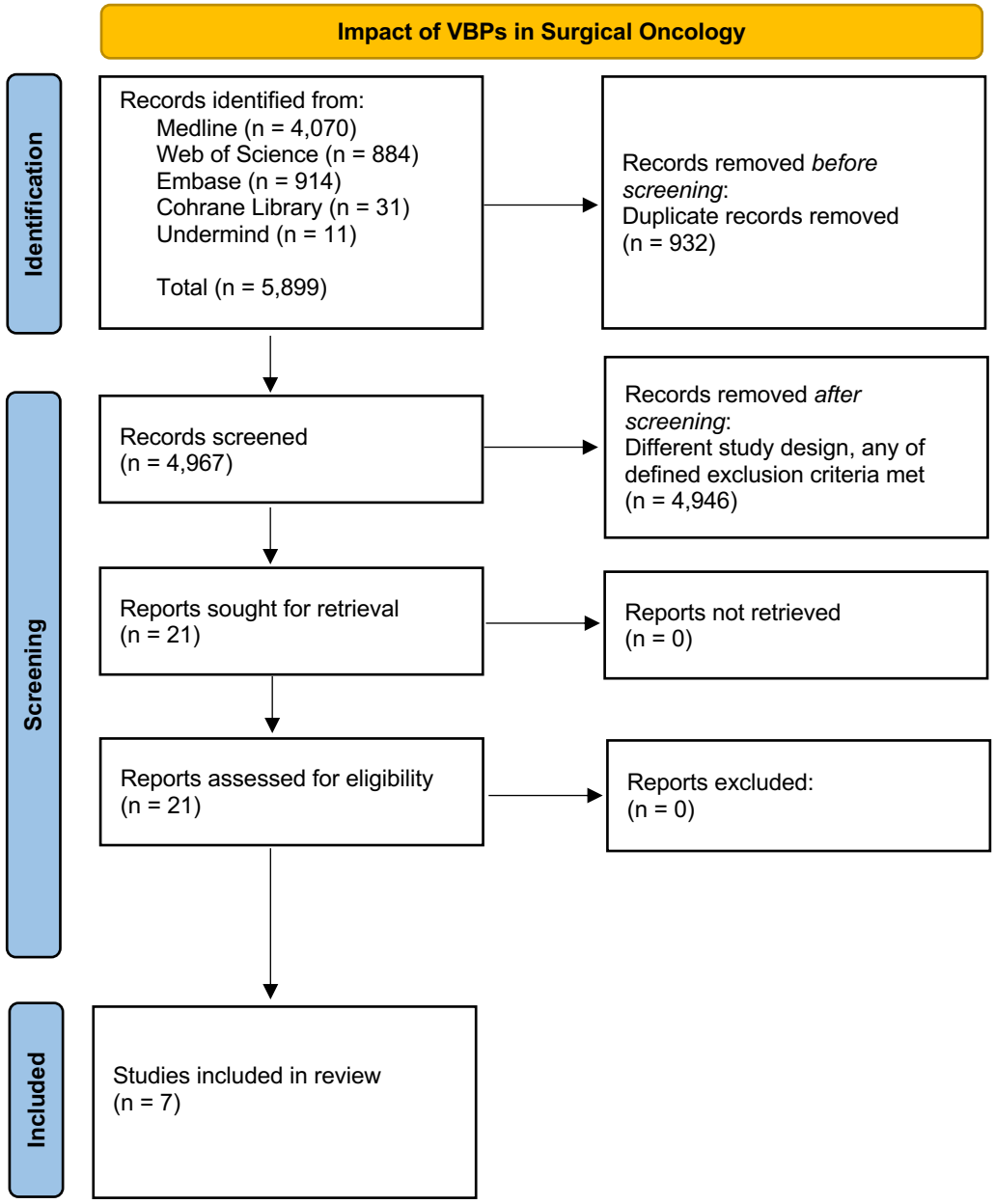


Figure 1: PRISMA flow-chart.

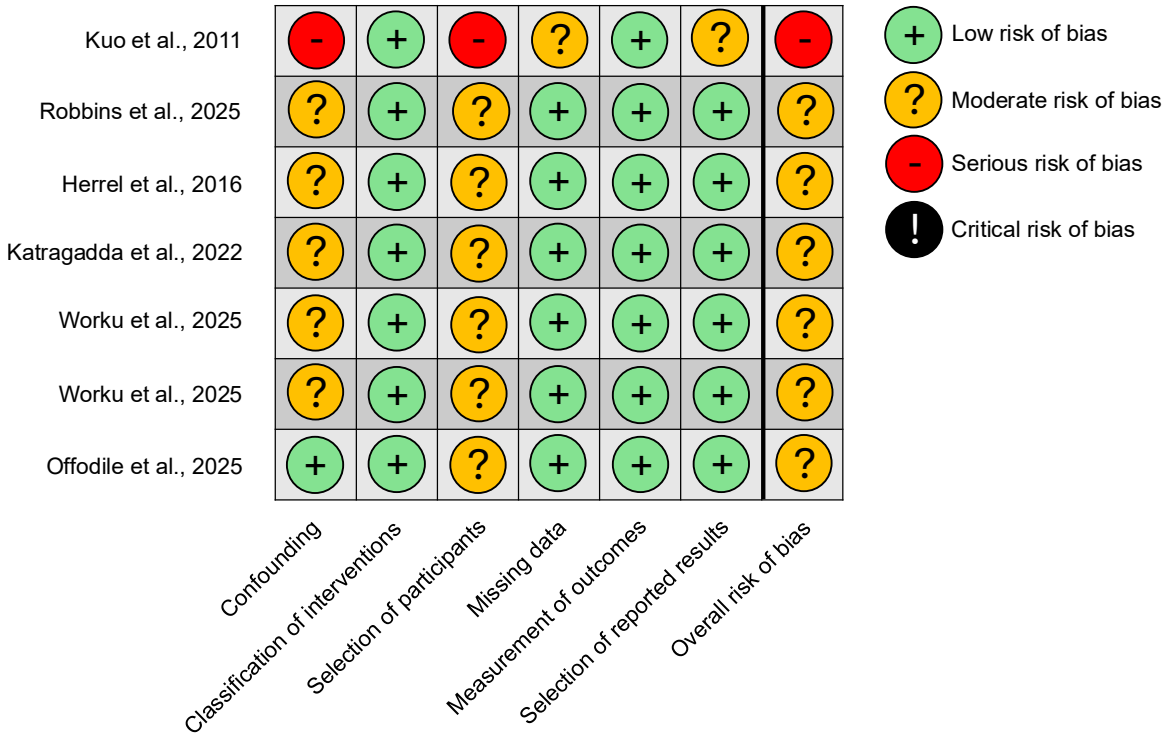


Figure 2: Risk of bias assessment. Assessment according to the Cochrane ROBINS-I tool (v2).

Authors	Country, Time period	VBP model	IG (VBP)	CG (N-VBP)	No pat. VBP	No pat. N-VBP	No hosp. VBP	No hosp. N-VBP	PSM	Cancer entities
Kuo et al. 2011 J. Oncol. Practice	Taiwan, 2003-2004	P4P	BC-P4P enrollment	No BC-P4P enrollment	1,393	3,135	947	2,700	No	Breast
Robbins et al. 2025 Ann. Surg.	USA, 2017-2019	Bundled payment	BPCI-A participation	No BPCI-A participation	3,339	1,884	86	86	Yes	Colorectal
Herrel et al. 2016 Cancer	USA, 2011-2013	ACO	ACO participation	No ACO participation	19,439	365,080	106	2,561	No	Colorectal Prostate Lung Kidney Ovarian Bladder Pancreas Esophagus Liver
Katragadda et al. 2022 Urology	USA, 2011-2017	ACO	ACO participation	No ACO participation	98,483	236,031	524	2,066	No	Prostate Bladder Kidney
Worku et al. 2025 J. Gastrointest. Surg. **Worku et al.** 2025 Surgery	USA, 2016-2020	ACO	ACO participation	No ACO participation	7,723	15,634	57	171	Yes	Biliary tract Liver Pancreas Colorectal

Offodile et al. 2022 JAMA Surg.	USA, 2011-2018	Capitation	GBR hospitals Maryland	Control hospitals in control states	20,320	4,737	35	101	Yes	Breast Colorectal Lung Prostate Bladder Female-Genitourin. Kidney Other

Table 1: Overview on included studies. ACO = Accountable Care Organization, BC-P4P = Breast-Cancer-Pay-for-Performance, BPCI-A = Bundles Payments for Care Improvement – Advanced, CG = Control Group, GBR = Global Budget Revenue, IG = Intervention Group, NA = Not Available, N-VBP = Non-Value-Based Payment, P4P = Pay-for-performance, PSM = Propensity Score Matching, VBP = Value-Based Payment.

	Time	IG (VBP)			CG (N-VBP)			DID		
Study	**Observed**	**Pre ($)**	**Post ($)**	**Diff. ($)**	**Pre ($)**	**Post ($)**	**Diff. ($)**	**DID ($)**	**DID ($) 95%CI**	***P*-value**
Robbins et al.	90 day episode	32,687	31,929	-758	34,006	32,932	-1,075	316	-2,076; 2,709	> 0.05
Worku et al. J. Surg. Oncol. Surgery	30 day episode	32,335	31,699	-636	33,210	32,820	-390	-246	-1,497; 1,005	0.69
Offodile et al.	30 day episode	31,720	28,312	NA	20,676	16,583	NA	646	-208; 1,542	0.14

Table 2: Short-term treatment expenditures. CG = Control Group, CI = Confidence Interval, Diff. = Difference, DID= Difference-In-Difference, IG = Intervention Group, NA = Not available, N-VBP = Non-Value-Based Payment, VBP = Value-Based Payment.

Mortality (%)										
	Time	**IG (VBP)**			**CG (N-VBP)**			**DID**		
	Period	**Pre**	**Post**	**Diff.**	**Pre**	**Post**	**Diff.**	**DID**	**95%CI**	***P*-value**
Robbins et al.	90d	5	5.3	0.31	4.4	5.9	1.53	-1.22	-3.60; 1.17	$P > 0.05$
Herrel et al.	30d	3.4	3.2	NA	3.4	3.3	NA	-2.06; 2.39 (range)	-4.75; 7.67 (range)	All $P > 0.05$
Worku et al. J. Surg. Oncol. Surgery	30d	6.67	6.89	1.01 (RR)	5.44	5.76	1.06 (RR)	0.96 (RRR)	0.77;1.21	0.745
	90d	9.75	11.05	1.13 (RR)	8.44	9.31	1.10 (RR)	1.03 (RRR)	0.85;1.24	0.777
Offodile et al.	30d	2.5	1.6	NA	2.1	1.3	NA	-0.4	-1.4; 0.5	0.360
Complications (%)										
	Time	**IG (VBP)**			**CG (N-VBP)**			**DID**		
	Period	**Pre**	**Post**	**Diff.**	**Pre**	**Post**	**Diff.**	**DID**	**95%CI**	***P*-value**
Herrel et al.	30d	44	42.5	NA	43.8	42.7	NA	-13.30; 0.45 (range)	-26.33; 5.56 (range)	All $P > 0.05$
Worku et al. J. Surg. Oncol. Surgery	Index	21.26	19.66	0.92 (RR)	19.83	17.88	0.90 (RR)	1.03 (RRR)	0.90;1.18	0.713
	30d	4.00	4.2	1.04 (RR)	3.61	3.94	1.09	0.96 (RRR)	0.72;1.27	0.774
	90d	6.71	11.05	0.98 (RR)	5.85	6.01	1.04	0.94 (RRR)	0.76; 1.17	0.580

Katragadda et al.	30d	22.8; 38.1 (range)	22.0; 36.9 (range)	NA	22.7; 37.1 (range)	22.4; 36.0 (range)	NA	-0.47; -0.13 (range)	NA	All $P > 0.05$
	90d	34.2; 43.1 (range)	30.0; 40.9 (range)	NA	35.3; 43.6 (range)	30.6; 40.9 (range)	NA	-0.60; 0.48 (range)	NA	All $P > 0.05$
Readmissions (%)										
	Time	**IG (VBP)**			**CG (N-VBP)**			**DID**		
	Period	**Pre**	**Post**	**Diff.**	**Pre**	**Post**	**Diff.**	**DID**	**95%CI**	***P*-value**
Robbins et al.	90d	21.00	22.00	1.03	21.10	18.40	-2.78	3.82	-1.12; 8.76	$P > 0.05$
Herrel et al.	30d	12.60	12.00	NA	12.50	12.10	NA	-8.67; 3.8 (range)	-21.88; 18.78 (range)	All $P > 0.05$
Katragadda et al.	30d	4.7; 24.5 (range)	3.7; 20.7 (range)	NA	5.0; 25.4 (range)	3.7; 21.3 (range)	NA	-0.38; 0.30 (range)	NA	All $P > 0.05$
	90d	10.6; 30.8 (range)	9.1;26.7 (range)	NA	10.7; 31.2 (range)	9.4; 27.3 (range)	NA	-0.23; -0.16 (range)	NA	All $P > 0.05$
Worku et al. J. Surg. Oncol. Surgery	30d	14.84	16.13	1.09 (RR)	15.64	15.42	1.01 (RR)	1.07 (RRR)	0.92;1.25	0.390
	90d	25.79	25.77	1.00 (RR)	25.68	25.23	0.98 (RR)	1.02 (RRR)	0.92; 1.13	0.688
Offodile et al.	30d	13.5	10.4	NA	8.9	7.60	NA	-2.2	-4.3; -0.1	**0.040**

Table 3: Clinical outcomes. CG = Control Group, CI = Confidence Interval, Diff. = Difference, DID= Difference-In-Difference, IG = Intervention Group, NA = Not available, N-VBP = Non-Value-Based Payment, VBP = Value-Based Payment.

Authors	**Common treatment timing**	Staggered timing addressed w/ heterogeneity-robust estimator	**PT assumption justified**	Event-study implemented	Formal pre-trend test reported	Sensitivity to PT violations reported	**Clustering-robust inference appropriate**
Kuo et al.	Not applicable (no DID)	Not applicable (no DID)	Not applicable (no DID)	Not applicable (no DID)	Not applicable (no DID)	Not applicable (no DID)	Not applicable (no DID)
Robbins et al.	Yes	Not necessary (common timing)	Yes (formally tested; some violations)	No	Yes	No	Partially (GEE; clustering level not fully specified)
Herrel et al.	No	No	No	No	No	No	Yes (hospital-level clustering)
Katragadda et al.	Yes	Not necessary (common timing)	No	No	No	No	Yes (hospital-level clustering)
Worku et al. J. Gastrointest. Surg. **Worku et al.** Surgery	No	No	No	No	No	No	Yes (hospital-level clustering)
Offodile et al.	Yes	Not necessary (common timing)	Yes (formal event-study and joint tests)	Yes	Yes (likelihood ratio tests)	No	Yes (hospital fixed effects and hospital-level clustering)

Supplementary Table 1: DID checklist for included studies. DID= Difference-In-Difference, PT = Parallel Trends.

	Age Mean (SD)			**Gender** Female, %			**Multimorbidity / Comorbidity**		
	IG (VBP)	**CG (N-VBP)**	***P*-value**	**IG (VBP)**	**CG (N-VBP)**	***P*-value**	**IG (VBP)**	**CG (N-VBP)**	***P*-value**
Kuo et al.	50.0 (11.4) (Mean (SD))	50.8 (11.6) (Mean (SD))	**0.022**	100	100	1	0.185 (0.412) (Mean comorbidities (SD))	0.225 (0.448) (Mean comorbidities (SD))	**0.004**
Robbins et al.	<65: 5.2% 65-80: 58.9% >80: 36.0%	<65: 6.3%; 65-80: 55.0% >80: 38.7%	NA	53	53.7	NA	33.7 (%)	33.5 (%)	NA
Herrel et al.	74.30 (Mean)	74.44 (Mean)	NA	42.23	41.99	0.42	NA	NA	NA
Katragadda et al.	<65: 10.1% 65-69: 33.05% 70-74: 25.25% 75-79: 14.9% 80-84: 9.5% >85: 7.2%	<65: 10.5% 65-69: 31.7% 70-74: 24.75% 75-79: 15.4% 80-84: 10% >85: 7.6%	**< 0.001**	14.15	13.4	**< 0.001**	0: 4.75% 1-3: 65.85% 4-6: 26.8% >=7: 2.7%	0: 5.05% 1-3: 66.3% 4-6: 26.05% >=7: 2.6%	**< 0.001**
Worku et al. J. Surg. Oncol. Surgery	76 (71-81) Median (IQR)	76 (71-82) Median (IQR)	0.07	52.5	52.1	0.5	60.1% (CCI >2)	60.4% (CCI >2)	0.7
Offodile et al.	74.9 (6.5) (Mean (SD))	74.9 (6.5) (Mean (SD))	NA	69.2	70.1	NA	NA	NA	NA

Supplementary Table 2: Patient characteristics of included studies. CG = Control Group, CCI = Charlson Comorbidity Index, DiD= Difference-In-Difference, IG = Intervention Group, IQR = Interquartile Range, NA = Not available, N-VBP = Non-Value-Based Payment, SD = Standard Deviation, VBP = Value-Based Payment.

	Hospital profit status, %			**Hospital size**			**Hospital volume**			**Teaching hospital**, %		
	IG (VBP)	**CG (N-VBP)**	***P*-value**	**IG (VBP)**	**CG (N-VBP)**	***P*-value**	**IG (VBP)**	**CG (N-VBP)**	***P*-value**	**IG (VBP)**	**CG (N-VBP)**	***P*-value**
Kuo et al.	NA	NA	NA	NA	NA	NA	NA	NA	NA	22.9	64.1	**<0.001**
Robbins et al.	For-profit: 22.00% Non-for-profit: 67.00% Public: 11.00%	For-profit: 23.30% Non-for-profit: 66.30% Public: 10.50%	NA	Small: 6.40% Medium: 65.50% Large: 27.30%	Small: 25.60% Medium: 51.20% Large: 18.60%	NA	Lowest qu.: 0.00% 2nd qu.: 21.80% 3rd qu.: 33.60% Highest qu.: 44.60% Rural: 0.90%	Lowest qu.: 26.7% 2nd qu.: 31.40% 3rd qu.: 17.40% Highest qu.: 24.40% Rural: 4.7%	NA	55.5	50.0	NA
Herrel et al.	For-profit: 4.70% Non-for-profit: 86.80% Public: 8.50%	For-profit: 17.60% Non-for-profit: 69.00% Public: 13.40%	**< 0.001**	<200: 42.5% 200-349: 24.5% 350-499: 19.8% ≥500 13.2%	<200: 52.5% 200-349: 27.0% 350-499: 11.3% ≥500: 9.3%	**0.02**	31.3 (10.7-90.3) (Median ann. oncoligcal surgical volume (IQR))	23.3 (9-58.7) (Median ann. oncoligcal surgical volume (IQR))	0.06	48.1	34.1	**0.002**
Katragadda et al.	For-profit: 5.50% Non-for-profit: 87.60% Public: 6.90%	For-profit: 24.10% Non-for-profit: 61.70% Public: 14.20%	**< 0.001**	<100: 10.9% 100-199: 31.1% 200-500: 44.3% >500: 13.7%	<100: 27.4% 100-199: 33.0% 200-500: 33.0% >500: 6.7%	**< 0.001**	NA	NA	NA	52.3	35.8	**< 0.001**

Worku et al. J. Surg. Oncol. Surgery	NA	NA		≤300: 31.6% >300: 68.4%	≤300: 31.6% >300: 68.4%	1.0	Low: 17.5% Medium: 61.4% High: 21.1%	Low: 17.5% Medium: 61.4% High: 21.1%	1.0	70.2	70.2	1.0
Offodile et al.	NA	NA	NA	5.46 (mean, ln)	5.42 (mean, ln)	NA	Inpatient: 8.13 Outpatient: 8.81 (Mean ln)	Inpatient: 8.12 Outpatient: 8.80 (Mean ln)	NA	42.9	40.6	NA

Supplementary Table 3: Hospital characteristics of included studies. CG = Control Group, IG = Intervention Group, IQR = Interquartile Range, ln = natural logarithm, NA = Not Available, N-VBP = Non-Value-Based Payment, VBP = Value-Based Payment.

	Time	IG (VBP)			CG (N-VBP)			DID		
Study	**Observed**	**Pre ($)**	**Post ($)**	**Diff. ($)**	**Pre ($)**	**Post ($)**	**Diff. ($)**	**DID ($)**	**DID ($) 95%CI**	***P*-value**
Robbins et al.	Index	18,353	17,941	-412	18,545	18,563	18	-430	-1,132; 272	> 0.05
	90d readmission	2,738	3,049	311	3,111	2,665	-446	757	-274; 1,787	> 0.05
Worku et al. J. Surg. Oncol. Surgery	Index + 30d readmission	29,321	28,920	-401	30,211	29,934	-277	-123	-1,280; 1,035	0.83

Supplementary Table 4: Treatment expenditures for index treatment and readmission. CG = Control Group, CI = Confidence Interval, Diff. = Difference, DID= Difference-In-Difference, IG = Intervention Group, NA = Not available, N-VBP = Non-Value-Based Payment, VBP = Value-Based Payment.

	Time	IG (VBP)			CG (N-VBP)			DID		
Study	**Observed**	**Pre ($)**	**Post ($)**	**Diff. ($)**	**Pre ($)**	**Post ($)**	**Diff. ($)**	**DID ($)**	**DID ($) 95%CI**	***P*-value**
Robbins et al.	90d	2,888	2,764	-124	3,501	3,581	80	-204	-1,210; 802	> 0.05
Worku et al. J. Surg. Oncol. Surgery	30d	8,599	7,938	-661	8,353	8,140	-213	-447	-1,335; 474	0.32
Katragadda et al.	NA	5,969; 8776	5,350; 7,455	NA	6,073; 8,860	5,414; 7496	NA	-92; 42.6	NA	All *P* > 0.05

Supplementary Table 5: Treatment expenditures for skilled nursing facility (SNF) treatment. CG = Control Group, CI = Confidence Interval, Diff. = Difference, DID= Difference-In-Difference, IG = Intervention Group, NA = Not available, N-VBP = Non-Value-Based Payment, VBP = Value-Based Payment.

	Specification	IG (VBP)			CG (N-VBP)			DID		
Study		Pre	Post	Diff.	Pre	Post	Diff.	DID	DID 95%CI	*P*-value
Herrel et al.	Prolonged length of stay (%)	10.0	8.9	NA	10.0	8.4	NA	-0.90; 2.73	-19.12 ; 19.3	> 0.05
Worku et al. J. Surg. Oncol. Surgery	Length of stay (days)	23.58	22.47	0.95 (RR)	19.71	18.08	0.92 (RR)	1.04 (RRR)	0.92; 1.67	0.523
Katragadda et al.	Length of SNF stay (mean, days)	21.4; 23.8	17.0; 20.9	NA	22.5; 23.7	18.3; 21.8	NA	-1.87; -0.25	NA	All *P* > 0.05

Supplementary Table 6: Length of stay measures. CG = Control Group, CI = Confidence Interval, Diff. = Difference, DID= Difference-In-Difference, IG = Intervention Group, NA = Not available, N-VBP = Non-Value-Based Payment, VBP = Value-Based Payment.