LEWINGROUP®

CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 7 Evaluation & Monitoring Annual Report

Final

Prepared for:

CMS

Submitted by:

The Lewin Group, Inc.

With our partners:

Abt Associates, GDIT, and Telligen

March 2021

CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 7 Evaluation & Monitoring Annual Report

The Lewin Group
With our partners Abt Associates, GDIT, and Telligen

Authors:

Grecia Marrufo, Aylin Bradley, Julie Somers, Laura Dummit, Jaclyn Marshall, Kyi-Sin Than, Ayah Fannoun, Anna Braendle, Rebecca Braun, Andrea Chung, Brady Durst, Gina Gerding, Sehreen Khan, Sharon Kim, Daniel Strubler, Brittani Thomas, Peter Weidner, Christine LaRocca, Colleen Kummet

Lewin's address:

3160 Fairview Park Dr., Suite 600, Falls Church, VA 22042

Federal Project Officer:

Daver Kahvecioglu
Division of Data, Research, and Analytic Methods (DRAM)
Research and Rapid Cycle Evaluation Group (RREG),
Center for Medicare and Medicaid Innovation (CMMI),
Centers for Medicare and Medicaid Services (CMS)
This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2014-00033I Task Order HHSM-500-T0009.

Table of Contents

List of Appendices

Executive Summary

A. Introduction

The Bundled Payments for Care Improvement (BPCI) initiative was designed to test whether linking payments for all providers that furnish Medicare-covered items and services during an episode of care related to an inpatient hospitalization can reduce Medicare expenditures while maintaining or improving quality of care. The BPCI initiative was comprised of four Models; this evaluation contract covers Models 2, 3, and 4. This seventh and final summative evaluation describes the experience under BPCI over the entire five-year performance period of the initiative, from October 2013 through September 2018.

BPCI was a voluntary initiative that allowed participants to choose among several key design options. BPCI Awardees, which could be hospitals, physician group practices (PGPs), post-acute care (PAC) providers, or other entities that convene health care organizations, entered into voluntary agreements with the Centers for Medicare & Medicaid Services (CMS) to be held accountable for total episode payments. The agreements also specified participants' choices among three payment Models, 48 clinical episodes, three episode lengths, and three risk tracks. Awardees submitted implementation protocols that specified care redesign interventions and if they would use available Medicare payment policy waivers, beneficiary engagement incentives, or financial arrangements that could be protected under specific waivers of fraud and abuse laws. Allowing Awardees to choose among key design options implicitly recognized the variability across health care markets, providers, and episodes of care. The resulting diversity in responses and impacts provides CMS with information on the approaches that show the most promise in achieving payment reductions while maintaining or improving quality.

This final report includes Medicare claims-based impact estimates for key outcomes for Models 2, 3, and 4, as well as estimated savings to the Medicare program over the entire five years of the initiative. Overall, our results are consistent with previous reports that demonstrated the initiative resulted in reductions to episode payments, which were driven by declines in institutional PAC use and payments, while maintaining quality of care. The report also includes an updated analysis of the estimate of net savings to the Medicare program. Although BPCI resulted in reductions in Medicare fee-for-service (FFS) payments, it did not result in net savings, i.e., overall savings to Medicare after taking into account reconciliation payments made to participants, a result consistent with previous reports. Finally, the report includes an analysis comparing Medicare FFS spending reductions that are calculated with the use of model benchmarks to those that are estimated by the evaluation. Medicare FFS spending reductions calculated with the use of model benchmarks were based on a national trend while spending reductions that are estimated by the evaluation were based on a retrospective comparison group trend. We found that Medicare FFS spending reductions calculated with the use of model benchmarks were larger than spending reductions estimated by the evaluation. Because reconciliation payments are based on model benchmarks, this finding provides insights into why the Medicare program did not achieve savings overall and may have implications for the design of future models. See Exhibit C.13 in **Appendix C** for a description of the measures used in the comparison of Medicare FFS spending reductions that are calculated with the use of model benchmarks to spending reductions that are estimated by the evaluation.

B. Results

Exhibit ES-1 summarizes participation rates, impact estimates, and estimates of savings to Medicare during the five-year initiative. Results in this final report are consistent with those of previous reports, which showed BPCI resulted in reductions to episode payments, although Medicare continued to experience net financial losses. Reductions in episode payments were driven by declines in institutional PAC use.[1]

[1] Previous reports are available for download at https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

Exhibit ES-1: Impact of BPCI on Key Claims-based Outcomes and Medicare Savings, Models 2-4, Q4 2013 – Q3 2018

METHODS: IMPACT OF BPCI

To estimate the impact of BPCI, the evaluation used a difference-in-differences (DiD) design, which calculates the differential change in outcomes between the baseline and intervention period for beneficiaries who received services from BPCI providers, relative to beneficiaries who received services from a comparison group of non-BPCI providers.

Time Period

Baseline: October 2011 - September 2012
Intervention: October 2013 - September 2018

Outcomes

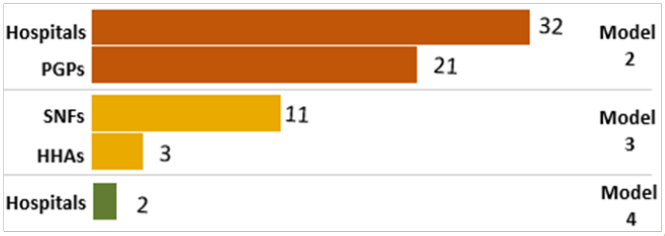
We estimated the impact on 17 claims- based outcomes related to payments, utilization, and quality

Comparison Group

Designed to be similar to BPCI participants with respect to baseline characteristics that could affect participation decisions and performance under BPCI, such as market-level characteristics and provider attributes

Clinical Episodes Evaluated

We constructed comparison groups and evaluated the impact of BPCI for differing numbers of clinical episodes by Model and participant type. We present results by Model and participant type overall, and by surgical and medical clinical episodes for Model 2 hospitals and PGPs.



METHODS: MEDICARE SAVINGS

Net Medicare savings were estimated for episodes initiated under Models 2 and 3, which together represent 99% of all BPCI episodes. Net Medicare savings are calculated by subtracting reconciliation payments made to or received from BPCI participants from the change in aggregate non-standardized paid amounts (excluding patient copays and deductibles). The change in aggregate non-standardized paid amounts is calculated by multiplying the DiD estimates by a standardized to non-standardized conversion factor.

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

We present the estimates of net savings to Medicare in two ways:

- As **BPCI was implemented**, in which CMS temporarily eliminated repayment responsibility (downside risk)
- As **BPCI was intended**, the hypothetical scenario in which CMS had not retrospectively eliminated repayment responsibility

SPENDING REDUCTIONS: MODEL VS EVALUATION

To understand why Medicare realized net losses even though the model was intended to achieve savings of 2% to 3%, we compared Medicare FFS spending reductions that are calculated with the use of model benchmarks to those estimated by the evaluation.

Spending Reductions (Model Benchmarks) = Benchmark price – Average FFS episode payments

Spending Reductions (Evaluation) = Estimate of payments absent BPCI – Average FFS episode payments

PARTICIPATION

Voluntary Participation

Participation during the five-year BPCI initiative varied by Model:

- **423** hospitals and **272** PGPs in **Model 2,**
- **873** SNFs and **117** HHAs in **Model 3,**
- **23** hospitals in **Model 4.**

Withdrawal Rates

But by the end of the initiative, some participants had withdrawn:

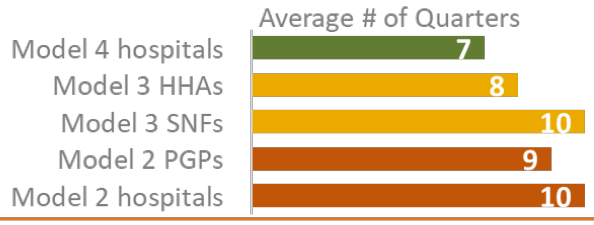
- **41%** of Model 2 hospitals
- **44%** of Model 2 PGPs
- **45%** of Model 3 SNFs
- **63%** of Model 3 HHAs
- **91%** of Model 4 hospitals – all but two

BPCI Episodes

Of the 1.4 million BPCI episodes initiated across all Models,

- **Model 2** hospitals and PGPs accounted for **88%,**
- **Model 3** SNFs and HHAs accounted for **8%,** and
- **Model 4** hospitals accounted for **1%.**

Participants could join BPCI over an extended period and stop participating at any time. The average length of participation varied by Model and participant type.



PAYMENTS

Model 2:

The total standardized allowed payment amount for the inpatient stay and 90 days post discharge declined **3.1% ($836)** for Model 2 hospital-initiated episodes and **4.9% ($1,260)** for Model 2 PGP-initiated episodes, relative to comparison group episodes. The declines were primarily due to relative reductions in SNF payments, which declined **13.5% ($700)** for hospital episodes and **14.2% ($693)** for PGP episodes.

- Comparing surgical and medical episodes under Model 2 hospitals, there were no statistically significant differences between payment results for the two types of episodes.
- For Model 2 PGP episodes, surgical episodes had larger relative declines than medical episodes in total payments, as well as in SNF, IRF, and HHA payments.

Model 3:

The total standardized allowed payment amount from episode start through 90 days declined by **7.6% ($2,171)** for Model 3 SNF-initiated episodes and by **7.5% ($981)** for Model 3 HHA-initiated episodes, relative to the comparison group. The decline in payments for SNF-initiated episodes was primarily due to the reduction in SNF payments, which declined **13.2% ($2,387)**. The decline in payments for HHA-initiated episodes was due to declines in Part A, as well as Part B, payments.

Model 4:

For Model 4, the episode includes the inpatient stay and readmissions 30 days post discharge, but not PAC services. Consistent with incentives, payments for readmissions within 30 days declined by **29.1% ($202)** for Model 4 hospital-initiated episodes relative to the comparison group, but the impact dissipated during the longer 90-day PDP. Payments for PAC either increased or were flat, and there was no statistically significant change in total payments.

UTILIZATION

Model 2: Among patients discharged to any PAC setting, the proportion discharged to an institutional PAC facility decreased by **3.6 percentage points** for Model 2 hospital-initiated episodes, while there was no statistically significant change for Model 2 PGP-initiated episodes relative to the comparison group. Of patients who had a SNF stay, the average number of days in a SNF was **2.9 days** shorter for hospital episodes and **2.3 days** shorter for PGP episodes.

Model 3: The average number of days in a SNF was **3.5 days** shorter for SNF-initiated episodes, relative to the comparison group. Among HHA-initiated episodes, there was no statistically significant change in the number of HH visits, relative to the comparison group.

Model 4: Unlike Model 2, there was no statistically significant change in the proportion of beneficiaries discharged to institutional PAC. However, the proportion of beneficiaries discharged to any PAC setting increased by **7.1 percentage points**, indicating an increase in the proportion of beneficiaries discharged to HHA and/or discharged home with no HHA services.

QUALITY

Quality of care, as measured by mortality, emergency department visits, and unplanned readmissions, generally did not change under BPCI.

MEDICARE SAVINGS

BPCI as Implemented

After accounting for reconciliation payments through the end of the initiative,

- **Model 2** resulted in an estimated **loss to Medicare of 1.3%** of what payments would have been absent BPCI ($418 million, $p<0.05$), ranging from 0.4% to 2.1% ($133 to $702 million).
- **Model 3** resulted in a **loss to Medicare of 3.1%** ($110 million, $p<0.05$), ranging from 1.6% to 4.7% ($55 to $165 million).

BPCI as Intended

If downside risk had not been eliminated and participants had been required to repay funds when their payments were above the target,

- Medicare could have realized **no change in net spending for Model 2** (0.0% in net spending, or $2 million), ranging from a loss of 0.9% ($283 million) to a savings of 0.9% ($286 million).
- The estimated **net loss to Medicare under Model 3 could have been 1.9%** ($66 million, $p<0.05$), ranging from 0.3% to 3.4% ($11 to $120 million).

For episodes initiated by both types of participants under Models 2 and 3, Medicare FFS spending reductions that are calculated with the use of model benchmarks are greater than spending reductions estimated by the evaluation. The differences are statistically significant for Model 2 hospital, Model 3 SNF, and Model 3 HHA episodes.

Reduction in Spending

	Model	Evaluation (95% CI)
Hospitals	6.0%	3.3% (2.3% to 4.3%)
PGPs	5.5%	4.7% (3.0% to 6.5%)
SNFs	11.9%	7.6% (5.7% to 9.5%)
HHAs	10.2%	5.8% (1.6% to 10.1%)

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers and CMS data on reconciliation payments.

C. Discussion and Conclusions

This report provides updated estimates of the impact of BPCI on select outcomes under Models 2, 3, and 4, as well as estimates of net savings to the Medicare program through the end of the BPCI initiative. Consistent with previous reports, BPCI participants responded to the initiative's incentives by reducing Medicare FFS payments. We continue to see this result from general patterns of reduced utilization and intensity of PAC use, with reductions in institutional care and decreases in the number of SNF days among patients who receive SNF care. This decline did not translate into overall savings to Medicare after taking into account reconciliation payments made to participants. There are few indications in claims-based results that BPCI affected quality of care. However, three out of the 11 Model 3 SNF clinical episodes analyzed had an increase in readmissions, ED use, or mortality, potentially indicating a decrease in quality. However, these three clinical episodes suffer from small sample sizes, both overall and per SNF, making it difficult to ascertain whether these results are a signal or simply noise.

This final report shows differences in findings between surgical and medical clinical episodes.[2] Under Model 2 hospitals, relative declines in total payments and payments for SNF, inpatient rehabilitation facility (IRF), and readmissions were similar for surgical and medical clinical episodes. Under Model 2 PGPs, relative declines in total payments, as well as SNF, IRF and HHA payments, were larger for surgical clinical episodes than medical clinical episodes. With respect to the effect of BPCI on quality of care as measured by readmissions, ED use, and mortality, there were no changes for Model 2 hospital-initiated episodes, and while there were a few indications of change for Model 2 PGP-initiated episodes, the direction was not consistent.

Results in the Year 5 Evaluation & Monitoring Annual Report also indicated that changes in functional status did not differ between beneficiaries in BPCI episodes and comparison beneficiaries, based on survey results, although fewer BPCI beneficiary respondents reported the highest level of satisfaction with their care. Quality of care was also maintained among vulnerable populations, including populations dually eligible for Medicare and Medicaid, with dementia, or with recent institutional PAC use.[3]

These findings suggest that holding providers financially accountable for episodes of care may successfully reduce payments without compromising quality of care. However, although per-episode payments declined under BPCI, net Medicare spending increased because the reconciliation payments made to participants were greater than the decrease in total payments. The final estimated net Medicare losses reported in this report for Model 2 and Model 3 episodes initiated during the five-year initiative were roughly similar to the estimated net losses for episodes initiated through the first four years, as reported in the Year 6 Evaluation and Monitoring Annual Report. However, based on the final estimates in this report, Model 2 and Model 3 would have resulted in net losses even under the hypothetical scenario in which downside risk was not temporarily eliminated; whereas, in the earlier report, Model 2 would have achieved a small, albeit not statistically significant, amount of net savings under the hypothetical scenario.

[2] See Exhibits C.7a through C.7e in **Appendix C** for the distribution of episodes by individual clinical episodes and by surgical and medical episode category.

[3] The report is available for download from: https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

I. Introduction

The Centers for Medicare & Medicaid Services (CMS), through the Center for Medicare & Medicaid Innovation (CMMI), implemented the Bundled Payments for Care Improvement (BPCI) initiative from October 2013 through September 2018, under the authority of section 1115A of the Social Security Act. The BPCI initiative, which was comprised of four Models, was designed to test whether linking provider payments for an episode of care could reduce Medicare payments while maintaining or improving the quality of care.

The Lewin Group, with our partners Abt Associates, Inc., GDIT, and Telligen, is under contract to CMS to evaluate the impact of BPCI Models 2, 3, and 4.[4] This is the seventh and final report under this contract. It therefore offers a summative evaluation of the five-year initiative, including results for key payment, utilization, and quality outcomes under each of the three Models, as well as analyses of net spending by the Medicare program. It also includes a comparison of Medicare fee-for-service (FFS) spending reductions under Models 2 and 3 that are calculated with the use of model benchmarks and that are estimated by the evaluation.

A. BPCI Initiative

The BPCI initiative incentivized participants financially for reducing Medicare payments for an episode of care in one of 48 clinical episodes relative to a target price. When a participant's aggregate Medicare episode payments were less than the target price, they could receive Net Payment Reconciliation Amounts (NPRA) from CMS.[5] When aggregate episode payments were higher, participants may have had to repay amounts to CMS.[6] Through this reconciliation process, BPCI was designed to achieve savings to Medicare ranging from 2% to 3.25% of what CMS estimated payments would have been absent the initiative (referred to as the benchmark price).[7]

The roles of the providers and organizations that participated in BPCI were distinguished by whether the entity bore financial risk, could initiate episodes, or served as an administrator or convener. Providers and other organizations that volunteered to participate could enter into the risk-bearing phase of the initiative during a two-year period from October 1, 2013 through September 2015, and enter additional clinical episodes into the risk-bearing phase through December 2015. Providers could stop participating in a given clinical episode at quarterly intervals

[4] Model 1 began earlier than Models 2, 3, and 4 and was evaluated separately; the evaluation and monitoring report found no impact on Medicare savings. The report is available for download from https://innovation.cms.gov/Files/reports/bpci-mdl1yr2annrpt.pdf.

[5] Only Model 2 and Model 3 participants were eligible to receive NPRA. Under Model 4, participants were paid a prospectively determined amount, and they, in turn, paid the providers involved in the episode. There is no NPRA for Model 4 because participants keep any difference between the prospectively determined amount and their payments to other providers for services furnished during the episode.

[6] CMS eliminated downside risk and did not require participants to repay Medicare for a portion of the initiative because of target price and episode attribution errors. CMS also offered participants amendments in participation agreements that limited participants' exposure to risk by applying stop loss and stop gain policies that limited gains and losses to 20% of the target price.

[7] In addition to NPRA, Model 2 and Model 3 participants, as well as Model 4 participants, could generate internal cost savings (ICS). For Model 2 and Model 4 participants, reducing resources used during the hospital stay could contribute to ICS for the hospital, but was unlikely to affect Medicare's payment (unless the hospital length of stay fell below a limit that triggers a per diem payment).

or terminate their participation in the initiative at any time with prior notice. Please refer to the Year 5 Evaluation & Monitoring Annual Report for additional details on the BPCI initiative.[8]

Awardees' agreements with CMS specified their Model choice as well as choices among the 48 clinical episodes, other episode characteristics, and multiple options for Medicare payment policy waivers and financial arrangements with other parties that could be protected under specific waivers of fraud and abuse laws. The clinical episodes were defined by the Medicare Severity-Diagnosis Related Group (MS-DRG) of the anchor or qualifying hospitalization (see **Appendix B** for a list of clinical episodes and associated MS-DRGs). The types of eligible episode-initiating (EI) providers, bundle length options, payment approach, and available waiver options varied by Model, as depicted in Exhibit 1. The services provided during the episode of care were bundled for payment purposes. The bundle definition for each Model is depicted in Exhibit 2. Throughout this report, we refer to Awardees and EIs collectively as "participants." See **Appendix A** for the definition of different types of Awardees and EIs. See the Year 5 Evaluation & Monitoring Annual Report for additional details on the role of conveners in BPCI.

BPCI Participation by Model and Participant Type

Out of the 1.4 million BPCI episodes initiated across Models 2, 3, and 4 during the five-year BPCI initiative,

- Model 2 hospital episodes accounted for 45.8%
- Model 2 PGP episodes accounted for 42.3%
- Model 3 SNF episodes accounted for 6.5%
- Model 3 HHA episodes accounted for 1.9%
- Model 3 PGP, LTCH, and IRF episodes accounted for 2.3%
- Model 4 hospital episodes accounted for 1.1%.

Exhibit 1: Characteristics of BPCI Models 2, 3, and 4

	Eligible Episode Initiator Types	Bundle Length Options	Payment Method	Potential Savings for Participants	Waiver Options
Model 2	✓ Hospital ✓ PGP	✓ 30 days ✓ 60 days ✓ 90 days	Providers continue to be paid on fee-for-service basis and total episode payments are reconciled retrospectively against the target price	✓ NPRA ✓ ICS	✓ SNF three-day rule ✓ Telehealth ✓ Post-discharge home visit ✓ Beneficiary Incentives ✓ Gainsharing
Model 3	✓ SNF ✓ HHA ✓ IRF ✓ LTCH ✓ PGP	✓ 30 days ✓ 60 days ✓ 90 days	Providers continue to be paid on fee-for-service basis and total episode payments are reconciled retrospectively against the target price	✓ NPRA ✓ ICS	✓ Telehealth ✓ Post-discharge home visit ✓ Beneficiary Incentives ✓ Gainsharing
Model 4	✓ Hospital	✓ 30 days	Participants paid a prospectively determined amount and they, in turn, pay the providers involved in the episode.	✓ ICS ✓ Difference between the prospectively determined amount and payments	✓ Beneficiary Incentives ✓ Gainsharing

Note: PGP = physician group practice; SNF = skilled nursing facility; HHA = home health agency; IRF = inpatient rehabilitation facility; LTCH = long-term care hospital; NPRA = net payment reconciliation amount; ICS = internal

[8] The report is available for download from: https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

cost savings. For details on waivers available to BPCI participants, see the Year 5 Evaluation & Monitoring Annual Report. The report is available for download from https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

Exhibit 2: Bundle Definitions under BPCI Models 2, 3, and 4

	Bundle Start	Bundle End	Anchor Hospital Services within the Bundle[1]	Post-discharge Services within the Bundle[1]
Model 2	Anchor hospital admission[2] date	Anchor hospital discharge date plus bundle length (30, 60, or 90 days)	Anchor hospitalization and concurrent professional services	Services furnished from anchor hospital discharge date to bundle end date, including PAC, hospital readmissions, and any professional services
Model 3	PAC admission date	PAC admission date plus bundle length (30, 60, or 90 days)	None[3]	Services furnished from PAC admission date to bundle end date, including PAC, readmissions, and any professional services
Model 4	Anchor hospital admission[4] date	Anchor hospital discharge date plus 30 days (or anchor hospital discharge date if no qualifying readmissions)	Anchor hospitalization and concurrent professional services	Hospital readmissions started within the 30-day readmission window and associated professional services

Note: PAC = post-acute care; MS-DRG = Medicare Severity-Diagnosis Related Group; PGP = physician group practice.
[1] Certain services, such as hospice, readmissions for certain MS-DRGs, and some Part B services were excluded.
[2] Admission to a Model 2 episode-initiating hospital, or the attending or operating physician for the hospitalization was associated with a Model 2 episode-initiating PGP.
[3] The qualifying hospitalization and concurrent professional services are not included within the bundle.
[4] Admission to a Model 4 episode-initiating hospital.

B. Research Questions

This final annual report provides an evaluation of the impact of BPCI on Medicare payments, utilization of services, and quality of care for Medicare beneficiaries under Models 2, 3, and 4, across the five years of the initiative, from October 1, 2013 through September 30, 2018. The report also includes an updated analysis of net savings to the Medicare program as well as an analysis comparing Medicare FFS spending reductions that are calculated with the use of model benchmarks and that are estimated by the evaluation under Model 2 and Model 3.

Three major research questions provide the framework for the analytic approach used in this report. These research questions are outlined in Exhibit 3.

Exhibit 3: BPCI Research Questions

Research Questions

- What was the impact of the five-year BPCI initiative on episode payments, utilization of post-acute care services, and the quality of care for Medicare beneficiaries under Models 2, 3, and 4?
- What was the impact of the five-year BPCI initiative on net savings to the Medicare program?
- How do Medicare FFS spending reductions under BPCI that are calculated with the use of model benchmarks and that are estimated by the evaluation compare?

C. Methods

This report includes impact estimates for key outcomes for episodes initiated by Model 2 hospitals and PGPs, Model 3 SNFs and HHAs, and Model 4 hospitals, as well as estimates of net savings to the Medicare program and an analysis comparing Medicare FFS spending reductions that are calculated with the use of model benchmarks and that are estimated by the evaluation under Model 2 and Model 3. The payment, post-acute care (PAC) utilization, and quality outcomes for the impact estimates and their descriptions are listed in Exhibit 4 below. (See **Appendix C** for additional details.)

Exhibit 4: Outcomes Definitions, by Model

Category	Outcome	Model 2	Model 3	Model 4
Payments	Total Standardized Allowed Payment Amount	Average total Medicare Parts A & B standardized allowed payment amount, during the IP stay and the 90-day PDP	Average total Medicare Part A & B standardized allowed payment amount, from the episode start through 90 days	Average total Medicare Part A & B standardized allowed payment amount, during the IP stay and the 30- and 90-day PDP
	Standardized Allowed Payment Amount, Part A Various Settings	Average Medicare Part A standardized allowed payment amount, during the 90-day PDP, for various settings (SNF, HHA, IRF, and readmissions)	Average Medicare Part A standardized allowed payment amount, from the episode start through 90 days, for various settings (SNF, HHA, IRF, and readmissions)	Average Medicare Part A standardized allowed payment amount, during the 30- and 90-day PDP, for various settings (SNF, HHA, IRF, and readmissions)
Utilization	Discharged to Any PAC Setting	The proportion of episodes that were discharged from the anchor hospital to any PAC setting	N/A	The proportion of episodes that were discharged from the anchor hospital to any PAC setting

Category	Outcome	Model 2	Model 3	Model 4
Utilization (cont'd)	Discharged to Institutional PAC Setting Relative to Discharged to any PAC Setting	The proportion of episodes discharged from the hospital to an institutional PAC setting (SNF, IRF, or long term care hospital) among episodes that were discharged to any PAC setting (including HHA)	N/A	The proportion of episodes discharged from the hospital to an institutional PAC setting (SNF, IRF, or long term care hospital) among episodes that were discharged to any PAC setting (including HHA)
	Number of Days in a SNF	Average number of SNF days of care during the 90-day PDP among episodes with at least one SNF day	Average number of SNF days of care during the 90-day PDP; Presented for SNF-initiated episodes	Average number of SNF days of care during the 90-day PDP among episodes with at least one SNF day
	Number of HHA Visits	N/A	Average number of HHA visits during the 90-day PDP; Presented for HHA-initiated episodes	N/A
Quality	All-cause Mortality	Death from any cause within the 90-day PDP	Death from any cause from the episode start through 90 days	Death from any cause within the 90-day PDP
	Emergency Department Use	Episodes with one or more ED visits for which the beneficiary requires medical treatment but is not admitted to the hospital, within the 90-day PDP	Episodes with one or more ED visits for which the beneficiary requires medical treatment but is not admitted to the hospital, from episode start through 90 days	Episodes with one or more ED visits for which the beneficiary requires medical treatment but is not admitted to the hospital, within the 90-day PDP
	Unplanned Readmission Rate	Episodes with one or more unplanned readmissions for any condition within the 90-day PDP	Episodes with one or more unplanned readmissions for any condition from the episode start through 90 days	Episodes with one or more unplanned readmissions for any condition within the 90-day PDP

Note: PDP = post-discharge period; SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency; IP = inpatient; PAC = post-acute care; ED = emergency department. See **Appendix C** for further information on the outcomes, technical definition, eligible sample, and other details.

The impact estimates for payment, utilization, and quality outcomes and the results of the Medicare savings analyses are presented at various levels in the report and appendices, as illustrated in Exhibit 5. We present estimates of net savings and losses to the Medicare program at the overall model level for Model 2 and Model 3, which represent 99% of BPCI episodes. We present impact estimates at the model and participant type level for Models 2, 3, and 4, based on the clinical episodes with sufficient volume for risk adjustment. We also present the results of the analysis comparing Medicare FFS spending reductions calculated with the use of model benchmarks and that are estimated by the evaluation at the model and participant type level for Models 2 and 3. We further disaggregate the results for Model 2 hospital and Model 2 PGP impact estimates and the estimates of net savings and losses to the Medicare program into surgical and medical clinical episodes. Finally, in **Appendices E through K**, we include results for each model and participant type for all clinical episodes with sufficient volume for risk adjustment. (See **Appendix C** for additional information on the disaggregation of results into surgical and medical episodes and for a

complete list of the clinical episodes with sufficient volume for risk adjustment for each Model and participant type.)

Exhibit 5: Analyses Included in the Report, by Model and Level of Results

Level of Results included in Report	Model 2 Results	Model 3 Results	Model 4 Results
Overall Model	✓ Net savings to Medicare	✓ Net savings to Medicare	
Model-participant type	✓ Impact estimates ✓ Comparison of Medicare FFS spending reductions	✓ Impact estimates ✓ Comparison of Medicare FFS spending reductions	✓ Impact estimates
Surgical/medical clinical episodes	✓ Impact estimates ✓ Net savings to Medicare		
Individual clinical episodes (in appendix)	✓ Impact estimates ✓ Net savings to Medicare ✓ Comparison of Medicare FFS spending reductions	✓ Impact estimates ✓ Net savings to Medicare ✓ Comparison of Medicare FFS spending reductions	✓ Impact estimates

The evaluation relied on multiple data sources. To describe BPCI-participating providers and to select comparison groups, we used provider-level data sources. Medicare claims and enrollment data were used to create outcome measures and beneficiary risk factors associated with the outcomes, and to construct episodes of care for patients at BPCI-participating sites and at matched comparison providers. To estimate net savings and losses to Medicare, we obtained the performance payments made to providers and any amounts owed to CMS (NPRA) from CMS's Healthcare Integrated General Ledger Accounting System (HIGLAS). The data to calculate Medicare FFS spending reductions that are based on model benchmarks came from the reconciliation contractor's reconciliation reports. See **Appendix C** for a detailed description of the data sources used in the evaluation.

The impact analysis uses a difference-in-differences (DiD) design to estimate the differential change in outcomes between the baseline and an intervention period for beneficiaries who received services from BPCI providers relative to beneficiaries who received services from a comparison group of non-BPCI providers. The DiD estimates compare changes in risk-adjusted payment, utilization, and quality outcomes from the baseline period (October 2011 through September 2012) to the intervention period (October 2013 through September 2018).[9] This approach controls for differences in health care service use before the hospitalization, and beneficiary, market, and provider characteristics between BPCI and comparison episodes; eliminates biases from time invariant differences between the BPCI and comparison episodes; and controls for common trends in the BPCI and comparison populations. We present payment results as a percentage of what payments would have been absent the initiative (also referred to as the counterfactual)[10] and in

[9] Because participants may have started to implement changes in preparation for BPCI, we exclude Phase 1 of BPCI, the one-year period from October 2012 through September 2013. During this time, participants could begin signing up for BPCI but no participants had entered Phase 2, the risk-bearing or intervention phase.

[10] The counterfactual is estimated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention.

dollars. We discuss results that are statistically significant at the 5% level, and in charts and tables, we display statistical significance at both the 5% and 10% level. See **Appendix C** for a detailed description of the DiD methodology, outcome definitions, methods for identifying comparison populations, statistical models, sensitivity analyses, and tests for parallel trends between BPCI and comparison episodes in the baseline period.

The change in net Medicare spending was calculated by subtracting reconciliation payments from the change in aggregate non-standardized payments for all 48 clinical episodes under Model 2 and Model 3, between the baseline period (October 2011 through September 2012) and the intervention period (October 2013 through September 2018). Medicare FFS spending reductions that are calculated with the use of model benchmarks is the difference between the benchmark price (participants' historical allowed amounts, based on data from Q3 2009 through Q2 2012, trended forward using a retrospective nation-wide trend and accounting for MS-DRG and provider characteristics) and average FFS episode payments to BPCI participants during the intervention. Medicare FFS spending reductions that are estimated by the evaluation was calculated with the use of a retrospective comparison group within a DiD framework, in which the outcome was the total standardized allowed payment amount during the bundle. The DiD is the difference between the counterfactual and risk-adjusted average FFS episode payments to BPCI participants during the intervention. We present all results as a percentage of the counterfactual and in dollars.

II. Model 2 Results

A. Impact of BPCI on Hospital-initiated Episodes

1. Key Findings

- **Impact on Payments**
  - During the five years of the BPCI initiative, total Medicare-allowed payments during the inpatient stay plus 90 days post discharge declined by 3.1% (or $836) for Model 2 hospital-initiated episodes, relative to a comparison group. Among surgical clinical episodes, total Medicare-allowed payments during the inpatient stay plus 90 days post discharge had a relative decline of 3.9% (or $1,195) while for medical clinical episodes there was a relative decline of 2.6% (or $652).
  - SNF payments in the 90 days post discharge had a relative decline of 13.5% (or $700) for hospital-initiated episodes, driving the reduction in total payments. For surgical clinical episodes, SNF payments in the 90 days post discharge had a relative decline of 15.4% (or $843), and medical clinical episodes had a relative decline of 11.6% (or $578).
  - HHA payments had a relative increase of 4.6% (or $75) for Model 2 hospital-initiated episodes. There was no statistically significant change in HHA payments for surgical clinical episodes, while among medical clinical episodes, HHA payments had a relative increase of 6.3% (or $81).
- **Impact on Post-Acute Care Utilization**
  - Among patients who were discharged to any PAC setting, there was a relative decline of 3.6 percentage points in the proportion discharged to institutional PAC settings for Model 2 hospital-initiated episodes. For surgical clinical episodes, there was a relative decline of 4.9 percentage points, while for medical clinical episodes there was a relative decline of 2.1 percentage points.
  - Among patients with at least one day in a SNF, there was a relative decline of 2.9 SNF days for Model 2 hospital-initiated episodes. The declines for surgical and medical clinical episodes were similar with a relative decline of 2.7 and 2.9 SNF days, respectively.
- **Impact on Quality**
  - In general, claims-based quality measures did not change under Model 2 for hospital-initiated episodes relative to the comparison group.

2. Sample Characteristics

Of the 423 hospitals that voluntarily participated in BPCI Model 2, we analyzed the baseline characteristics of 419 hospitals that received Medicare certification by 2011. These 419 hospitals were different from non-participating hospitals based on key characteristics, as described in Exhibits 6a-6b. BPCI Model 2 hospitals were more likely to be non-profit and located in urban areas compared to non-participating hospitals. BPCI-participating hospitals also had higher bed counts and larger teaching programs, as indicated by the higher resident-to-bed ratios.

Standardized Part A payments for the inpatient stay plus the 90-day post-discharge period (PDP) averaged 6% higher in 2011 across all clinical episodes for BPCI-participating hospitals relative to non-participating hospitals. See **Appendix C** for further details about these measures and **Appendix E** of the Year 5 Evaluation & Monitoring Annual Report for more details and additional sample characteristics.

Exhibit 6a and 6b: Baseline Characteristics of All BPCI-participating Hospitals and Non-participating Hospitals, Model 2

Domain	Characteristic	BPCI Hospitals (N)	BPCI Hospitals (%)	Non-participating Hospitals (N)	Non-participating Hospitals (%)
Ownership	For Profit	66	16%	638	23%
	Government	32	8%	542	20%
	Non-Profit	321	77%	1,594	57%
Urban/Rural	Rural	32	8%	872	31%
	Urban	387	92%	1,902	69%
Part of Chain	Yes	216	52%	1,469	53%

Characteristic	BPCI Hospitals (mean)	Non-participating Hospitals (mean)
Bed Count	311	175
Number of Discharges for BPCI Episode MS-DRGs, 2011	3,004	1,598
Medicare Days Percent	39%	42%
Resident-to-bed Ratio	0.12	0.05
Disproportionate Share Percent	27%	29%
Hospital Market Share	21%	27%

Note: Data from 419 BPCI hospitals and 2,774 non-participating hospitals. MS-DRG = Medicare Severity-Diagnosis Related Group.

Source: Lewin analysis of 2013 Provider of Service (POS) files and 2011 Medicare claims. BPCI-participating hospitals are defined as hospitals participating in Model 2. Non-participating hospitals are all other hospitals not participating in any BPCI initiative that reported values for all measures listed above and are not in Maryland. Please note that BPCI-participating hospitals that received Medicare certification after 2011 are not included in this table.

Of the 423 BPCI-participating hospitals, we were able to identify comparison hospitals for 406 hospitals in the analysis of the impact estimates. Exhibit 7 describes the sample of Model 2 hospitals included in the analysis, with characteristics for Model 2 hospitals for all clinical episodes, for surgical clinical episodes, and for medical clinical episodes. The BPCI hospitals included in the impact analysis initiated 576,124 episodes during the five-year initiative and participated for an average of ten quarters. By the end of the initiative, 317 (78.1%) Model 2 hospitals in the analytical sample stopped participating in at least one clinical episode, and

162 (39.9%) terminated their participation in BPCI completely.[11] Thirty percent of the episodes in the analytical sample were initiated by hospitals that stopped participating in the clinical episode.

Exhibit 7: Characteristics of the BPCI Providers Included in the BPCI Impact Estimates, Model 2 Hospitals, Q4 2013 – Q3 2018

Clinical Episodes	BPCI Hospitals (N)	BPCI Episodes (N)	Average Length of Participation (Quarters)[a]	Hospitals that Stopped Participating in at least one Clinical Episode (N)	Proportion of Episodes from Hospitals that Stopped Participating in the Clinical Episode (%)
Model 2 Hospitals Overall	**406**	**576,124**	**10**	**317**	**30.0%**
Surgical Clinical Episodes	344	259,819	10	233	27.5%
Medical Clinical Episodes	284	316,305	9	239	32.1%

Note: Model 2 Hospitals Overall represents hospital-initiated episodes in the analytical sample in any of the 32 clinical episodes that had sufficient volume for risk adjustment. The analytical sample includes 89% of the episodes initiated in the 32 clinical episodes and 88% of all BPCI Model 2 hospital-initiated episodes. The number of BPCI hospitals that stopped participating in the clinical episode represents unique hospitals in the analytical sample across the 32 clinical episodes. Average length of participation and the proportion of episodes from hospitals that stopped participating in the clinical episode are calculated as an average of all hospital/clinical episode combinations in the analytical sample across the 32 clinical episodes.

[a] The average length of participation varies because providers and other organizations that volunteered to participate in BPCI could enter into the risk-bearing phase of the initiative during a two-year period through September 2015, and they could enter additional clinical episodes through December 2015. Providers could stop participating in a given clinical episode at quarterly intervals or terminate their participation in the initiative completely at any time with 60 days advance notice.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI providers.

3. Payment, Utilization, and Quality Outcomes

This section presents the BPCI impact estimates for payments, utilization, and quality for episodes initiated by acute care hospitals over the course of the five-year initiative (Q4 2013 through Q3 2018). We present the total standardized allowed payment amount for the inpatient stay and the 90-day PDP, as well as the standardized allowed payment amount for SNF, IRF, and HHA services and for readmissions during the 90-day PDP. We present the results separately for Model 2 hospital-initiated episodes, surgical clinical episodes, and medical clinical episodes. Detailed results by clinical episode are located in **Appendix E**.

a. How have the average standardized payments changed under BPCI?

By the end of the five-year BPCI initiative, the total payment amount declined from baseline to the intervention period for BPCI Model 2 hospital episodes more than the comparison group by an estimated 3.1% ($836, $p<0.05$) of what payments would have been absent BPCI. Both surgical and medical clinical episodes had relative declines in total payments as well. Surgical clinical episodes

[11] By the end of the five-year initiative, approximately 41% of Model 2 hospitals that had ever participated in BPCI (regardless of whether they were in the analytical sample) withdrew completely from the initiative and no longer participated in any clinical episodes.

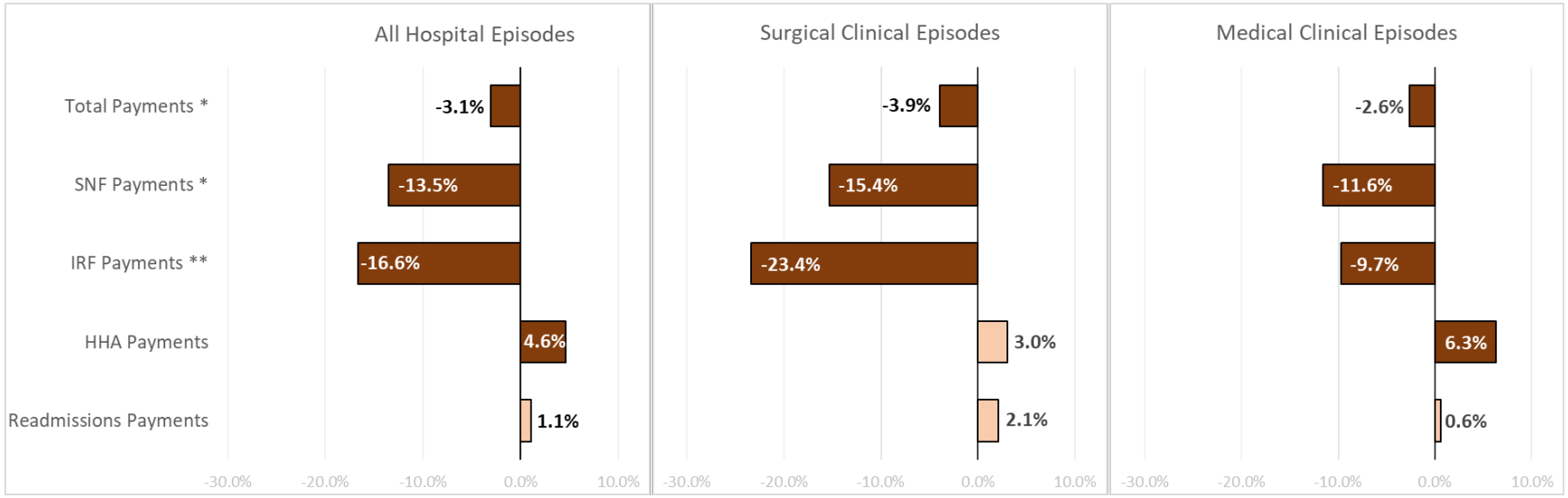
had a decline of 3.9% ($1,195, $p<0.05$), and medical clinical episodes had a decline of 2.6% ($652, $p<0.05$) (Exhibits 8 and 9).

The declines in total payments were driven by declines in SNF and IRF payments. Across all clinical episodes, SNF payments declined an estimated 13.5% ($700, $p<0.05$) relative to what payments would have been absent the initiative. SNF payments declined approximately 15.4% ($843, $p<0.05$) for surgical clinical episodes and 11.6% ($578, $p<0.05$) for medical clinical episodes. While both surgical and medical clinical episodes had declines in IRF payments, the decline was statistically significantly greater for surgical clinical episodes. IRF payments decreased an estimated 16.6% ($203, $p<0.05$) for all clinical episodes, 23.4% ($349, $p<0.05$) for surgical clinical episodes, and 9.7% ($102, $p<0.05$) for medical clinical episodes.

While SNF and IRF payments declined for all clinical episodes, HHA payments increased an estimated 4.6% ($75, $p<0.05$). HHA payments also increased for medical clinical episodes by an estimated 6.3% ($81), while the increase for surgical clinical episodes was not statistically significant.

Although the magnitude of the declines in SNF and IRF payments are greater for surgical clinical episodes than they are for medical clinical episodes, HHA payments only increased for medical clinical episodes. The increase in HHA payments suggest a shift towards less costly services for medical clinical episodes, whereas there is no evidence of this shift for surgical clinical episodes.

Exhibit 8: Percent Change in Standardized Allowed Payment Amount Outcomes, Inpatient Stay and 90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 32 clinical episodes that had sufficient volume to allow for risk-adjustment. The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is calculated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. PDP = post-discharge period; SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

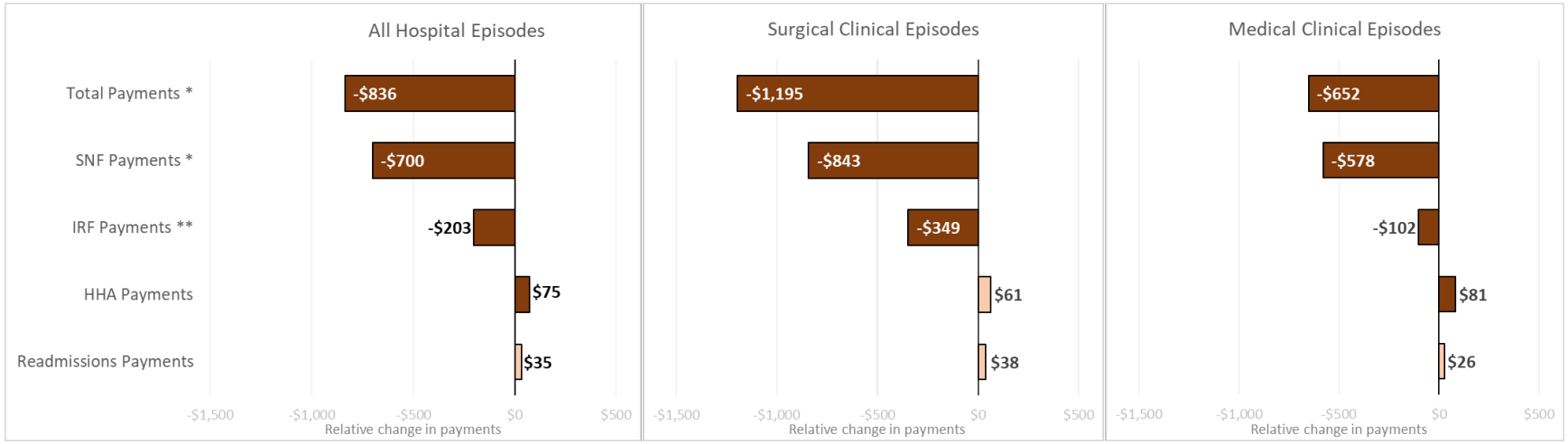
* The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 10% level of significance.

** The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 5% level of significance.

[a] Total Payments includes Part A and B payments during the inpatient stay and 90-day PDP. All other outcomes include payments during the 90-day PDP. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

Exhibit 9: Impact of BPCI on Standardized Allowed Payment Amount Outcomes, Inpatient Stay and 90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 32 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. PDP = post-discharge period; SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

* The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 10% level of significance.

** The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 5% level of significance.

[a] Total Payments includes Part A and B payments during the inpatient stay and 90-day PDP. All other outcomes include payments during the 90-day PDP. Payment measures are not conditional upon the use of the service.

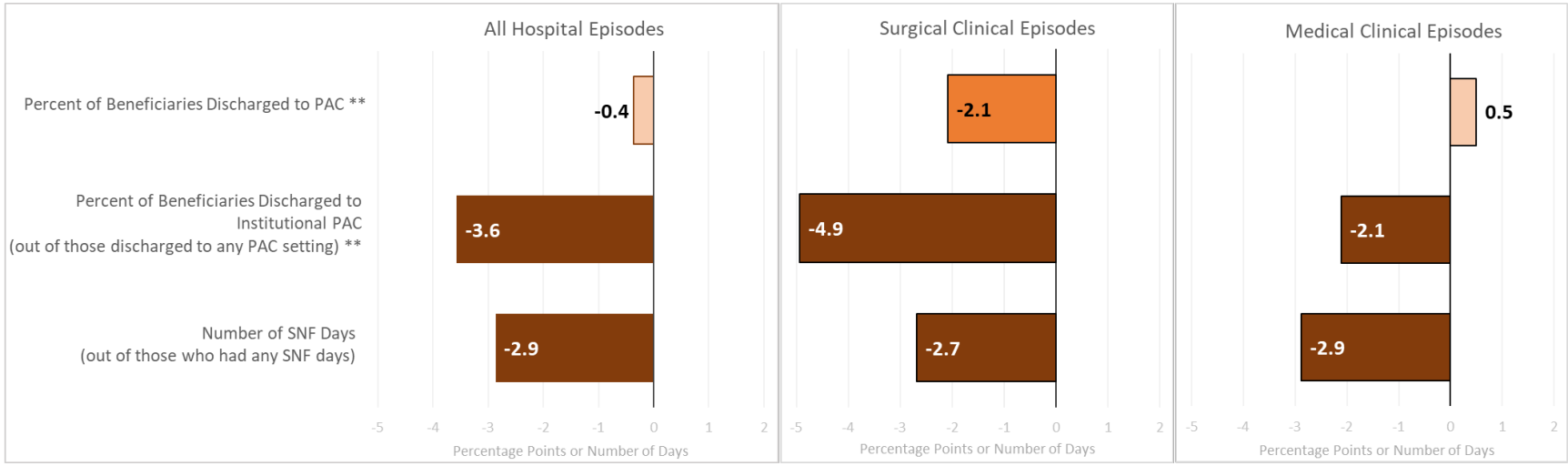
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

b. How have the services changed under BPCI?

The changes in service use were consistent with the changes in payments. There was a relative decrease in the share of episodes discharged to institutional PAC (SNF, IRF, or long-term care hospitals (LTCH), among those discharged to any PAC, for BPCI Model 2 hospital-initiated episodes overall and for surgical and medical clinical episodes (Exhibit 10). The share of institutional PAC discharges decreased by a relative 3.6 percentage points ($p<0.05$) for all hospital-initiated episodes. For surgical clinical episodes, the share of institutional PAC discharges decreased 4.9 percentage points ($p<0.05$), which was statistically significantly greater than the decline of 2.1 percentage points ($p<0.05$) for medical clinical episodes. For BPCI episodes with at least one day in a SNF, the number of SNF days during the 90-day PDP declined by 2.9 days for all hospital-initiated episodes ($p<0.05$). The relative decrease in SNF days was similar for surgical and medical clinical episodes; SNF days declined 2.7 days for surgical clinical episodes and 2.9 days for medical clinical episodes ($p<0.05$).

For medical clinical episodes, the changes in service use indicate a shift from more expensive institutional PAC discharges to less expensive HHA discharges. For surgical clinical episodes, the changes in service use indicate a reduction in both institutional and overall PAC discharges.

**Exhibit 10: Impact of BPCI on PAC Utilization,
90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes,
Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018**



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 32 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant. PAC = post-acute care; SNF = skilled nursing facility.

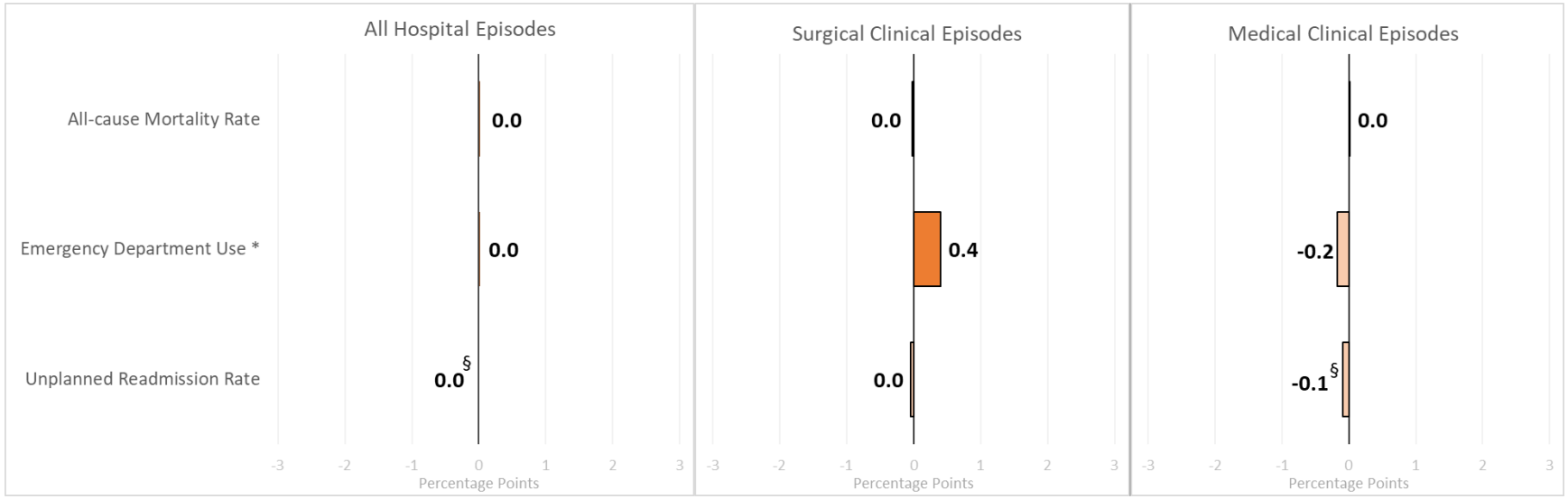
** The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 5% level of significance.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

c. How has quality of care changed under BPCI?

Under BPCI, the incentive to lower episode payments could lead to changes in service use that lower the quality of care. We examined three key claims-based measures during the 90-day PDP to assess changes in quality and found no statistically significant changes in mortality, ED use, or readmission rates at the 5% level for hospital-initiated episodes overall or for surgical or medical clinical episodes (Exhibit 11).

**Exhibit 11: Impact of BPCI on Quality Outcomes,
90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes,
Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018**



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 32 clinical episodes that had sufficient volume to allow for risk-adjustment. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant.

§ Data from the baseline period shows BPCI and matched comparison episodes were not on parallel trends for this outcome, which is required for an unbiased estimate.

* The results for this outcome are statistically significantly different for surgical and medical episodes at the 10% level of significance.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

B. Impact of BPCI on PGP-initiated Episodes

1. Key Findings

- **Impact on Payments**
  - During the five years of the BPCI initiative, total Medicare-allowed payments during the inpatient stay plus 90 days post discharge declined by an estimated 4.9% (or $1,260) for episodes initiated by BPCI Model 2 PGPs relative to a comparison group. Among surgical clinical episodes, total Medicare-allowed payments during the inpatient stay plus 90 days post discharge had a relative decline of 7.4% (or $2,015), while medical clinical episodes had a relative decline of 2.0% (or $491).
  - SNF payments in the 90 days post discharge had a relative decline of 14.2% (or $693) for PGP-initiated episodes, contributing to the reduction in total payments. For surgical clinical episodes, SNF payments had a relative decline of 21.3% (or $1,100), while medical clinical episodes had a relative decline of 5.7% (or $265).
  - IRF payments in the 90 days post discharge had a relative decline of 32.3% ($300) for all PGP-initiated episodes. For surgical clinical episodes, IRF payments had a relative decline of 51.7% ($575), while there was no statistically significant change for medical clinical episodes.
  - HHA payments had a relative decrease of 12.7% (or $196) for PGP-initiated episodes. For surgical clinical episodes, HHA payments had a relative decrease of 21.1% (or $415), while there was no statistically significant change for medical clinical episodes.
- **Impact on Post-Acute Care Utilization**
  - There was a 5.6 percentage point relative decline in the share of beneficiaries discharged to PAC settings for BPCI Model 2 PGP-initiated episodes, relative to a comparison group. This was driven by a 10.6 percentage point relative decline in the share of PAC setting discharges among surgical clinical episodes; there was no relative change for medical clinical episodes.
  - Among patients who were discharged to any PAC setting, there was a relative decline of 3.2 percentage points in the proportion discharged to institutional PAC settings for surgical clinical episodes, while the change was not statistically significant at the 5% level for Model 2 PGP-initiated episodes overall or for medical clinical episodes.
  - Among patients with at least one day in the SNF, there was a relative decline of 2.3 SNF days for Model 2 PGP-initiated episodes. There was a relative decrease of 3.6 SNF days for surgical clinical episodes and 0.9 SNF days for medical clinical episodes.
- **Impact on Quality**
  - In general, claims-based quality measures did not change under Model 2 for PGP-initiated episodes relative to the comparison group. However, for surgical

clinical episodes, there was a relative decrease of 0.8 percentage points in ED use.

2. Sample Characteristics

We analyzed characteristics of BPCI-participating PGPs, such as the composition of physician specialties and volume of discharges.[12] Of the 272 PGPs participating in Model 2 of BPCI, 245 were included in the analysis of PGP characteristics.[13] Of these 245 PGPs, 20% were classified as hospitalist practices, 26% were single-specialty practices and 54% were multi-specialty practices. Most single-specialty practices were in the surgical specialty category, making up 18% of all PGPs. The proportion of primary care physicians among BPCI-participating physicians rose between 2012 and 2016, from 33% to 40%. Finally, the average number of discharges from BPCI-participating PGPs varied from 0 to more than 10,000 per quarter for MS-DRGs included in the 48 BPCI clinical episodes. See **Appendix D** for further details about participant characteristics, including physician specialty categories.

Of the 272 BPCI-participating Model 2 PGPs, we were able to identify comparison groups for 189 PGPs in the analysis of the impact estimates. Exhibit 12 describes the sample of Model 2 PGPs included in the impact analysis, with characteristics for Model 2 PGPs across all clinical episodes, for surgical clinical episodes, and for medical clinical episodes. The BPCI PGPs included in the impact analysis initiated 266,589 episodes during the five-year initiative and participated for an average of nine quarters. By the end of the initiative, 140 of the 189 (74.1%) Model 2 PGPs in the analytical sample stopped participating in at least one clinical episode, and 55 (29.1%) terminated their participation in BPCI completely.[14] Thirty-two percent of the episodes in the analytical sample were initiated by PGPs that stopped participating in the clinical episode.

[12] In this report, we summarize the main findings of the analysis from the Year 5 Evaluation & Monitoring Annual Report. See the Year 5 Evaluation & Monitoring Annual Report for the full analysis of PGP characteristics. The report is available for download from https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

[13] We required at least one physician to be associated with the PGP in the baseline and one physician in the intervention period for the PGP to be included in the descriptive analysis. The number of PGPs included in this analysis (245) differs from the number of PGPs in the analytical sample (248) because this requirement does not apply to the analytical sample. See **Appendix C** for further details about the analytical sample.

[14] By the end of the five-year initiative, approximately 44% of Model 2 PGPs that had ever participated in BPCI (regardless of whether they were in the analytical sample) withdrew completely from the initiative and no longer participated in any clinical episodes

Exhibit 12: Characteristics of the BPCI Providers Included in the BPCI Impact Estimates, Model 2 PGPs, Q4 2013 – Q3 2018

Clinical Episodes	BPCI PGPs (N)	BPCI Episodes (N)	Average Length of Participation (Quarters)[a]	PGPs that Stopped Participating in at least one Clinical Episode (N)	Proportion of Episodes from PGPs that Stopped Participating in the Clinical Episode (%)
Model 2 PGPs Overall	**189**	**266,589**	**9**	**140**	**32.2%**
Surgical Clinical Episodes	156	129,940	10	93	23.3%
Medical Clinical Episodes	102	136,649	9	95	40.6%

Note: Model 2 PGPs Overall represents the PGP-initiated episodes in the analytical sample in any of the 21 clinical episodes that had sufficient volume for risk adjustment. The analytical sample includes 54% of the episodes initiated in the 21 clinical episodes and 44% of all BPCI Model 2 PGP-initiated episodes. The number of BPCI PGPs that stopped participating in the clinical episode represents unique PGPs in the analytical sample across the 21 clinical episodes. Average length of participation and the proportion of episodes from PGPs that stopped participating in the clinical episode are calculated as an average of all PGP/clinical episode combinations in the analytical sample across the 21 clinical episodes. PGP = physician group practice.

[a] The average length of participation varies because providers and other organizations that volunteered to participate in BPCI could enter into the risk-bearing phase of the initiative during a two-year period through September 2015, and they could enter additional clinical episodes through December 2015. Providers could stop participating in a given clinical episode or terminate their participation in the initiative at any time.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI providers.

3. Payment, Utilization, and Quality Outcomes

This section presents the BPCI impact estimates for payments, utilization, and quality for PGP-initiated episodes over the course of the five-year initiative (Q4 2013 through Q3 2018). We present the total standardized allowed payment amount for the inpatient stay and the 90-day PDP, as well as the standardized allowed payment amount for SNF, IRF, and HHA services and for readmissions during the 90-day PDP. We present the results separately for Model 2 PGP-initiated episodes overall, surgical clinical episodes, and medical clinical episodes. Detailed results by clinical episode are located in **Appendix F**.

a. How have the average standardized payments changed under BPCI?

During the five years of the BPCI initiative, the total payment amount declined from the baseline to the intervention period for BPCI Model 2 PGP episodes more than the comparison group by 4.9% of what payments would have been absent the BPCI initiative ($1,260, $p<0.05$). Total payments declined by an estimated 7.4% ($2,015, $p<0.05$) for surgical clinical episodes, which was a statistically significantly greater decline than the 2.0% ($491, $p<0.05$) for medical clinical episodes (Exhibits 13 and 14).

Reductions in payments for services in PAC settings during the 90-day PDP were responsible for the decline in total payments. SNF payments drove the declines in total payments for all clinical episodes, surgical clinical episodes, and medical clinical episodes. SNF payments declined 14.2% ($693, $p<0.05$) for all PGP-initiated episodes. The decline in SNF payments was statistically significantly greater for surgical clinical episodes at the 5% level than they were for medical

clinical episodes; SNF payments declined an estimated 21.3% ($1,100, $p<0.05$) for surgical clinical episodes and 5.7% ($265, $p<0.05$) for medical clinical episodes.

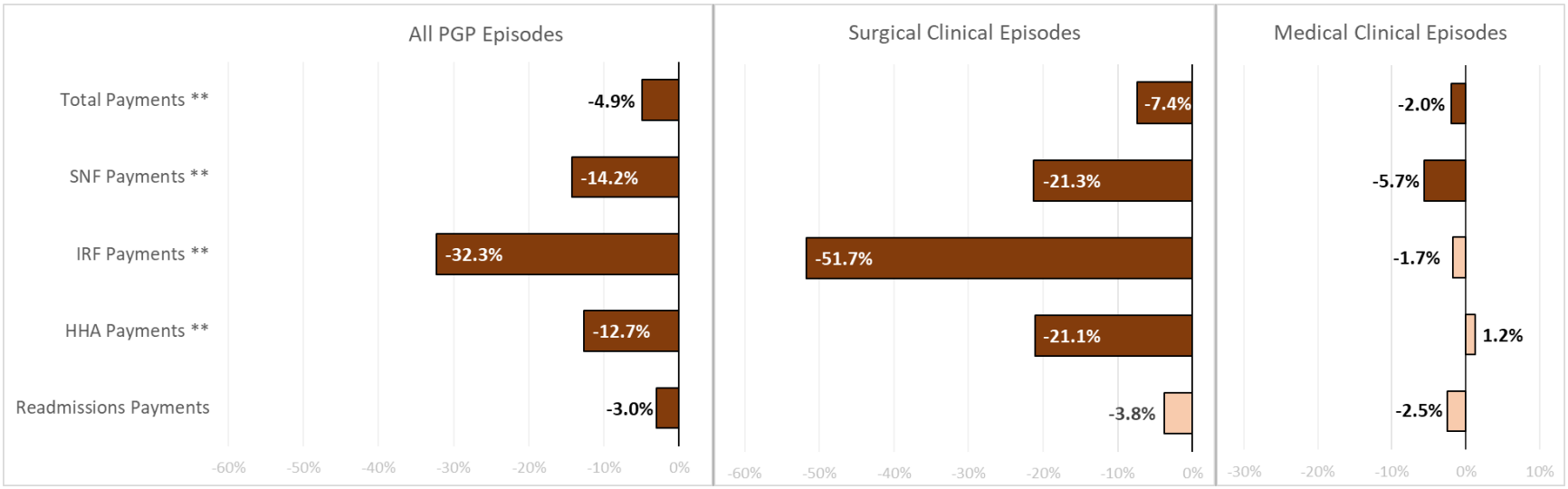
For surgical clinical episodes, declines in SNF payments were accompanied by declines in IRF and HHA payments, but this was not the case for medical clinical episodes. IRF payments decreased an estimated 32.3% ($300, $p<0.05$) for all PGP-initiated episodes and 51.7% ($575, $p<0.05$) for surgical clinical episodes. HHA payments decreased approximately 12.7% ($196, $p<0.05$) for all episodes and 21.1% ($415, $p<0.05$) for surgical clinical episodes. The changes in IRF and HHA payments were not statistically significant for medical clinical episodes.

Readmissions payments declined an estimated 3.0% ($88, $p<0.05$) for PGP episodes overall, but the changes for surgical and medical clinical episodes were not statistically significant.[15]

The changes in payments for Model 2 PGP episodes showed a decrease in all three PAC settings, SNF, IRF, and HHA, and these changes were driven by surgical clinical episodes; there were no statistically significant changes in IRF or HHA payments for medical PGP episodes.

[15] The surgical- and medical-level estimates are not statistically significant, and have larger p-values (0.1 and 0.2, respectively) due to smaller sample sizes.

Exhibit 13: Percent Change in Standardized Allowed Payment Amount Outcomes, Inpatient Stay and 90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes, Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



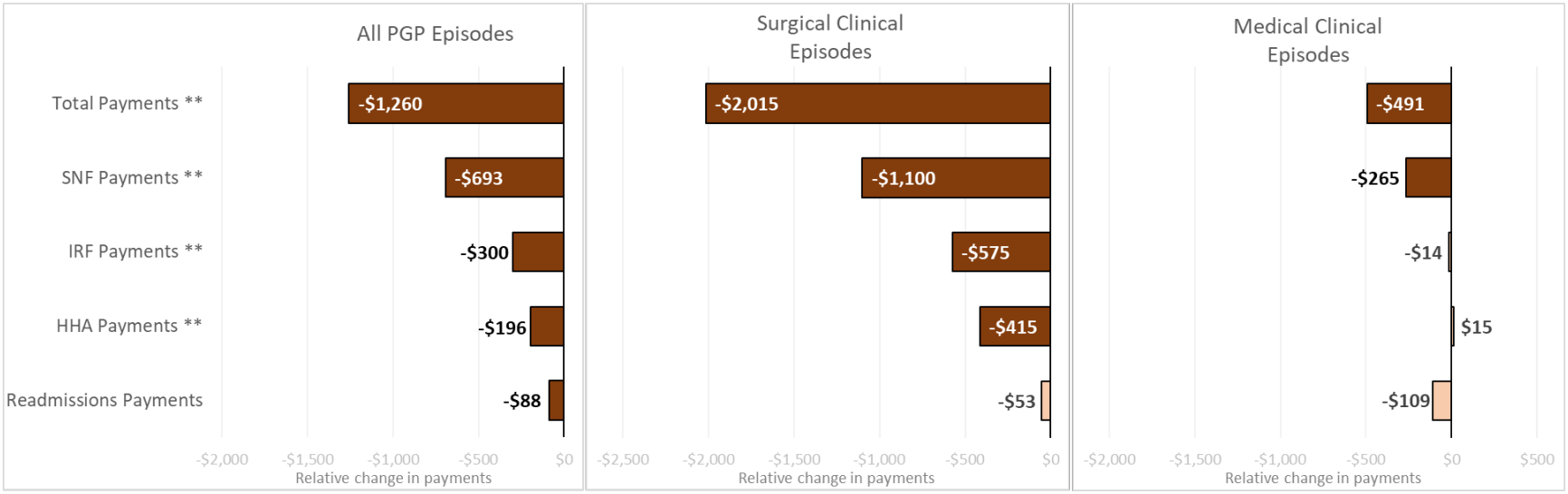
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 21 clinical episodes that had sufficient volume to allow for risk-adjustment. The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is calculated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. PGP = physician group practice; SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

** The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 5% level of significance.

[a] Total Payments includes Part A and B payments during the inpatient stay and 90-day post-discharge period. All other outcomes include payments during the 90-day PDP. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

Exhibit 14: Impact of BPCI on Standardized Allowed Payment Amount Outcomes, Inpatient Stay and 90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes, Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 21 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. PDP = post-discharge period. PGP = physician group practice; SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

** The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 5% level of significance.

[a] Total Payments includes Part A and B payments during the inpatient stay and 90-day PDP. All other outcomes include payments during the 90-day PDP. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

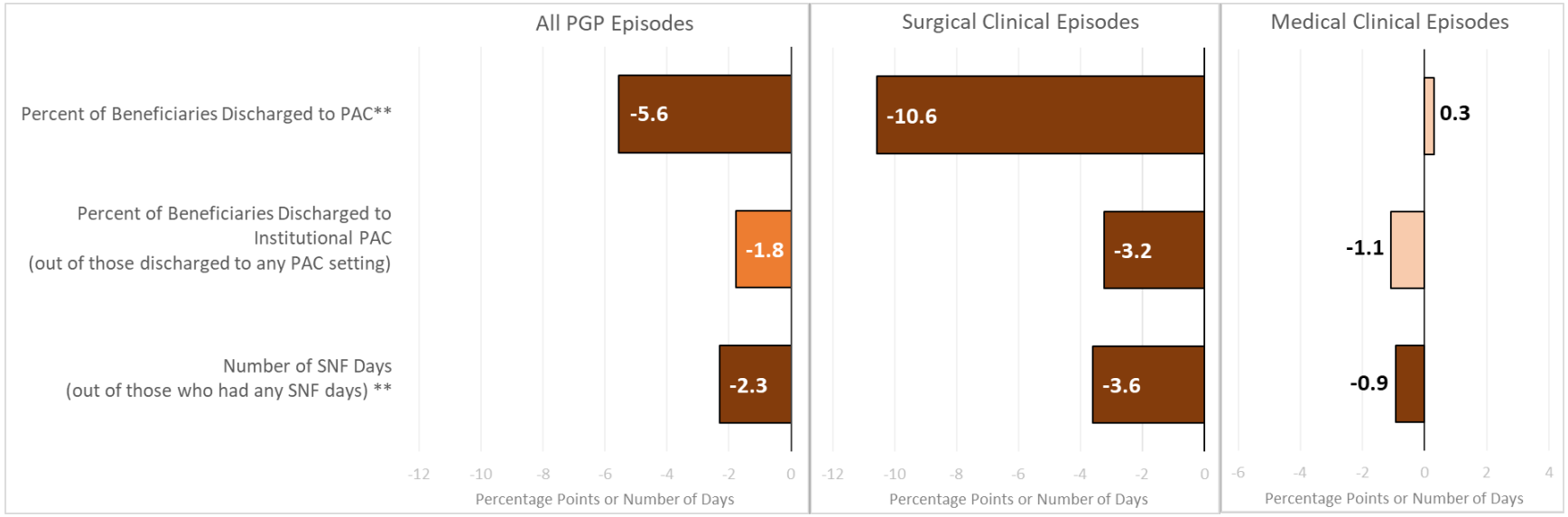
b. How have the services changed under BPCI?

The changes in service utilization were consistent with the changes in payments. For surgical clinical episodes, there was a decline in any PAC use, institutional PAC use, and the number of SNF days. For medical clinical episodes, the only statistically significant change in service use was the decline in the number of SNF days (Exhibit 15).

The proportion of patients discharged to any PAC setting declined an estimated 5.6 percentage points for all PGP episodes and an estimated 10.6 percentage points for surgical clinical episodes ($p<0.05$). There was no statistically significant change in the proportion of patients discharged to PAC settings for medical clinical episodes. Among those discharged to a PAC setting, the proportion of beneficiaries discharged to an institutional PAC setting decreased an estimated 3.2 percentage points for surgical clinical episodes ($p<0.05$), with no statistically significant change for all PGP episodes or medical PGP episodes. Among those who spent any time in a SNF, the length of stay in a SNF decreased an estimated 2.3 days for all clinical episodes, 3.6 days for surgical clinical episodes, and 0.9 days for medical clinical episodes ($p<0.05$). The declines in SNF days were statistically significantly different for surgical and medical clinical episodes.

The changes in service utilization for surgical and medical clinical episodes indicate surgical clinical episodes reduced the share of patients discharged to any PAC setting, the share discharged to institutional PAC settings, and the number of SNF days, while medical clinical episodes only reduced the number of SNF days.

**Exhibit 15: Impact of BPCI on PAC Utilization,
90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes,
Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2018**



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 21 clinical episodes that had sufficient volume to allow for risk-adjustment Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant. PGP = physician group practice; PAC = post-acute care; SNF = skilled nursing facility.

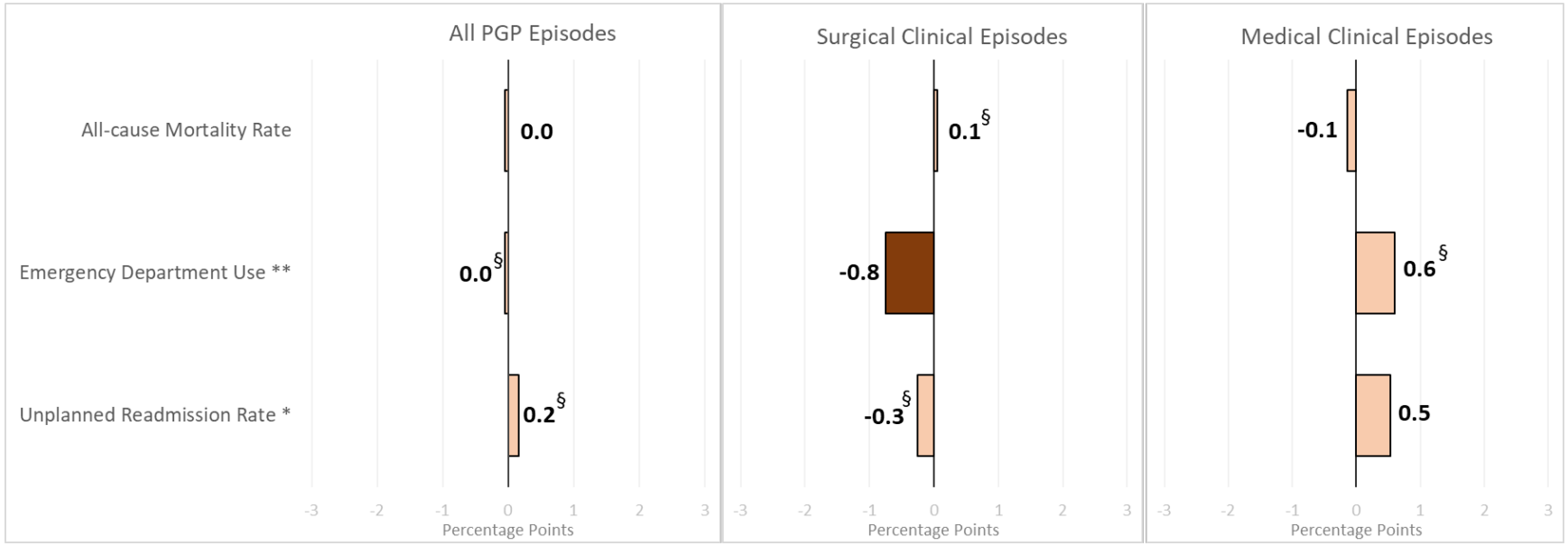
** The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 5% level of significance.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

c. How has quality of care changed under BPCI?

The results from the claims-based quality measures show no clear pattern of change in the quality of care under BPCI for Model 2 PGP episodes. ED use during the 90-day PDP declined an estimated 0.8 percentage points for surgical clinical episodes ($p<0.05$), but there were no statistically significant changes in ED use for medical clinical episodes or for PGP episodes overall. Furthermore, there were no statistically significant changes in the mortality or unplanned readmission rates during the 90-day PDP for surgical, medical, or PGP episodes overall (Exhibit 16).

**Exhibit 16: Impact of BPCI on Quality Outcomes,
90-day PDP, for All Episodes, Surgical Clinical Episodes, and Medical Clinical Episodes,
Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2018**



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 21 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant. PDP = post-discharge period; PGP = physician group practice.

§ Data from the baseline period shows BPCI and matched comparison episodes were not on parallel trends for this outcome, which is required for an unbiased estimate.

* The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 10% level of significance.

** The results for this outcome are statistically significantly different for surgical and medical clinical episodes at the 5% level of significance.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

III. Model 3 Results

A. Impact of BPCI on SNF-initiated Episodes

1. Key Findings

- **Impact on Payments**
  - During the five years of the BPCI initiative, the total standardized allowed payment amount from episode start through 90 days declined by 7.6% (or $2,171) for Model 3 SNF-initiated episodes overall, relative to a comparison group.
  - SNF allowed payment amounts decreased by a relative 13.2% (or $2,387) for SNF-initiated episodes, contributing to the reduction in total payments.
  - Home health allowed payment amounts increased by a relative 9.2% (or $170) for SNF-initiated episodes.
- **Impact on Post-Acute Care Utilization**
  - There was a relative decline of 3.5 SNF days for Model 3 SNF-initiated episodes.
- **Impact on Quality**
  - In aggregate, claims-based quality measures did not change relative to the comparison group under Model 3 for SNF-initiated episodes. However, three out of 11 clinical episodes had an increase in the mortality rate, ED use, or unplanned readmissions, potentially indicating a decrease in quality. However, these three clinical episodes suffer from small sample sizes both overall and per participant making it difficult to ascertain whether these results are a signal or simply noise.

2. Sample Characteristics

Of the 873 SNFs that voluntarily participated in BPCI Model 3, we analyzed the characteristics of 864 SNFs that received Medicare certification by 2011 and of 493 SNFs for which we were able to identify comparison groups for the impact estimates (referred to as "matched SNFs").[16] The 493 matched SNFs were similar to the group of all BPCI Model 3 SNFs based on key characteristics, as described in Exhibits 17a-17b, but these 864 BPCI SNFs were different from non-participating SNFs. A higher proportion of BPCI-participating SNFs were for-profit organizations (86%) compared with non-participating SNFs (70%) and were more likely to be part of a chain (52% vs. 22%). They had a higher average bed count (122 vs. 112), and they averaged more admissions for BPCI episode MS-DRGs in 2011 (136 vs. 94) than non-participating SNFs. (See **Appendix D** for further details about these measures and **Appendix E** of the Year 5 Evaluation & Monitoring Annual Report for more details and additional sample characteristics.)

[16] **Appendix C** provides further information on the methods for identifying comparison groups for the analysis of the impact estimates.

Exhibits 17a and 17b: Baseline Characteristics of All BPCI-participating SNFs and Non-participating SNFs, Model 3

Domain	Characteristic	Matched BPCI SNFs (N)	Matched BPCI SNFs (%)	All BPCI SNFs (N)	All BPCI SNFs (%)	Non-participating SNFs (N)	Non-participating SNFs (%)
Ownership	For Profit	399	81%	740	86%	9,374	70%
	Government	3	1%	3	0%	617	5%
	Non-Profit	91	18%	121	14%	3,311	25%
Urban/Rural	Rural	59	12%	140	16%	3,938	30%
	Urban	434	88%	724	84%	9,364	70%
IRF in CBSA	Yes	272	55%	488	56%	7,238	54%
Hospital-Based	Yes	5	1%	7	1%	586	4%
Part of Chain	Yes	115	23%	216	25%	2,946	22%

Characteristic	Matched BPCI SNFs (mean)	All BPCI SNFs (mean)	Non-participating SNFs (mean)
Bed Count	129	122	112
Number of Admissions for BPCI Episode MS-DRGs, 2011	166	136	94
SNF Market Share	7%	6%	6%
Nursing Home Overall Score[a]	3.68	3.45	3.32

Note: Data from 864 BPCI SNF episode initiators and 13,302 non-participating SNFs are included in this exhibit. SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; CBSA = core-based statistical area; MS-DRG = Medicare Severity-Diagnosis Related Group.

[a] This indicates the number of points out of 5 in overall rating and in three areas: Quality, Survey/Health Inspections, and Staffing. The closer to 5 the better the quality, inspections, and staffing.

Source: Lewin analysis of 2013 Provider of Service (POS) files and 2011 Medicare claims. BPCI-participating SNFs are defined as SNFs participating in Model 3. Non-participating SNFs are all other SNFs not participating in any BPCI initiative and that reported values for all measures listed in the table above. BPCI-participating SNFs that received Medicare certification after 2011 are not included in this table.

Exhibit 18 describes the sample of Model 3 SNFs included in the analysis, with characteristics for Model 3 SNFs across all clinical episodes. The BPCI SNFs included in the impact analysis initiated 45,694 episodes during the five-year initiative and participated for an average of ten quarters. By the end of the initiative, 335 (68.0%) Model 3 SNFs in the analytical sample stopped participating in at least one clinical episode, and 156 (31.6%) terminated their participation in BPCI completely.[17] Twenty-eight percent of the episodes in the analytical sample were initiated by SNFs that stopped participating in the clinical episode.

[17] By the end of the five-year initiative, approximately 45% of Model 3 SNFs that had ever participated in BPCI (regardless of whether they were in the analytical sample) withdrew completely from the initiative and no longer participated in any clinical episodes.

Exhibit 18: Characteristics of the BPCI Providers Included in the BPCI Impact Estimates, Model 3 SNF, Q4 2013 – Q3 2018

Clinical Episodes	BPCI SNFs (N)	BPCI Episodes (N)	Average Length of Participation (Quarters) [a]	SNFs that Stopped Participating in at least one Clinical Episode (N)	Proportion of Episodes from SNFs that Stopped Participating in the Clinical Episode (%)
Model 3 SNFs Overall	493	45,694	10	335	27.9%

Note: Model 3 SNFs Overall represents SNF-initiated episodes in the analytical sample in any of the 11 clinical episodes that had sufficient volume for risk adjustment. The analytical sample includes 66% of the episodes initiated in the 11 clinical episodes and 49% of all BPCI Model 3 SNF-initiated episodes. The number of BPCI SNFs that stopped participating in the clinical episode represents unique SNFs in the analytical sample across the 11 clinical episodes. Average length of participation and the proportion of episodes from SNFs that stopped participating in the clinical episode are calculated as an average of all SNF/clinical episode combinations in the analytical sample across the 11 clinical episodes. SNF = skilled nursing facility.

[a] The average length of participation varies because providers and other organizations that volunteered to participate in BPCI could enter into the risk-bearing phase of the initiative during a two-year period through September 2015, and they could enter additional clinical episodes through December 2015. Providers could stop participating in a given clinical episode at quarterly intervals or terminate their participation in the initiative completely at any time with 60-days advance notice.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI providers.

3. Payment, Utilization, and Quality Outcomes

This section presents the BPCI impact estimates for payments, utilization, and quality for episodes initiated at BPCI SNFs over the course of the five-year initiative (Q4 2013 through Q3 2018). We present the total standardized allowed payment amount from the start of the episode through 90 days, as well as payments for SNF, IRF, and HHA services and readmissions from the episode start through 90 days. We present the payments during this period for Model 3 episodes because this is the period in which participants have more control over the delivery of care and, therefore, may have greater ability to influence changes in outcomes; the episode does not include the inpatient stay for Model 3. We present results for Model 3 SNF-initiated episodes overall below; detailed results by clinical episode are located in **Appendix G.**

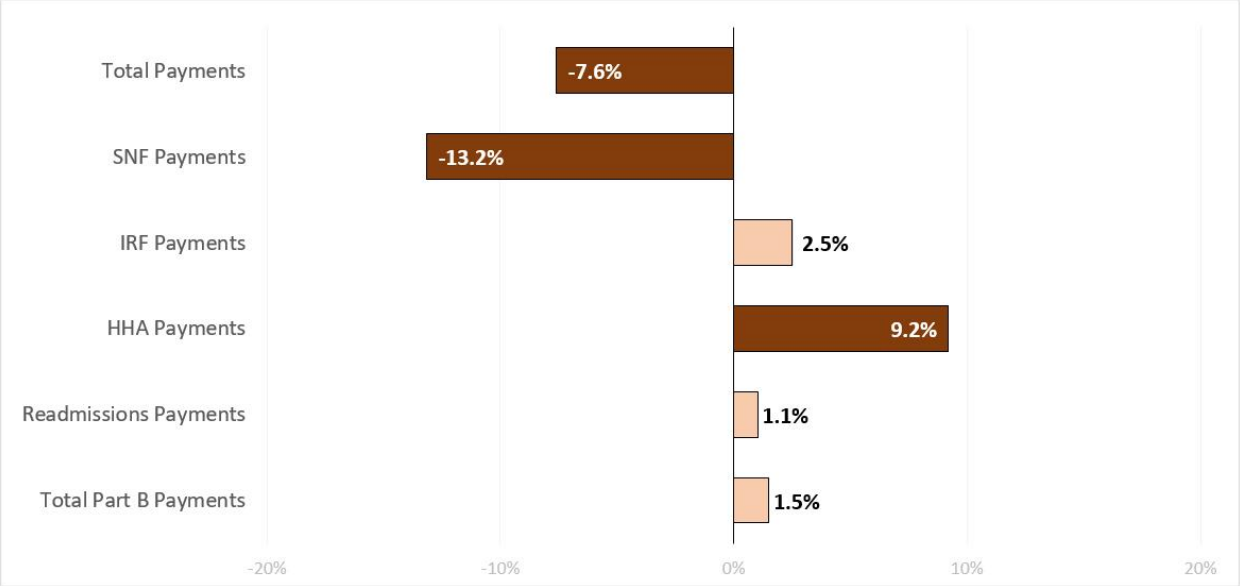
a. How have the average standardized payments changed under BPCI?

By the end of the five-year BPCI initiative, the total payment amount declined from baseline to the intervention period for BPCI Model 3 SNF-initiated episodes more than the comparison group by an estimated 7.6% ($2,171, $p<0.05$) of what payments would have been absent BPCI (Exhibits 19 and 20).

This was driven by a relative decline in SNF payments, which declined 13.2% ($2,387, $p<0.05$) for BPCI Model 3 SNF episodes relative to the comparison group. While there was a relative decline in SNF payments, HHA payments increased 9.2% ($170, $p<0.05$) for BPCI episodes relative to the comparison group.

These changes in payments provide insight into how SNFs responded to the incentives of the BPCI initiative, namely by decreasing SNF payments and increasing payments for home health services, which are less expensive on average.

Exhibit 19: Percent Change in Standardized Allowed Payment Outcomes from Episode Start through 90 Days, Model 3 SNFs, Baseline to Intervention, Q4 2013 – Q3 2018 [a]

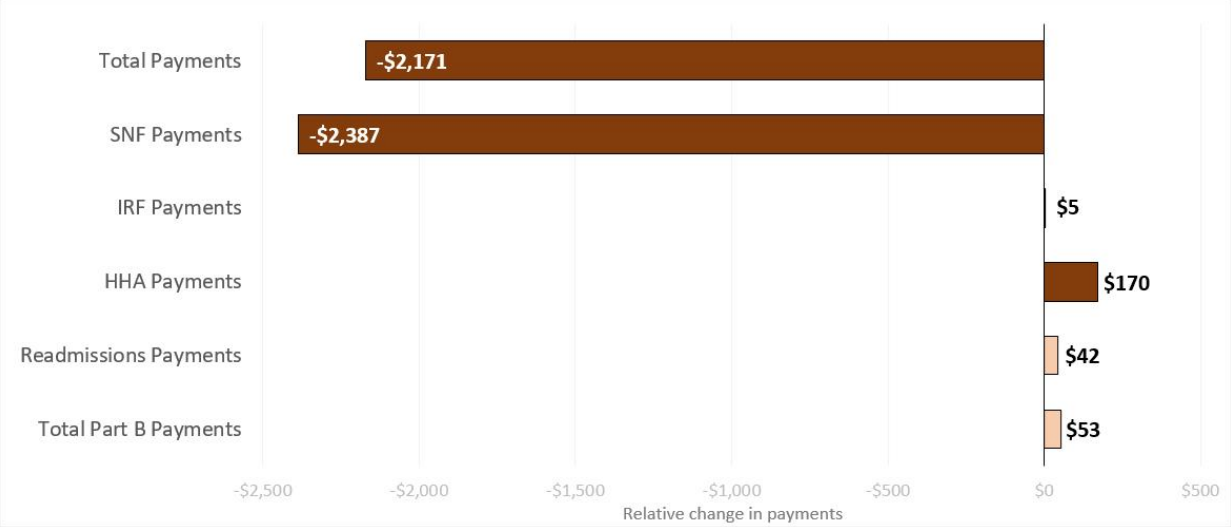


Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 11 clinical episodes that had sufficient volume to allow for risk-adjustment. The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is calculated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

[a] Total Payments includes Part A and B payments from the episode start through 90 days. All other outcomes include payments from the episode start through 90 days. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

Exhibit 20: Impact of BPCI on Standardized Allowed Payment Outcomes from Episode Start through 90 Days, Model 3 SNFs, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 11 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

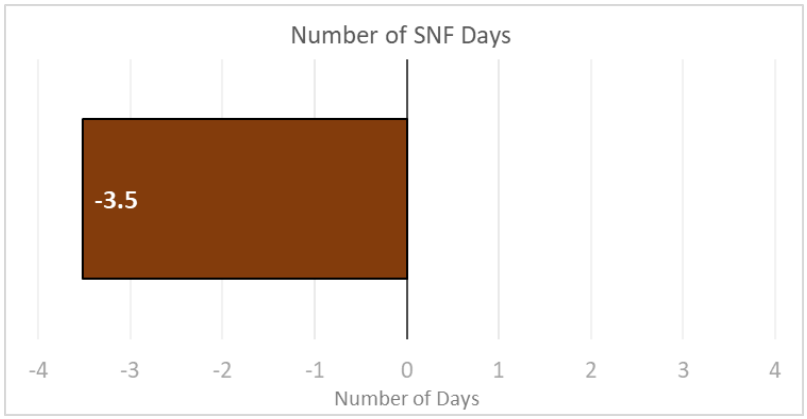
[a] Total Payments includes Part A and B payments from the episode start through 90 days. All other outcomes include payments from the episode start through 90 days. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

b. How have the services changed under BPCI?

The changes in SNF use were consistent with changes in payments. We observed a decline in the number of SNF days in the 90-day PDP of 3.5 days ($p<0.05$) for BPCI Model 3 SNF episodes relative to the comparison group (Exhibit 21).

Exhibit 21: Impact of BPCI on PAC Utilization, Model 3 SNFs, Baseline to Intervention, Q4 2013 – Q3 2018



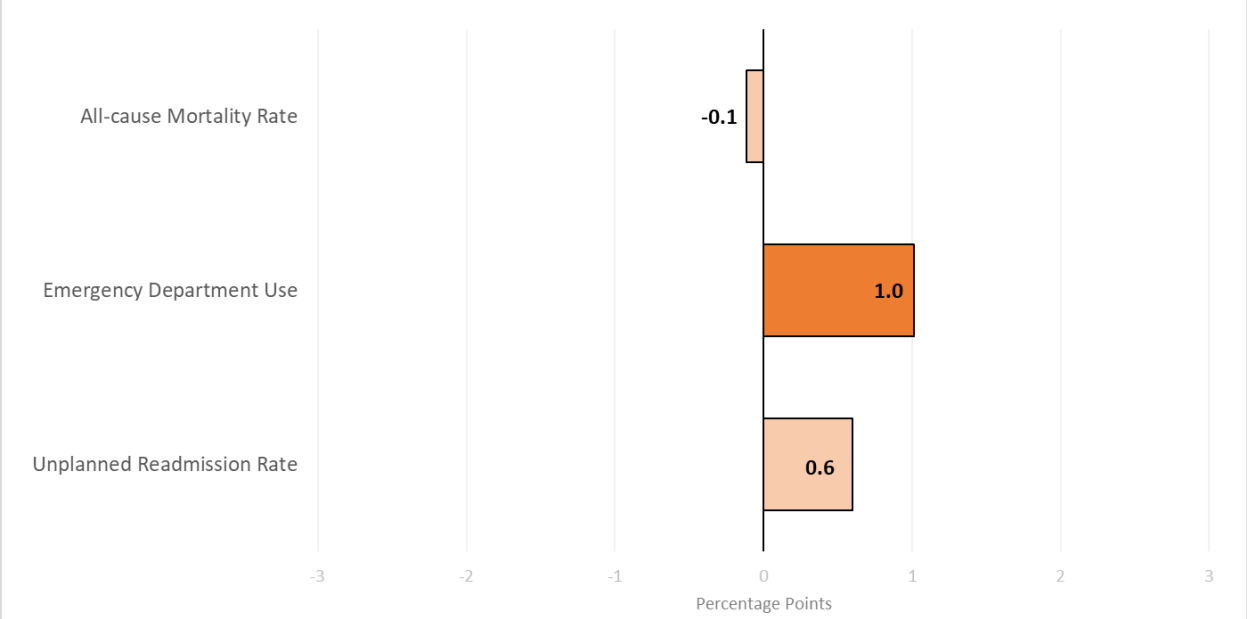
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 11 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. PAC = post-acute care; SNF = skilled nursing facility.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

c. How has quality of care changed under BPCI?

In aggregate, quality of care generally did not change under BPCI Model 3 for SNF-initiated episodes. There were no statistically significant changes at the 5% level in the mortality rate, ED use, or unplanned readmission rate from episode start through 90 days for Model 3 SNF-initiated episodes relative to the comparison group (Exhibit 22). However, for three out of the 11 clinical episodes analyzed, there was a statistically significant relative increase in one of the quality measures, potentially indicating a decrease in quality. There was a statistically significant increase in the 90-day mortality rate for chronic obstructive pulmonary disease (COPD), bronchitis, and asthma episodes relative to the comparison group. There was a statistically significant increase in 90-day ED use for other respiratory episodes. And there was a statistically significant increase in the 90-day unplanned readmission rate for stroke episodes. However, these three clinical episodes suffer from small sample sizes, both overall and per SNF, making it difficult to ascertain whether these results are a signal or simply noise. See **Appendix G** for clinical episode-level results.

Exhibit 22: Impact of BPCI on Quality Outcomes from the Episode Start through 90 days, Model 3 SNFs, Baseline to Intervention, Q4 2013 – Q3 2018



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 11clinical episodes that had sufficient volume to allow for risk-adjustment. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant. SNF = skilled nursing facility.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

B. Impact of BPCI on HHA-initiated episodes

1. Key Findings

- **Impact on Payments**
  - During the five years of the BPCI initiative, the total standardized allowed payment amount from episode start through 90 days declined by 7.5% (or $981) for Model 3 HHA-initiated episodes relative to the comparison group.
  - The IRF payment amount decreased by 29.2% (or $106) for HHA-initiated episodes relative to the comparison group. This decline was driven by a decrease in the share of episodes with any IRF payment.
- **Impact on Post-Acute Care Utilization**
  - There was no statistically significant relative change in the number of HHA visits for HHA-initiated episodes relative to the comparison group.
- **Impact on Quality**
  - There was a relative decrease of 2.0 percentage points in the readmission rate for HHA-initiated episodes relative to the comparison group.

2. Sample Characteristics

We analyzed the characteristics of the 116 HHAs that voluntarily participated in BPCI Model 3 and the 71 HHAs for which we were able to identify comparison groups for the impact estimates (referred to as "matched HHAs").[18] The 71 matched HHAs were similar to the group of all BPCI Model 3 HHAs based on key characteristics as described in Exhibits 23a-23b, but these 116 BPCI HHAs were different from non-participating HHAs. A higher proportion of participating HHAs were part of a chain (73% vs. 32%) and for-profit (81% vs. 76%). BPCI-participating HHAs had more employed nurses on average than did non-participating HHAs (29 vs. 9), although the BPCI average is driven by one large HHA that had 1,558 nurses. BPCI-participating HHAs also had a greater number of BPCI-eligible patients for their MS-DRGs during 2011 (374 vs. 101).[19] (See **Appendix D** for further details about these measures and **Appendix E** of the Year 5 Evaluation & Monitoring Annual Report for more details and additional sample characteristics.)

[18] **Appendix C** provides further information on the methods for identifying comparison groups for the analysis of the impact estimates.

[19] After excluding the large HHA, BPCI participating HHAs employed an average of 16 nurses and had an average of 257 admissions for BPCI episode MS-DRGs during 2011.

Exhibits 23a and 23b: Baseline Characteristics of All BPCI-participating HHA and Non-participating HHAs, Model 3

Domain	Characteristic	Matched BPCI HHAs (N)	Matched BPCI HHAs (%)	BPCI HHAs (N)	BPCI HHAs (%)	Non-participating HHAs (N)	Non-participating HHAs (%)
Ownership	For Profit	54	74%	94	81%	7,458	76%
	Government	0	0%	0	0%	612	6%
	Non-Profit	19	26%	22	19%	1,699	17%
Urban/Rural	Rural	13	18%	25	22%	1,886	19%
	Urban	60	82%	91	78%	7,883	81%
Part of Chain	Yes	50	68%	85	73%	3,110	32%

Characteristic	Matched BPCI HHAs (mean)	BPCI HHAs (mean)	Non-participating HHAs (mean)
Number of Employed Nurses in HHA	41	29	9
Number of Admissions for BPCI Episode MS-DRGs, 2011	534	374	101

Note: Data from a total of 116 BPCI HHA episode initiators and 9,769 non-participating HHAs are included in this exhibit. HHA = home health agency; MS-DRG = Medicare Severity-Diagnosis Related Group.

Source: Lewin analysis of 2013 Provider of Service (POS) files and 2011 Medicare claims. BPCI-participating HHAs are defined as HHAs participating in Model 3. Non-participating HHAs are all other HHAs not participating in any BPCI initiative that reported values for all measures listed in the table above.

Exhibit 24 describes the sample of Model 3 HHAs included in the analysis, across all three clinical episodes with sufficient volume for risk adjustment. The BPCI HHAs included in the impact analysis initiated 15,050 episodes during the five-year initiative and participated for an average of eight quarters. By the end of the initiative, 57 (80.3%) Model 3 HHAs in the analytical sample stopped participating in at least one clinical episode, and 46 (64.8%) terminated their participation in BPCI completely.[20] Eighteen percent of the episodes in the analytical sample were initiated by HHAs that stopped participating in the clinical episode.

[20] By the end of the five-year initiative, approximately 63% of Model 3 HHAs that had ever participated in BPCI (regardless of whether they were in the analytical sample) withdrew completely from the initiative and no longer participated in any clinical episodes

Exhibit 24: Characteristics of the BPCI Providers Included in the BPCI Impact Estimates, Model 3 HHA, Q4 2013 – Q3 2018

Clinical Episodes	BPCI HHAs (N)	BPCI Episodes (N)	Average Length of Participation (Quarters) [a]	HHAs that Stopped Participating in at least one Clinical Episode (N)	Proportion of Episodes from HHAs that Stopped Participating in the Clinical Episode (%)
Model 3 HHAs Overall	71	15,050	8	57	17.6%

Note: Model 3 HHAs Overall represents HHA-initiated episodes in the analytical sample in any of the three clinical episodes that had sufficient volume for risk adjustment. The analytical sample includes 89% of the episodes initiated in the 3 clinical episodes and 55% of all BPCI Model 3 HHA-initiated episodes. The number of BPCI HHAs that stopped participating in the clinical episode represents unique HHAs in the analytical sample across the three clinical episodes. Average length of participation and the proportion of episodes from HHAs that stopped participating in the clinical episode are calculated as an average of all HHA/clinical episode combinations in the analytical sample across the clinical episodes. HHA = home health agency.

[a] The average length of participation varies because providers and other organizations that volunteered to participate in BPCI could enter into the risk-bearing phase of the initiative during a two-year period through September 2015, and they could enter additional clinical episodes through December 2015. Providers could stop participating in a given clinical episode at quarterly intervals or terminate their participation in the initiative completely at any time with 60-days advance notice.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI providers.

3. Payment, Utilization, and Quality Outcomes

This section presents the BPCI impact estimates for payments, utilization, and quality for episodes initiated at BPCI HHAs over the course of the five-year initiative (Q4 2013 through Q3 2018). We present the total standardized allowed payment amount from the start of the episode through 90 days, as well as payments for SNF, IRF, and HHA services and readmissions from the episode start through 90 days. We present the payments during this period for Model 3 episodes because this is the period in which participants have more control over the delivery of care and, therefore, may have greater ability to influence changes in outcomes; the episode does not include the inpatient stay for Model 3. We present results for Model 3 HHA-initiated episodes overall below; detailed results by clinical episode are located in **Appendix H.**

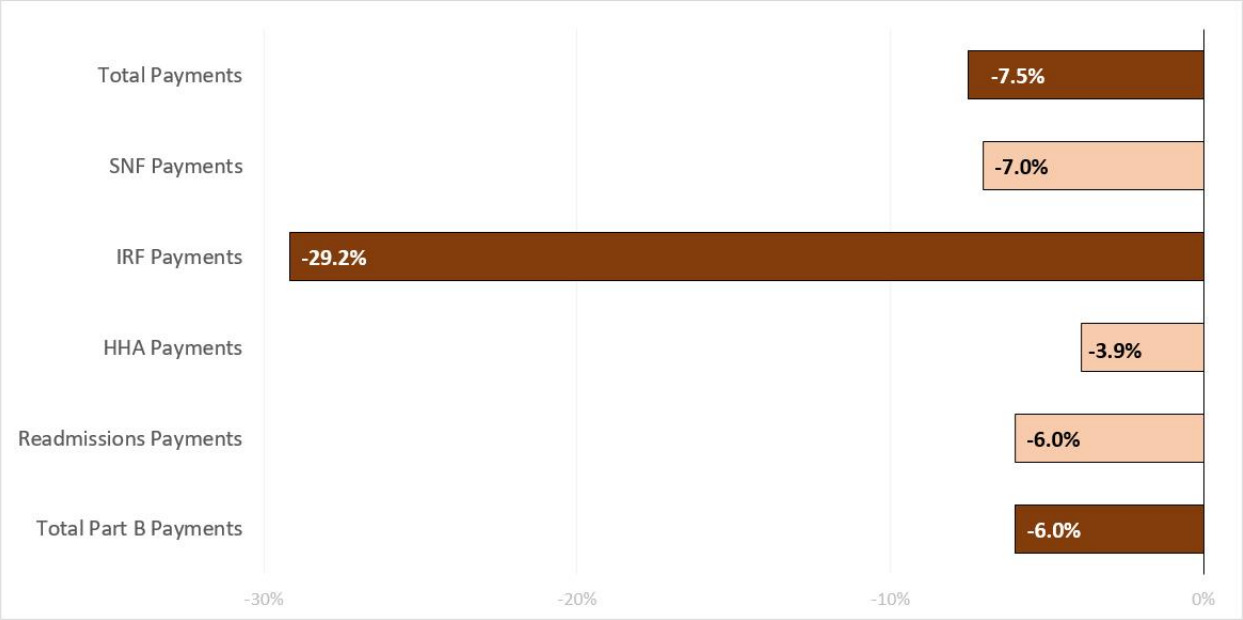
a. How have the average standardized payments changed under BPCI?

By the end of the five-year BPCI initiative, the total payment amount declined from baseline to the intervention period for BPCI Model 3 HHA-initiated episodes more than the comparison group by an estimated 7.5% ($981, $p<0.05$) of what payments would have been absent BPCI (Exhibits 25 and 26).

This relative decline in total payments was associated with relative declines in all other payment outcomes presented, but only the declines in IRF and Part B payments were statistically significant. The decline in IRF payments (of 29.2% or $106, $p<0.05$) was driven by a small but statistically significant decrease in the share of episodes with any IRF use during the intervention period for

BPCI episodes relative to the comparison group.[21] Part B payments declined 6.0% ($203, p<0.05) for BPCI episodes relative to the comparison group.

Exhibit 25: Percent Change in Standardized Allowed Payment Outcomes from Episode Start through 90 Days, Model 3 HHAs, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



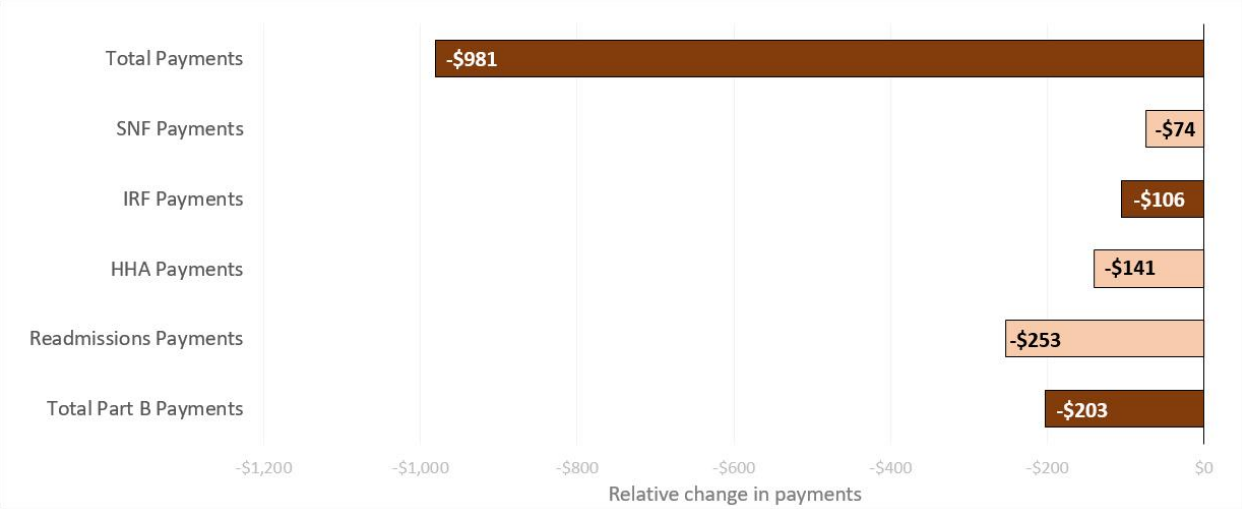
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the three clinical episodes that had sufficient volume to allow for risk-adjustment. The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is calculated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

[a] Total Payments includes Part A and B payments from the episode start through 90 days. All other outcomes include payments from the episode start through 90 days. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

[21] Approximately 1% to 2% of episodes had an IRF payment. There was a 0.5 percentage point decrease in the proportion of BPCI Model 3 HHA episodes with IRF relative to the comparison group. Because IRF is more expensive than the other services that comprise the total payment for HHA-initiated episodes, even a small decrease in IRF utilization had a detectable impact on the reduction in total payments overall.

Exhibit 26: Impact of BPCI on Standardized Allowed Payment Outcomes from Episode Start through 90 Days, Model 3 HHAs, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 3 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

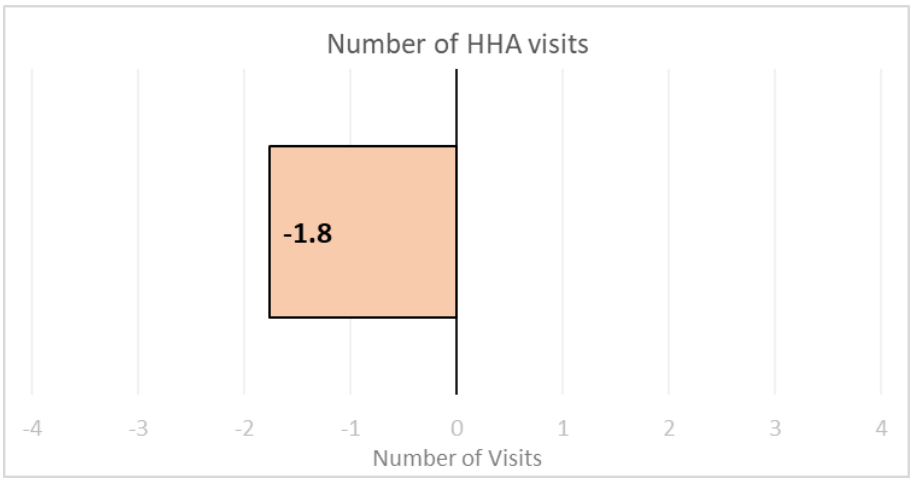
[a] Total Payments includes Part A and B payments from the episode start through 90 days. All other outcomes include payments from the episode start through 90 days. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

b. How have the services changed under BPCI?

As with payments, there was a decline in the average number of HHA visits for BPCI Model 3 HHA-initiated episodes relative to the comparison group of 1.8 visits, but the result was not statistically significant (Exhibit 27).

Exhibit 27: Impact of BPCI on PAC Utilization, Model 3 HHAs, Baseline to Intervention, Q4 2013 – Q3 2018



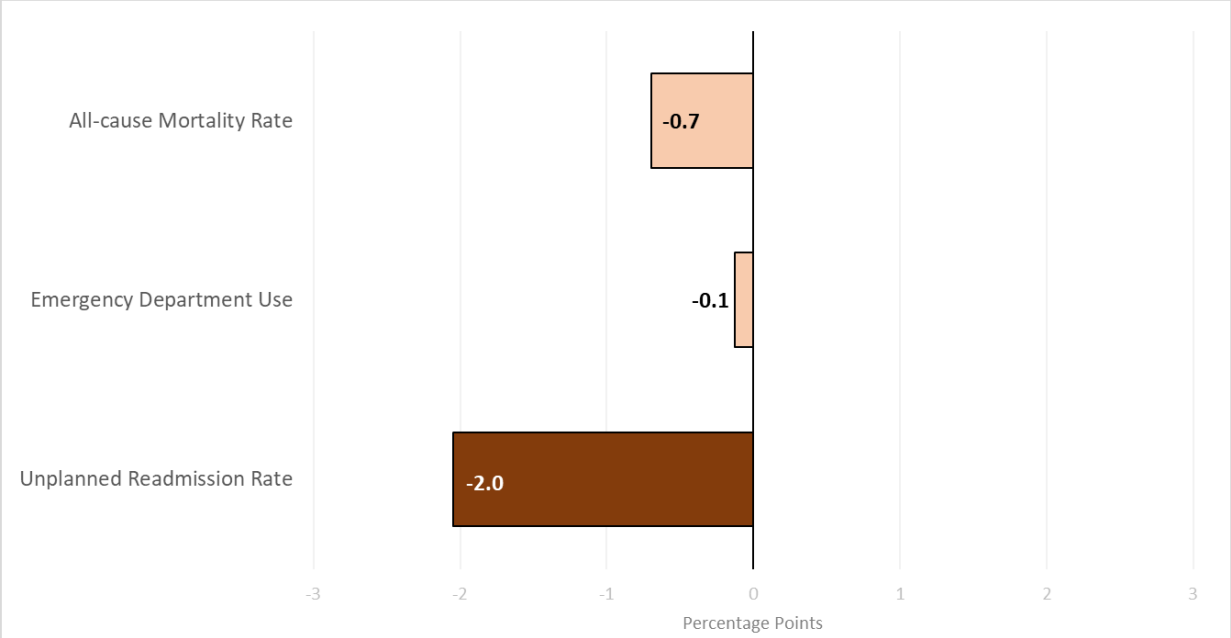
Note: The estimate in this exhibit is the result of a difference-in-differences (DiD) model and include the 3 clinical episodes that had sufficient volume to allow for risk-adjustment. Light orange bars indicate DiD estimates that are not statistically significant. PAC = post-acute care; HHA = home health agency.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

c. How has quality of care changed under BPCI?

Quality of care was not adversely affected under BPCI Model 3 for HHA-initiated episodes (Exhibit 28). There was a 2.0 percentage point reduction in readmissions from the episode start through 90 days ($p<0.05$), which was driven by a statistically significant decrease in readmissions for congestive heart failure episodes. Readmissions decreased by 1.8 percentage points for MJRLE episodes, but the data from the baseline period shows BPCI and matched comparison providers were not on parallel trends for readmissions for MJRLE episodes, which is required for an unbiased estimate and therefore this result is biased. There were no changes in mortality or ED use. Since the results predominantly show no change in the measures of quality, they suggest that there was no consistent impact on quality, but BPCI Model 3 HHA episodes were not associated with a worsening of care for beneficiaries.

Exhibit 28: Impact of BPCI on Quality Outcomes from the Episode Start through 90 days, Model 3 HHAs, Baseline to Intervention, Q4 2013 – Q3 2018



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the 3 clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. HHA = home health agency.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

IV. Model 4 Results

Model 4 differs from Model 2 and Model 3 in several key ways that resulted in different impacts on payment, utilization, and quality outcomes. Under Model 4, a participating episode initiator was an acute-care hospital, and the episode started when the beneficiary was admitted to the participating hospital for the chosen clinical episode. However, the bundle included only the inpatient hospital stay and any readmissions and associated professional services that occurred within 30 days of hospital discharge that were not explicitly excluded from the bundle, and it did not include PAC services. The admitting hospital was paid a prospectively determined amount, and the hospital in turn paid the providers that furnished services included in the episode.

Model 4 was chosen by few hospitals and experienced substantial withdrawals. From a high of 23 hospitals, only two remained at the end of the five-year initiative. Three hospitals switched participation from Model 4 to Model 2. Participants indicated that several BPCI design features made it difficult for them to succeed and contributed to their decision to withdraw, including difficulties in developing and maintaining the infrastructure required to pay participating providers under the prospective payment design. Interviewees also reported selecting Model 2 over Model 4 because Model 2 included the PAC stay in the bundle, and therefore offered the greatest opportunity to achieve savings. On the other hand, those that entered under Model 4 expressed doubts that they would be able to influence PAC providers sufficiently to improve care coordination and management, and interviewees cited the ability to mitigate risks as a reason for participating under Model 4, instead of Model 2.

A. Key Findings

- **Impact on Payments**
  - By the end of the five-year BPCI initiative, there was no change in the total standardized allowed payments during the inpatient stay plus 30 days post discharge or 90 days post discharge for BPCI Model 4 hospital-initiated episodes relative to a comparison group.
  - Readmission payments decreased by an estimated 29.1% (or $202) for BPCI Model 4 episodes relative to a comparison group during the 30 days post discharge, the time period included in the bundle definition for Model 4. However, there was no change in readmission payments during the 90-day PDP.
  - HHA payments on the other hand, experienced a strengthening of the impact during the longer, 90-day PDP. HHA payments during the 30-day PDP increased by an estimated 17.6% (or $188, $p<0.10$) for BPCI Model 4 episodes, relative to a comparison group, while they increased by an estimated 19.3% (or $303, $p<0.05$) during the 90-day PDP.
- **Impact on Post-Acute Care Utilization**
  - There was an increase for BPCI Model 4 episodes relative to the comparison group of 7.1 percentage points in the proportion of patients discharged to PAC settings, services which were not included in the Model 4 bundle definition.

- There was no statistically significant change in the percent of beneficiaries discharged to an institutional PAC setting, among those discharged to any PAC setting. Similarly, there was no change in the average number of SNF days among those who had any SNF days.

- **Impact on Quality**
  - In general, claims-based quality measures did not change for Model 4 hospital-initiated episodes relative to the comparison group.

B. Sample Characteristics

We analyzed the characteristics of the 23 hospitals that voluntarily participated in BPCI Model 4. These 23 hospitals were different from non-participating hospitals based on key characteristics, as described in Exhibits 29a-29b. In regard to ownership, 70% of BPCI-participating hospitals in Model 4 were non-profit compared to 57% of non-participants. A lower proportion of Model 4 participating hospitals were part of a chain (43%) than non-participating hospitals (53%). Nearly all Model 4 BPCI-participating hospitals were located in urban areas (91%), compared with 69% of non-participating hospitals. The average bed count for participating hospitals was more than double that of non-participants (405 vs. 175). Participating hospitals had a higher average resident-to-bed ratio than non-participating hospitals (0.14 vs. 0.05) and over twice as many admissions during 2011 for MS-DRGs included in BPCI episodes (3,460 vs. 1,598). (See **Appendix D** for further details about these measures.)

Exhibits 29a and 29b: Baseline Characteristics of All BPCI-participating Hospitals and Non-participating Hospitals, Model 4

Domain	Characteristic	BPCI Hospitals (N)	BPCI Hospitals (%)	Non-participating Hospitals (N)	Non-participating Hospitals (%)
Ownership	For Profit	6	26%	638	23%
	Government	1	4%	542	20%
	Non-Profit	16	70%	1,594	57%
Urban/Rural	Rural	2	9%	872	31%
	Urban	21	91%	1,902	69%
Part of Chain	Yes	10	43%	1,469	53%

Characteristic	BPCI Hospitals (mean)	Non-participating Hospitals (mean)
Bed Count	405	175
Number of Admissions for BPCI Episode MS-DRGs, 2011	3,460	1,598
Medicare Days Percent	31%	42%
Resident-bed ratio	0.14	0.05
Disproportionate Share Percent	29%	29%
Hospital Market Share	21%	26%

Note: Data from 23 BPCI hospitals and 2,774 non-participating hospitals. MS-DRG=Medicare Severity-Diagnosis Related Group.

Source: Lewin analysis of 2013 Provider of Service (POS) files and 2011 Medicare claims. BPCI participating hospitals are defined as hospitals participating in Model 4. Non-participating hospitals are all other hospitals not participating in any BPCI initiative that reported values for all measures listed above and are not from Maryland.

Of the 23 BPCI-participating hospitals, we were able to identify comparison hospitals for 17 hospitals in the analysis of the impact estimates. Exhibit 30 describes the sample of Model 4 hospitals included in the analysis, with characteristics for Model 4 hospitals across all clinical episodes. The BPCI hospitals included in the impact analysis initiated 8,246 episodes during the five-year initiative and participated for an average of seven quarters. By the end of the initiative, only 2 of the 17 Model 4 hospitals in the analytical sample were still participating in at least one clinical episode; the other 15 (88.2%) terminated their participation in BPCI completely.[22] Fifty-eight percent of the episodes in the analytical sample were initiated by the hospitals that stopped participating in the clinical episode.

Exhibit 30: Characteristics of the BPCI Providers included in the BPCI Impact Estimates, Model 4 Hospitals, Q4 2013 – Q3 2018

Clinical Episodes	BPCI Hospitals (N)	BPCI Episodes (N)	Average length of participation (Quarters) [a]	Hospitals that Stopped Participating in at least one Clinical Episode (N)	Proportion of Episodes from Hospitals that Stopped Participating in the Clinical Episode (%)
Model 4 Hospitals Overall	17	8,246	7	16	58.3%

Note: Model 4 Hospitals Overall represents hospital-initiated episodes in the analytical sample in either of the two clinical episodes that had sufficient volume for risk adjustment. The analytical sample includes 93% of the episodes initiated in the two clinical episodes and 52% of all BPCI Model 4 hospital-initiated episodes. The number of BPCI hospitals that stopped participating in the clinical episode represents unique hospitals in the analytical sample across the two clinical episodes. Average length of participation and the proportion of episodes from hospitals that stopped participating in the clinical episode are calculated as an average of all hospital/clinical episode combinations in the analytical sample across the two clinical episodes.

[a] The average length of participation varies because providers and other organizations that volunteered to participate in BPCI could enter into the risk-bearing phase of the initiative during a two-year period through September 2015, and they could enter additional clinical episodes through December 2015. Providers could stop participating in a given clinical episode at quarterly intervals or terminate their participation in the initiative completely at any time with 60-days advance notice.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI providers.

C. Payment, Utilization, and Quality Outcomes

This section presents the BPCI impact estimates for payment, utilization, and quality outcomes for episodes initiated by Model 4 participating acute-care hospitals over the course of the five-year initiative (Q4 2013 through Q3 2018). We present the total standardized allowed payment amount for the inpatient stay plus 30 days post discharge and for the inpatient stay plus 90 days post discharge. We also present the standardized allowed payment amount for SNF, IRF, and HHA services and for readmissions during both the 30 and 90 days post discharge. While the Model 4

[22] By the end of the five-year initiative, these two Model 4 hospitals were the only hospitals still participating at the end of the initiative out of all 23 Model 4 hospitals that had ever participated in BPCI (regardless of whether they were in the analytical sample).

episode included the inpatient hospital stay and readmissions and associated professional services that occurred within 30 days of discharge, we present results for both 30 and 90 days in order to be consistent with Model 2 and Model 3, and because the impact on some outcomes may not be detected during the shorter, 30-day period. Detailed results of the BPCI impact estimates by clinical episode for the two clinical episodes with sufficient sample size for analysis are located in **Appendix I**.

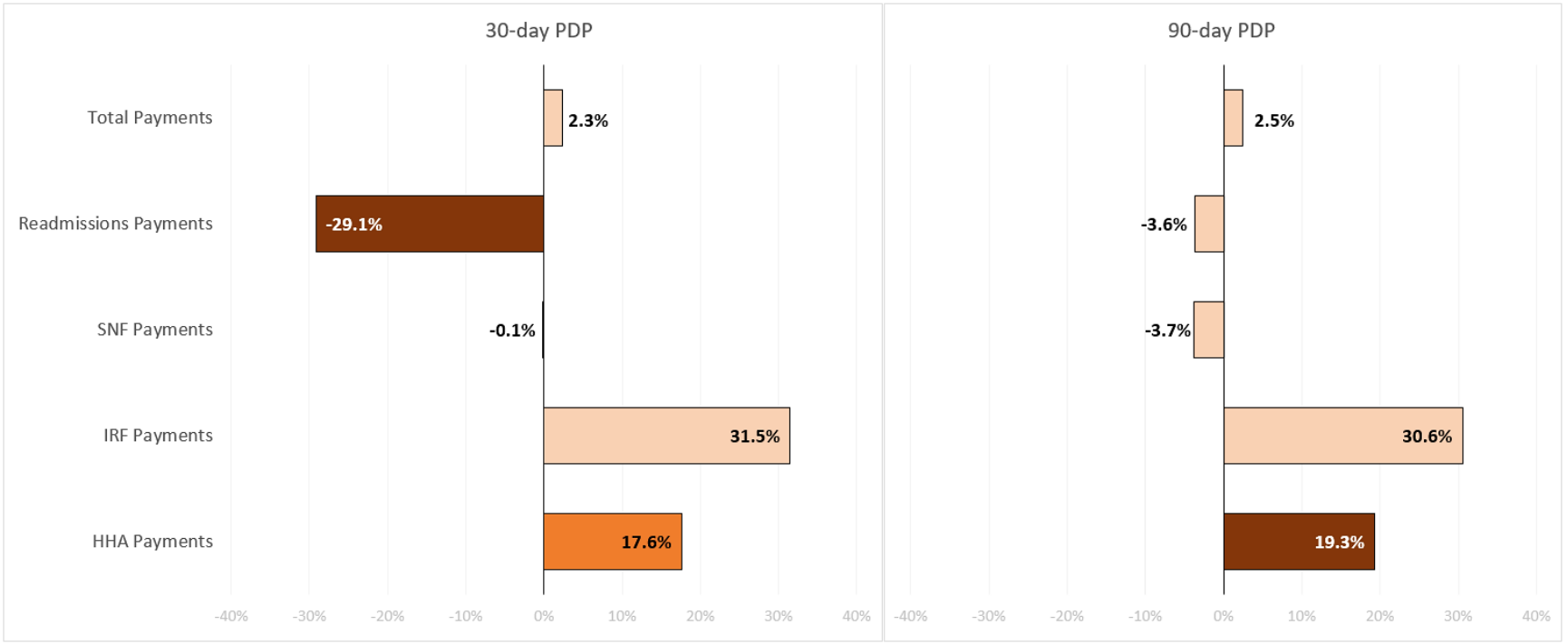
1. How have the average standardized payments changed under BPCI?

By the end of the five-year BPCI initiative, there were no statistically significant changes at the 5% level in total payments for the inpatient stay plus 30 days post discharge or 90 days post discharge for BPCI Model 4 hospital-initiated episodes relative to the comparison group. Total payments during the inpatient stay plus 30 days increased 2.3% of what payments would have been absent BPCI ($579) from baseline to the intervention period for BPCI Model 4 episodes relative to comparison group episodes, and total payments during the inpatient stay plus 90 days increased 2.5% ($728) (Exhibits 31 and 32).

Readmission payments during the 30 days post discharge decreased by 29.1% ($202, $p<0.05$). However, the change in readmission payments during the 90-day PDP was not statistically significant. Furthermore, there were no statistically significant changes in SNF or IRF payments during either the 30 or 90-day PDP. There was, however, some evidence of an increase in HHA payments of 17.6% ($188, $p<0.10$) during the 30-day PDP. This increase grew to 19.3% ($303, $p<0.05$) during the 90-day PDP.

The results provide some evidence that there were changes in payments for BPCI episodes that were consistent with the intended effect of BPCI Model 4, which was to reduce total payments by targeting readmissions. However, these payment reductions were offset by PAC payments, which were not part of the episode, and the effect of reduced readmission payments did not endure beyond the 30-day PDP.

Exhibit 31: Percent Change in Standardized Allowed Payment Amount, Inpatient Stay and 30- and 90-day PDPs, Model 4 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018 [a]

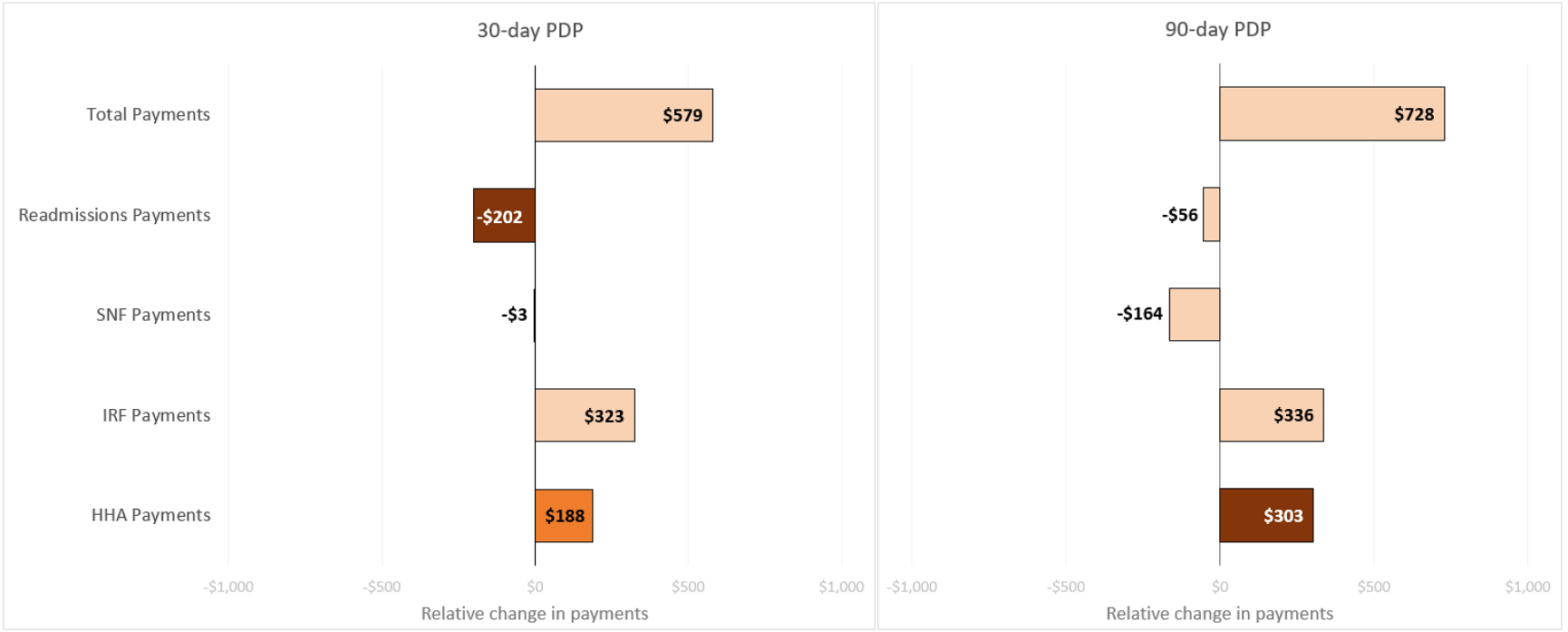


Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the two clinical episodes that had sufficient volume to allow for risk-adjustment. The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is calculated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant. PDP = post-discharge period; SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

[a] Total Payments includes Part A and B payments during the inpatient stay and 30- and 90-day PDP. All other outcomes include payments during the 30- and 90-day PDP. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

Exhibit 32: Impact of BPCI on Standardized Allowed Payment Amount, Inpatient Stay and 30- and 90-day PDPs, Model 4 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018 [a]



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the two clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant. PDP = post-discharge period; SNF = skilled nursing facility; IRF = inpatient rehabilitation facility; HHA = home health agency.

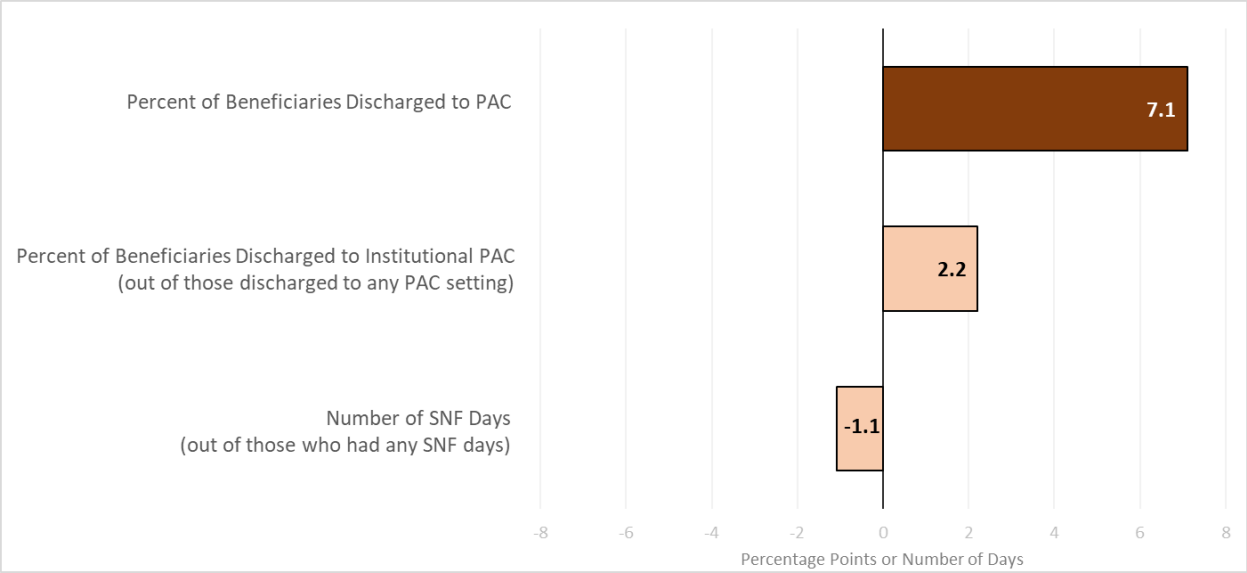
[a] Total Payments includes Part A and B payments during the inpatient stay and 30- and 90-day PDP. All other outcomes include payments during the 30- and 90-day PDP. Payment measures are not conditional upon the use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

2. How have the services changed under BPCI?

The changes in service use were consistent with the changes in payments, namely an increase in PAC use overall and no reduction in institutional PAC use, services which were not included in the Model 4 bundle definition. There was a 7.1 percentage point ($p<0.05$) increase in the share of patients discharged to any PAC setting for BPCI Model 4 hospital episodes relative to the comparison group, and no change for the share of patients discharged to institutional PAC facilities among those discharged to PAC settings. There was also no relative change in the average number of SNF days during the 90-day PDP, among episodes with at least one day in a SNF (Exhibit 33).

Exhibit 33: Impact of BPCI on PAC Utilization Outcomes, Model 4 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018



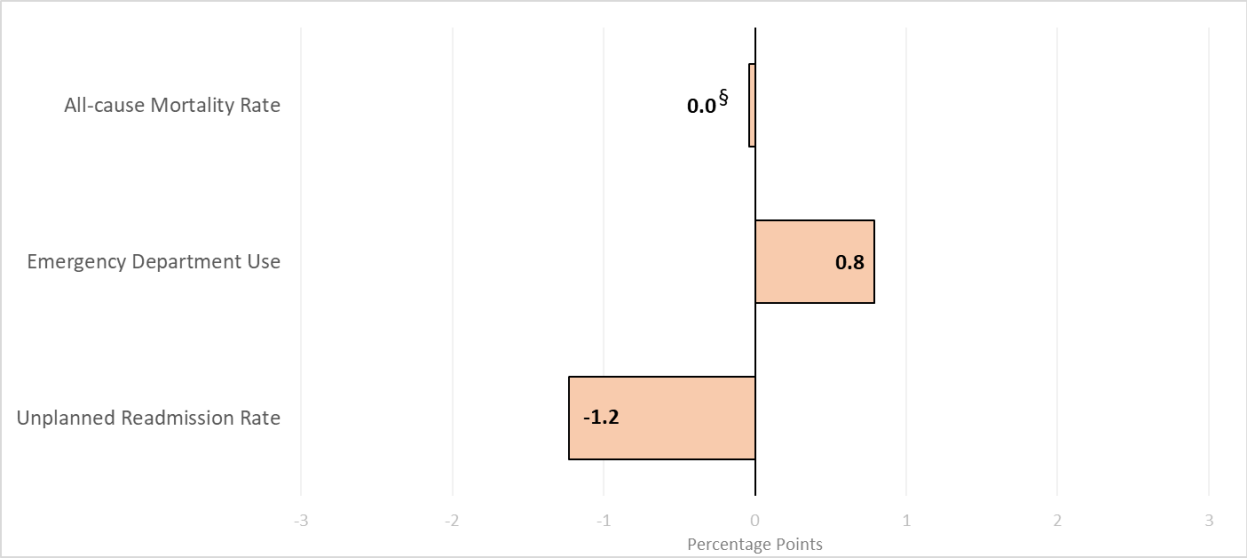
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the two clinical episodes that had sufficient volume to allow for risk-adjustment. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. PAC = post-acute care; SNF = skilled nursing facility.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

3. How has quality of care changed under BPCI?

Under BPCI, the incentive to lower episode payments could lead to changes in service use that lower the quality of care. We examined three key claims-based measures during the 90-day PDP to assess changes in quality and found no statistically significant changes in quality for Model 4 episodes. The results suggest that BPCI Model 4 was not associated with a worsening of care for beneficiaries (Exhibit 34).

Exhibit 34: Impact of BPCI on Quality Outcomes, 90-day PDP, Model 4 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2018



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and include the two clinical episodes that had sufficient volume to allow for risk-adjustment. Light orange bars indicate DiD estimates that are not statistically significant. PDP = post-discharge period.

§ Data from the baseline period shows BPCI and matched comparison episodes were not on parallel trends for this outcome, which is required for an unbiased estimate.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers.

V. Medicare Program Savings

As documented above, Medicare fee-for-service (FFS) payments decreased under BPCI Model 2 and Model 3. However, payment reductions may not translate into net savings to Medicare because they do not account for the reconciliation payments that Medicare paid to or recovered from participants.

The first section of this chapter presents estimates of the net change to spending by the Medicare program after accounting for reconciliation payments made to or received from BPCI participants. Even though the model was intended to save 2% to 3%, we found that BPCI resulted in net losses for the Medicare Program. To better understand this, the second section presents an analysis of the Medicare FFS spending reductions that are calculated with the use of the model benchmarks and that are estimated by the evaluation.

A. Estimates of Net Savings to Medicare

1. Key Findings

- **Model 2:**
  - After accounting for reconciliation payments, BPCI Model 2 resulted in an estimated net loss to Medicare of 1.3% of what payments would have been absent BPCI ($418 million, $p<0.05$), ranging from a loss of 0.4% to 2.1% ($133 million to $702 million).
  - Had CMS not eliminated downside risk and had required participants to return funds when payments were above the target price (benchmark price with 2% to 3% CMS discount applied), reconciliation payments would have been lower, and Medicare would have realized no change in net spending for Model 2 after accounting for reconciliation payments (0.0% in net spending, or $2 million), ranging from a loss of 0.9% ($283 million) to a savings of 0.9% ($286 million).
  - Both surgical and medical clinical episodes initiated by Model 2 hospitals were associated with an estimated net loss to Medicare, even if CMS had not eliminated downside risk, though the estimates were not statistically significant at the 5% level. Surgical clinical episodes initiated by Model 2 PGPs, however, would have been associated with an estimated net savings to Medicare, had CMS not eliminated downside risk, of 2.5% ($159 million, $p<0.05$), ranging from a savings of 0.9% to 4.1% ($59 million to $259 million). Medical clinical episodes initiated by Model 2 PGPs would have been associated with an estimated net loss to Medicare even if CMS had not eliminated downside risk, though the estimate was not statistically significant at the 5% level.
- **Model 3:**
  - After accounting for reconciliation payments through the end of the five-year initiative, BPCI Model 3 resulted in an estimated net loss to Medicare of 3.1% of what payments would have been absent BPCI ($110 million, $p<0.05$), ranging from a loss of 1.6% to 4.7% ($55 million to $165 million).

- Had CMS not eliminated downside risk, reconciliation payments would have been lower, and the estimated net loss to Medicare would have been 1.9% ($66 million, $p<0.05$), ranging from a loss of 0.3% to 3.4% ($11 million to $120 million).

2. Methods

Net Medicare savings for both Model 2 and Model 3 were defined as the difference between the change in aggregate non-standardized paid amounts[23] and reconciliation payments made to or received from BPCI participants following the formula below:

Net Medicare savings = change in aggregate non-standardized paid amounts – reconciliation payments

The change in aggregate non-standardized paid amounts is calculated by multiplying the estimates from the DiD model by a standardized to non-standardized conversion factor. For Model 2, the DiD model estimates the change in per-episode standardized paid amounts during the inpatient stay and 90-day PDP. For Model 3, the DiD model estimates the change in per-episode standardized paid amounts from the PAC admission through 90 days. For each Model, the per-episode change in standardized payments was converted to non-standardized payments and multiplied by the total number of BPCI episodes.[24] We also present the estimated net savings to Medicare, the change in non-standardized paid amounts, and reconciliation payments as a percentage of what episode payments would have been absent BPCI, which is calculated as the BPCI baseline payment amount plus the change in the episode payment amount for the comparison group from baseline to intervention. See **Appendix C** for additional details on the definitions and calculations for each component.

Because CMS eliminated downside risk for some episodes over the course of the BPCI initiative, we present estimated net savings to Medicare in two ways. We present results as the initiative was implemented, in which CMS eliminated repayment responsibility, and we present results as the initiative was initially designed, in which CMS had not retrospectively eliminated repayment responsibility.

We present results for Model 2 overall and Model 3 overall both with and without downside risk eliminated. Our overall Model 2 estimates are based on 1,260,141 BPCI Model 2 episodes, and our Model 3 estimates are based on 154,106 Model 3 episodes. Together, Model 2 and Model 3 represent over 99% of all BPCI episodes. We also present results for Model 2 hospitals and Model 2 PGPs by surgical and medical clinical episodes with downside risk eliminated. The Model 2 hospital estimates are based on 655,461 episodes and the Model 2 PGPs estimates are based on

[23] Non-standardized paid amounts vary from the standardized allowed amounts that we use in the DiD analyses. We use non-standardized paid amounts for this analysis, which approximate the payments made from Medicare to providers incorporating geographic and other payment adjustments and excluding beneficiary cost sharing. We use standardized allowed amounts in the DiD analyses—amounts that exclude payment adjustments and include beneficiary cost sharing—in order to isolate the impact of BPCI on Medicare payments.

[24] The number of BPCI episodes used to estimate the net savings to Medicare does not necessarily match the number of episodes in the analytical sample used for the impact estimates because it includes all clinical episodes and provider types, as well as all providers and all eligible episodes, whether or not they met the additional criteria for inclusion in the impact estimates. Thus, our method provides a projection of results of the entire model.

604,680 episodes.[25] All reported confidence intervals use the 95% confidence interval. The clinical episode level estimates, also calculated with downside risk eliminated, are shown in **Appendix J**.

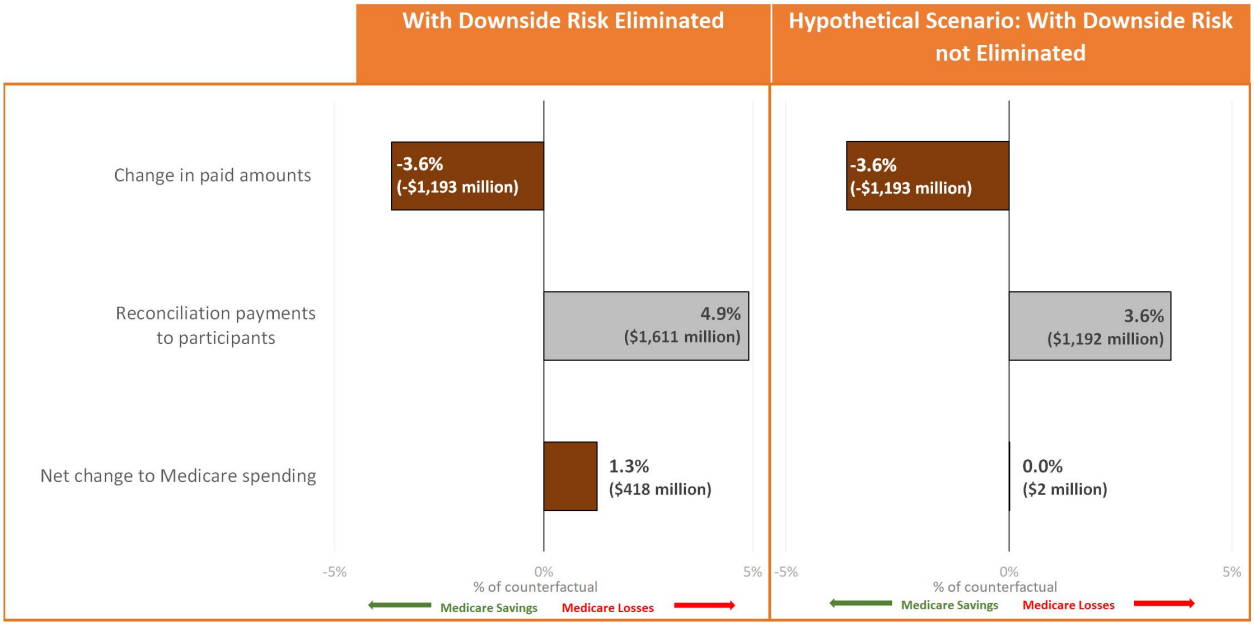
3. Results

Model 2

During the five-year BPCI initiative, Model 2 was associated with an estimated net loss to Medicare of 1.3% of what payments would have been absent BPCI ($418 million, $p<0.05$), ranging from a loss of 0.4% to 2.1% ($133 million to $702 million, 95% confidence interval) (Exhibit 35). The net loss is equivalent to a per-episode loss to Medicare of $332. While aggregate non-standardized paid amounts declined an estimated 3.6% ($1,193 million, $p<0.05$), reconciliation payments for Model 2 were 4.9% ($1,611 million), which were larger than the reductions achieved in aggregate payments.

If CMS had not eliminated downside risk for some episodes, Medicare would have realized no change in net spending after accounting for reconciliation payments, essentially breaking even. In this scenario, reconciliation payments would have fallen from 4.9% to 3.6% ($1,192 million). Subtracting this amount from the reduction in aggregate non-standardized paid amounts results in an estimated change in net Medicare spending of 0.0% ($2 million), ranging from a loss of 0.9% ($283 million) to a savings of 0.9% ($286 million) (Exhibit 35).

Exhibit 35: Estimated Change in Net Medicare Spending with and without Downside Risk Eliminated, Model 2, Q4 2013 – Q3 2018



Note: The estimates of the change in aggregate non-standardized paid amounts are from a difference-in-differences (DiD) model of standardized paid amounts during the qualifying inpatient stay and 90-day post-discharge period that were converted to non-standardized paid amounts using a standardized to non-standardized ratio. "With downside risk eliminated" depicts estimates of net savings to Medicare as the model was implemented in which CMS did not require participants to repay all funds. "Hypothetical scenario: With downside risk not eliminated" depicts estimates of net

[25] Note that the episode counts listed here may differ from those reported in the impact estimates due to differences in the source data. See **Appendix D** for more information about the samples.

savings to Medicare in the hypothetical scenario that the model was implemented as designed (i.e., repayments to Medicare were collected throughout the entire intervention period). Net savings to Medicare is the difference between the change in aggregate non-standardized paid amounts and reconciliation payments. The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is calculated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. Grey indicates values are not an estimate and do not have a standard error.

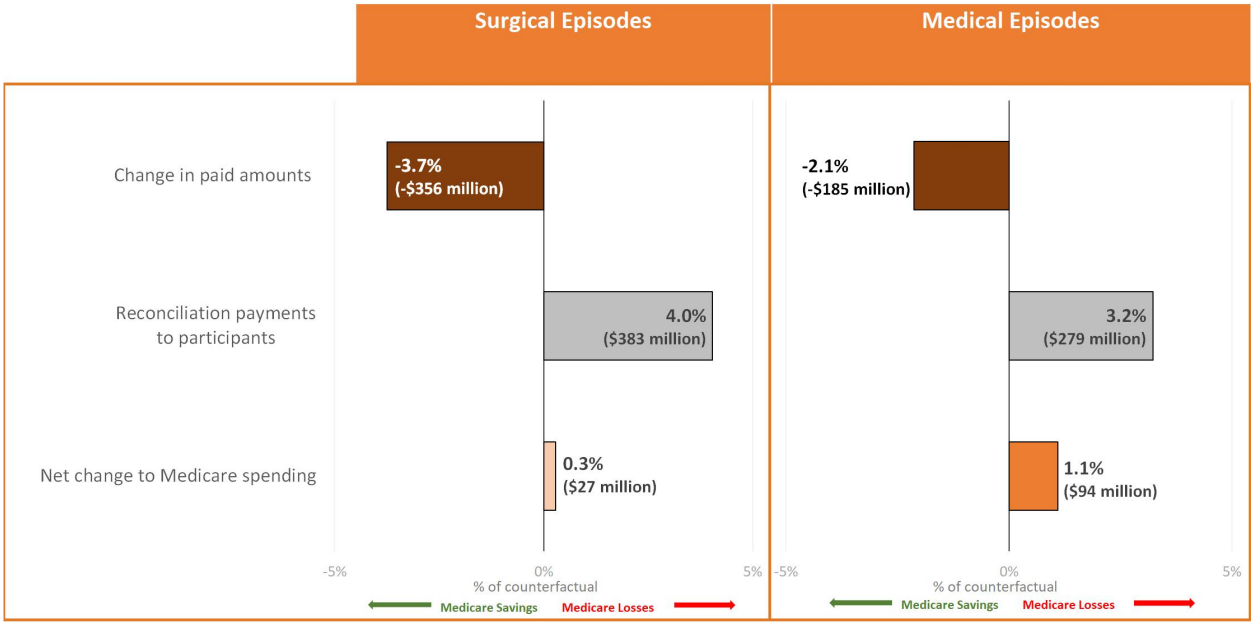
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers and CMS data on reconciliation payments.

a. Hospitals

In this section, we present the estimates of the net change to Medicare spending for Model 2 hospital surgical and medical clinical episodes assuming downside risk had not been eliminated. Surgical clinical episodes were associated with an estimated net loss to Medicare of 0.3% ($27 million), ranging from a savings of 1.0% ($91 million) to a loss of 1.5% ($145 million). Medical clinical episodes were associated with a slightly larger estimated net loss to Medicare of 1.1% ($94 million), ranging from a savings of 0.1% ($8 million) to a loss of 2.3% ($197 million) (Exhibit 36).

The change in net Medicare spending varied greatly by clinical episode. Clinical episodes that showed statistically significant losses per episode include two medical clinical episodes, sepsis and urinary tract infection, and two surgical clinical episodes, cardiac valve and percutaneous coronary intervention. No clinical episodes showed statistically significant estimates of net savings. (Detailed results are located in **Appendix J**).

Exhibit 36: Change in Net Medicare Spending, by Surgical and Medical Clinical Episodes, Model 2 Hospitals, Q4 2013 – Q3 2018



Note: The estimates of the change in aggregate non-standardized paid amounts are from a difference-in-differences (DiD) model of standardized paid amounts during the qualifying inpatient stay and 90-day post-discharge period that were converted to non-standardized paid amounts using a standardized to non-standardized ratio. Net savings to Medicare is the difference between the change in aggregate non-standardized paid amounts and reconciliation payments.

The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is estimated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Bright orange bars indicate DiD estimates are statistically significant at the 10% level. Light orange bars indicate DiD estimates that are not statistically significant. Grey indicates values are not an estimate and do not have a standard error.

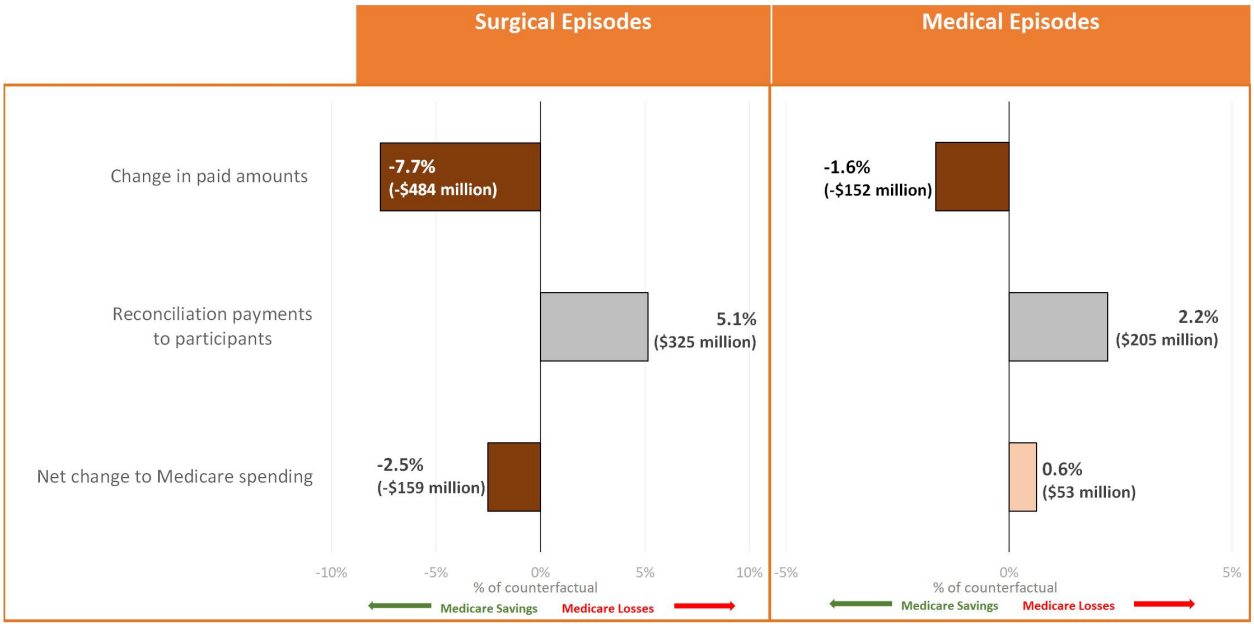
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers and CMS data on reconciliation payments.

b. PGPs

In this section, we present the estimates of the change in net Medicare spending for Model 2 PGP surgical and medical clinical episodes assuming downside risk had not been eliminated. Surgical clinical episodes were associated with an estimated net savings to Medicare of 2.5% ($159 million, $p<0.05$), ranging from a savings of 0.9% to 4.1% ($59 million to $259 million). Medical clinical episodes were associated with an estimated net loss to Medicare of 0.6% ($53 million), ranging from a savings of 0.9% ($82 million) to a loss of 2.0% ($188 million) (Exhibit 37).

The change in net Medicare spending varied greatly by clinical episode. Clinical episodes that showed statistically significant net savings per episode include major joint replacement of the lower extremity, gastrointestinal hemorrhage, and medical non-infectious orthopedic, while sepsis showed losses. (Detailed results are located in **Appendix J**.)

Exhibit 37: Change in Net Medicare Spending, by Surgical and Medical Clinical Episodes, Model 2 PGPs, Q4 2013 – Q3 2018



Note: The estimates of the change in aggregate non-standardized paid amounts are from a difference-in-differences (DiD) model of standardized paid amounts during the qualifying inpatient stay and 90-day post-discharge period that were converted to non-standardized paid amounts using a standardized to non-standardized ratio. Net savings to Medicare is the difference between the change in aggregate non-standardized paid amounts and reconciliation payments. The estimates are presented as a percentage of what episode payments would have been absent BPCI, which is estimated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at

the 5% level. Light orange bars indicate DiD estimates that are not statistically significant. Grey indicates values are not an estimate and do not have a standard error.

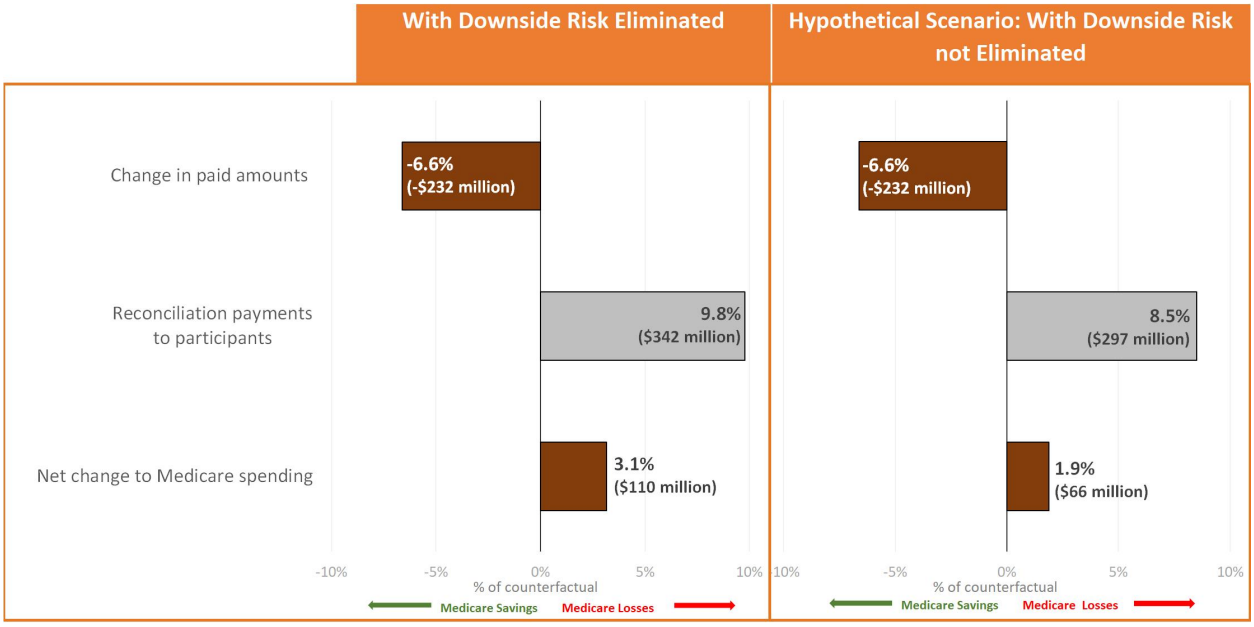
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers and CMS data on reconciliation payments.

Model 3

Model 3 was associated with an estimated net loss to Medicare of 3.1% ($110 million, $p<0.05$), ranging from a loss of 1.6% to 4.7% ($55 million to $165 million). This is equivalent to a per-episode net loss to Medicare of $714. While the estimated decline in aggregate non-standardized paid amounts was 6.6% ($232 million, $p<0.05$), reconciliation payments were 9.8% ($342 million), which far exceeded the estimated decline in Medicare payments (Exhibit 38).

If CMS had not eliminated downside risk, net losses to Medicare would have been smaller. In this hypothetical scenario, reconciliation payments would have fallen from 9.8% to 8.5% ($297 million). Subtracting this amount from the reduction in aggregate non-standardized paid amounts results in a net estimated loss to Medicare of 1.9% ($66 million, $p<0.05$), ranging from a loss of 0.3% to 3.4% ($11 million to $120 million) (Exhibit 38). In this scenario, the net loss is equivalent to a per-episode loss to Medicare of $426. (Results by clinical episode are located in **Appendix J**.)

Exhibit 38: Estimated Change in Medicare Spending with and without Downside Risk Eliminated, Model 3, Q4 2013 – Q3 2018



Note: The estimates of the change in aggregate non-standardized paid amounts are from a difference-in-differences (DiD) model of standardized Medicare paid amounts during the qualifying inpatient stay and 90-day post-discharge period that were converted to non-standardized paid amounts using a standardized to non-standardized ratio. "With downside risk eliminated" depicts estimates of net savings to Medicare as the model was implemented in which CMS did not require participants to repay all funds. "Hypothetical scenario: With downside risk not eliminated" depicts estimates of net savings to Medicare in the hypothetical scenario that the model was implemented as designed (i.e., repayments to Medicare were collected throughout the entire performance period). Net savings to Medicare is the difference between the change in aggregate non-standardized payments and reconciliation payments. The estimates are

presented as a percentage of what episode payments would have been absent BPCI, which is estimated as the average BPCI baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Grey indicates values are not an estimate and do not have a standard error.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers and CMS data on reconciliation payments.

B. Comparison of Medicare FFS Spending Reductions that are Calculated with the Use of Model Benchmarks and that are Estimated by the Evaluation

This section aims to understand why Medicare realized net losses under BPCI even though the model was intended to save the Medicare Program 2% to 3% of what payments would have been absent BPCI.[26] To produce an estimate of what payments would have been absent BPCI, the model calculated a historical payment for each participant from historical allowed amounts and then used a national trend to update it to an intervention period benchmark. This represents the model estimate of what payments should have been absent BPCI, and is also referred to as the benchmark. This was then discounted to become the target price. Model 2 and Model 3 participants with intervention episode payments below their target price received the difference as reconciliation payments. Participants with intervention episode payments above their target price repaid the difference to CMS. This reconciliation process was intended to ensure that CMS achieved savings equal to 2% to 3% of the benchmark. Medicare FFS spending reductions that are calculated with the use of model benchmarks is the difference between the benchmark and average FFS episode payments in the intervention period.[27]

The evaluation, on the other hand, used a DiD approach with a retrospective, matched comparison group to produce an estimate of what payments would have been absent BPCI. The trend of the comparison group was used to update BPCI participants' baseline payments to the intervention period. This is also referred to as the counterfactual. Medicare FFS spending reductions that are estimated by the evaluation is the difference between the counterfactual and average FFS episode payments in the intervention period. The comparison group is more representative of BPCI participants than providers nationally because the comparison group was selected to have similar characteristics (e.g., ownership status, rural/urban, bed size).[28]

1. Key Findings

- Medicare FFS spending reductions that are calculated with the use of model benchmarks were larger than those estimated by the evaluation across all four Model and provider combinations studied, and the difference was statistically significant at the 5% level in three of the four.
- For Model 2 hospital-initiated episodes, the spending reduction calculated with the use of model benchmarks was 6.0%, and the evaluation estimate was 3.3% (2.3% to 4.3%, $p<0.05$). This difference of 2.7 percentage points was statistically significant.

[26] Episodes under Model 2 and 3 were discounted 2% to 3%. Episodes under Model 4 were discounted 3.25%.
[27] Note that the model framework does not account for changes in patient characteristics from the baseline to the intervention period.
[28] Note that the evaluation framework *does* account for changes in patient characteristics from the baseline to the intervention period.

- For Model 2 PGP-initiated episodes, the spending reduction calculated with the use of model benchmarks was 5.5%, and the evaluation estimate was 4.7% (3.0% to 6.5%, $p<0.05$). This difference was not statistically significant.
- For Model 3 SNF-initiated episodes, the spending reduction calculated with the use of model benchmarks was 11.9%, and the evaluation estimate was 7.6% (5.7% to 9.5%, $p<0.05$). This difference of 4.3 percentage points was statistically significant.
- For Model 3 HHA-initiated episodes, the spending reduction calculated with the use of model benchmarks was 10.2%, and the evaluation estimate was 5.8% (1.6% to 10.1%, $p<0.05$). This difference of 4.4 percentage points was statistically significant.

2. Methods

We compared Medicare FFS spending reductions that are calculated with the use of model benchmarks for Model 2 hospital, Model 2 PGP, Model 3 SNF, and Model 3 HHA episodes using data for episodes initiated during the five-year BPCI intervention (Q4 2013 through Q3 2018). A key difference in the two estimates of spending reductions is the estimate of what payments would have been absent BPCI (Exhibit 39).

Medicare FFS spending reductions that are calculated with the use of model benchmarks: The model estimate of what payments would have been absent BPCI is called the benchmark and is based on trending forward historical allowed amounts (from Q3 2009 through Q2 2012) using a nation-wide trend and for each MS-DRG/provider combination. The spending reduction as indicated by the model is the difference between the benchmark and average FFS episode payments in the intervention period. The data are based on the reconciliation contractor EI-level reconciliation reports from the entire intervention period (Q4 2013 through Q3 2018).

Medicare FFS spending reductions that are estimated by the evaluation: The evaluation estimate of what payments would have been absent BPCI is called the counterfactual and is based on a DiD estimate of the historical paid amounts using a retrospective comparison-group trend and accounting for MS-DRG, provider and patient characteristics. The spending reduction that are estimated by the evaluation is the difference between the counterfactual and average FFS episode payments in the intervention period. The data are based on Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018). (See **Appendix C** for additional information methods and definitions.)

Exhibit 39. Definition of Model and Evaluation Components included in Medicare FFS Spending Reductions

Estimate	Average FFS Episode Payments in the Intervention Period	Payments Absent BPCI
Model	• Risk-adjusted (for MS-DRG and provider characteristics) • Allowed Amounts within the bundle	• Benchmark: Historical Allowed Amounts, trended to performance year using a retrospective national trend and risk-adjusted (for MS-DRG and provider characteristics)
Evaluation	• Risk-adjusted (for MS-DRG, provider and patient characteristics) • Allowed Amounts within the bundle	• Counterfactual: Historical paid amounts within the bundle, risk-adjusted (for MS-DRG, provider and patient characteristics) and trended to performance year using a comparison-group trend

Note: MS-DRG = Medicare Severity-Diagnosis Related Group; FFS = fee-for-service.

3. Results

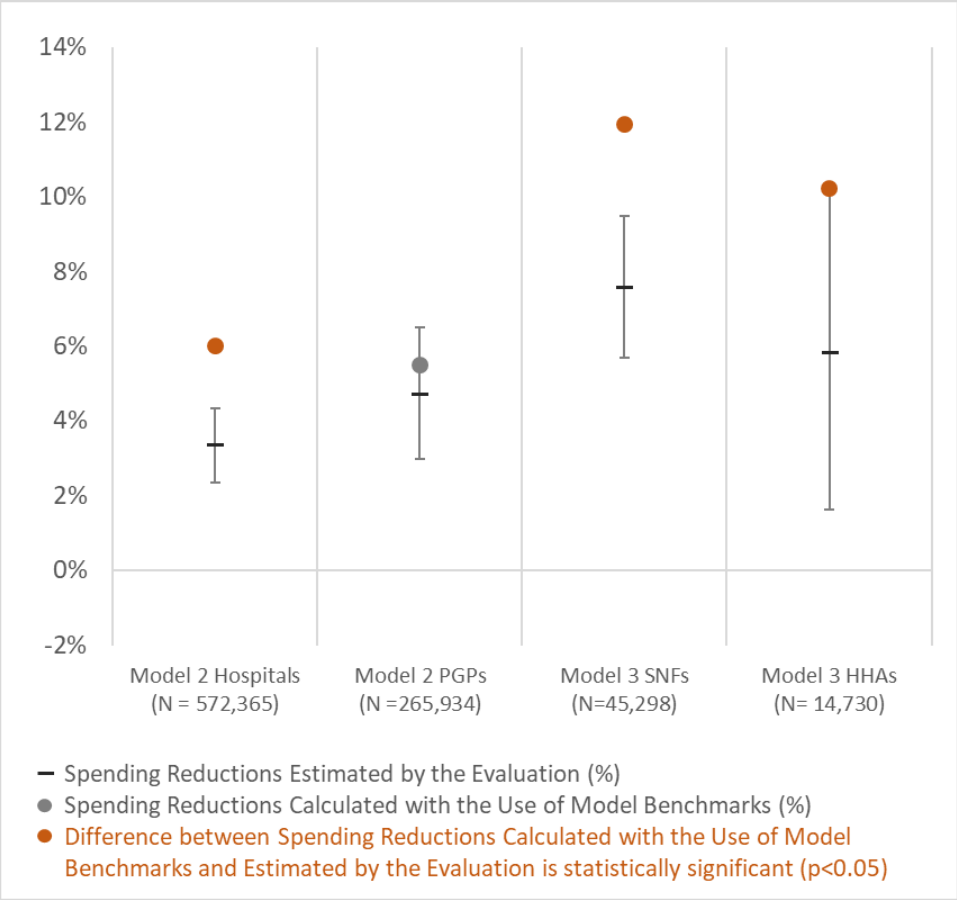
For Model 2 hospital-initiated episodes, Medicare FFS spending reductions that are calculated with the use of model benchmarks were 6.0% ($1,578), and the evaluation estimate was 3.3% ($877, $p<0.05$) ranging from 2.3% to 4.3% ($615 to $1,098, 95% confidence interval). This difference of 2.7 percentage points ($700) was statistically significant, as spending reductions that are calculated with the use of model benchmarks were greater than the upper limit of the 95% confidence interval of the evaluation estimate (Exhibit 40). Medicare FFS spending reductions that are calculated with the use of model benchmarks were higher than the evaluation estimate for all 32 clinical episodes with sufficient volume for analysis, and they were statistically significantly higher for 10 of the 32 clinical episodes. The clinical episode level estimates are shown in **Appendix K**.

For Model 2 PGP-initiated episodes, Medicare FFS spending reductions that are calculated with the use of model benchmarks were 5.5% ($1,374), and the evaluation estimate was 4.7% ($1,184, $p<0.05$), ranging from 3.0% to 6.5% ($747 to $1,622). Although spending reductions that are calculated with the use of model benchmarks were larger than the evaluation estimate, the two estimates were not statistically significantly different from each other. Medicare FFS spending reductions that are calculated with the use of model benchmarks were higher than the evaluation estimate for 14 of the 21 clinical episodes with sufficient volume for analysis. They were statistically significantly higher for two clinical episodes, and they were statistically significantly lower for two other clinical episodes.

For Model 3 SNF-initiated episodes, Medicare FFS spending reductions that are calculated with the use of model benchmarks were 11.9% ($3,100), and the evaluation estimate was 7.6% ($1,973, $p<0.05$), ranging from 5.7% to 9.5% ($1,481 to $2,465). This difference of 4.3 percentage points ($1,127) was statistically significant. Medicare FFS spending reductions that are calculated with the use of model benchmarks were higher than the evaluation estimate for all 11 clinical episodes with sufficient volume for analysis, and they were statistically significantly higher for six clinical episodes.

For Model 3 HHA-initiated episodes, Medicare FFS spending reductions that are calculated with the use of model benchmarks were 10.2% ($1,201), and the evaluation estimate was 5.85% ($688, $p<0.05$), ranging from 1.6% to 10.1% ($190 to $1,186). This difference of 4.4 percentage points ($512) was statistically significant. Medicare FFS spending reductions that are calculated with the use of model benchmarks were higher than the evaluation estimate for two of three HHA clinical episodes with sufficient volume for analysis, and they were statistically significantly higher for one clinical episode.

Exhibit 40: Medicare FFS Spending Reductions that are Calculated with the Use of Model Benchmarks and that are Estimated by the Evaluation, Q4 2013 – Q3 2018 (%)



Note: Medicare FFS spending reductions that are calculated with the use of model benchmarks are the difference between the benchmark price (participants' historical allowed amounts trended forward using a nation-wide trend and accounting for MS-DRG and provider characteristics) and average FFS episode payments in the intervention period. Medicare FFS spending reductions that are as estimated by the evaluation are based on a difference-in-differences (DiD) model and are the difference between the counterfactual (participants' historical paid amounts trended forward using a comparison-group trend and accounting for MS-DRG, provider, and patient characteristics) and average FFS episode payments (standardized allowed payment covered within the bundle) in the intervention period. Estimates are calculated as a share of the evaluation's counterfactual. The error bars around the evaluation estimate represent 95% confidence intervals. FFS = fee-for-service; PGP = physician group practice; SNF = skilled nursing facility; HHA = home health agency.

Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2018) for BPCI and comparison providers and CMS data on reconciliation payments.

VII. Discussion and Conclusion

A. Discussion

This final annual BPCI evaluation report presents updated results using data for episodes initiated through the fifth and final year of the model, which reflect an average of ten quarters of participation experience. For Model 2, Model 3, and Model 4, we estimated the impact of BPCI on payments, utilization, and quality of care. For Model 2 and Model 3, we also estimated the impact of BPCI on net Medicare savings. The final results for Model 2 and Model 3 remain consistent with earlier reports, BPCI participants responded to the initiative's incentives by reducing Medicare payments. We continue to see general patterns of reduced intensity of PAC use, with reductions in institutional care and decreases in the number of SNF days among patients who receive SNF care. Unlike Model 2 and Model 3, Model 4 BPCI participants did not reduce Medicare payments and did not reduce intensity of PAC use. As reported in the Year 3 Evaluation & Monitoring Annual Report, interviewees stated that Model 4 was selected over Model 2 because they operate in an integrated system, which meant that reducing costs in the PAC portion of the clinical episode would have resulted in reduced revenues to its SNF and HHA. Also consistent with the model's incentives, Model 4 BPCI participants did reduce readmission payments during the 30 days post discharge, the time period included in the bundle definition for Model 4. However, readmission payments did not decline during the 90-day PDP.

For the most part, the results of the evaluation indicate that BPCI did not affect quality of care. There are a few exceptions, however. For Model 2 PGP-initiated episodes there was a decrease in ED use, potentially pointing to an improvement in quality. There are a few concerning signs for Model 3 SNF-initiated episodes. Three out of 11 clinical episodes had an increase in the mortality rate, ED use, or unplanned readmissions, potentially indicating a decrease in quality. However, these three clinical episodes suffer from small sample sizes, both overall and per SNF, making it difficult to ascertain whether these results are a signal or simply noise. To obtain reliable results in future models, CMMI may want to consider setting minimum thresholds for participation.

This report examined outcomes for Model 2 separately for hospital and PGP episodes. The analysis found reductions in total payments for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes. Both types of providers had reductions in institutional PAC payments; however, hospital episodes were more likely to substitute HHA for institutional PAC, while PGP episodes tended to have reductions in both institutional PAC and HHA. Furthermore, the reduction in total payments for hospital-initiated episodes was due to both surgical and medical clinical episodes, whereas the reduction in total payments for PGP-initiated episodes was driven largely by surgical clinical episodes.

These differences may stem from the different levers available to redesign care for hospitals and PGPs. For example, hospitals and PGPs may differ in terms of the services under their control. Hospitals may have more control over inpatient care protocols, which could impact recovery and the type and duration of PAC needed. PGPs and hospitals may also have different relationships with patients. It is possible that PGPs were able to reduce home health use because they have a better understanding of the home supports available to patients.

This report also updates the estimates of the impact of BPCI on Model 4 episodes that was presented in the Year 3 Evaluation and Monitoring Annual Report.[29] In this report, we evaluated payments during the inpatient stay plus 30-day and 90-day PDP. For Model 4 episodes, BPCI participants responded to the incentives of the initiative by reducing readmission payments within the 30-day PDP. However, during the longer 90-day PDP there was no change in readmission payments and there was an increase in HHA payments, highlighting the importance of the design of financial incentives in shaping participant behavior. Further, the low participation levels in Model 4 and the high attrition rates of providers that did participate point to other model design features that providers take into account when deciding whether to participate in voluntary payment models.

The differences in the estimated impact of BPCI between Model 4 hospital-initiated episodes and Model 2 hospital-initiated episodes under surgical clinical episodes highlight how hospitals responded to the different financial incentives under Models 2 and 4. The bundle for Model 4 included only the inpatient hospital stay and readmissions and associated professional services that occurred within 30 days of hospital discharge; it excluded PAC services. Model 2, on the other hand, had the most comprehensive bundle, including the inpatient hospital stay, all concurrent professional services and post discharge services, including hospital readmissions, delivered within the chosen episode length of 30, 60, or 90 days. The results for Model 4 hospital-initiated episodes, which are based on two surgical clinical episodes, indicate that readmission payments decreased relative to the comparison group during the 30-days post discharge. This was offset by an increase in overall PAC use and PAC payments. Consequently, total payments including the inpatient stay, readmissions, and PAC use during the 30- and 90-days post discharge did not change. For Model 2 hospital-initiated surgical clinical episodes, readmission payments during the 90-day PDP did not change, and lower intensity (and lower cost) HHA services were substituted for higher intensity (and higher cost) SNF services. This resulted in a reduction in PAC payments overall, and subsequently a reduction in total payments overall. These changes in payments and utilization provide insights into how hospitals under each model responded to the initiative's incentives to reduce Medicare expenditures for a bundle of items and services while maintaining or improving quality of care.[30]

In previous reports, we examined other outcomes and subsets of the BPCI population, as described in Exhibit 41 below. Please refer to the Year 5 Evaluation & Monitoring Annual Report and published manuscripts for additional details.[31,32,33]

[29] The report is available for download from: https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

[30] We compare estimates for Model 4 hospital-initiated episodes and the estimates for Model 2 hospital-initiated surgical clinical episodes because both are based on the same types of episodes. The Model 4 estimates include two clinical episodes, major joint replacement of the lower extremity (MJRLE), which represents 86% of Model 4 episodes, and coronary artery bypass graft (CABG), which represents 14% of Model 4 episodes. The Model 2 hospital surgical clinical episode estimates include 13 clinical episodes, of which MJRLE represents 72% and CABG represents 2%. See **Appendix C** for a listing of all clinical episodes analyzed and the shares of episodes.

[31] The report is available for download from: https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

[32] Maughan et al. (2019). Medicare's Bundled Payments for Care Improvement initiative maintained quality of care for vulnerable patients. *Health Affairs*, 38(4), 561–568.

[33] Trombley et al. (2019). Association of Medicare's Bundled Payments for Care Improvement initiative with patient-reported outcomes. *Health Services Research*, 1–12.

Exhibit 41: Previous BPCI Research Findings

- Results in previous reports indicated that changes in functional status did not differ between beneficiaries in BPCI episodes and comparison beneficiaries, based on survey results, although fewer BPCI beneficiary respondents reported the highest level of satisfaction with their care.
- Quality of care was maintained among vulnerable populations, including populations dually eligible for Medicare and Medicaid, with dementia, or with recent institutional PAC use.
- BPCI had no impact on total market volume of non-fracture MJRLE procedures.
- Among Model 2 BPCI patients discharged to a SNF, the proportion who went to a SNF with a high star rating increased.
- Use of program rule waivers, beneficiary incentives, and gainsharing was limited.

In this report, we also estimated net savings to the Medicare program for Model 2 and Model 3. Consistent with previous estimates, even though BPCI resulted in reductions in episode payments, Medicare experienced net losses under Model 2 and Model 3 after accounting for reconciliation payments that Medicare paid to or recovered from participants. Medicare net losses under Model 2 and Model 3 in this report during the five-year initiative were roughly similar to the estimated net losses through September 2017 in the Year 6 Evaluation & Monitoring Annual Report. However, in this report, Model 2 and Model 3 would still have resulted in net losses under the hypothetical scenario that downside risk was not temporarily eliminated; whereas, in the earlier report, Model 2 would have achieved a small (although not statistically significant) amount of net savings under the hypothetical scenario.

This report includes an analysis comparing Medicare FFS spending reductions that are calculated with the use of model benchmarks and that are estimated by the evaluation to understand why Medicare realized net losses under BPCI even though the model was intended to achieve savings of 2% to 3.25%. As described above, during the reconciliation process, intervention spending was compared with a discounted benchmark or target price. CMS created a participant-specific benchmark by updating historical episode payments with national spending trends, and then discounted it 2% to 3% to create a target price. Model 2 and Model 3 participants with episode payments below their target price received the difference as reconciliation payments. Conversely, participants with episode payments above their target price repaid the difference to CMS. Medicare savings, therefore, depended on benchmarks accurately reflecting what episode payments would have been absent BPCI. National trends, however, may not have accurately captured the payment changes for BPCI participants that would have happened absent the BPCI initiative. This may have led to inaccurate target prices in some cases. Given the voluntary nature of the model, participants with favorable target prices would be more likely to participate and continue participation, whereas participants with unfavorable target prices would be more likely to exit, tipping the financial balance against CMS.[34] The evaluation, in contrast, uses a comparison group instead of trended historical payments to construct a counterfactual, or what spending would have been absent BPCI. A comparison group accounts for changing market and policy factors that may have affected the episode payments of BPCI participants. Further, the evaluation counterfactual incorporates risk adjustment to account for changes in patient mix from baseline to intervention period as well as

[34] Participants could stop participation in BPCI by notifying CMS.

any potential cost-shifting to services not covered by the episode; the CMS benchmark does not account for these factors.

By comparing Medicare FFS spending reductions that are calculated with the use of model benchmarks and that are estimated by the evaluation under BPCI, we found that the model benchmarks indicated a larger reduction in episode payments than that estimated by the evaluation across the four model/provider combinations studied, and the difference was statistically significant in three of the four.

Other features of the initiative also contributed to the lack of net Medicare savings. CMS eliminated downside risk during periods of the initiative to accommodate start-up challenges experienced by CMMI and participants. Episode payments had a substantial amount of variability for some clinical episodes, especially for episode initiators with small episode volume (which was more likely in Model 3). Thus, even if benchmarks were accurate on average, they were not accurate for every participant. This inherent payment variability within clinical episodes may have had adverse financial effects on CMS, again given the voluntary nature of the model. Participants with average payments above the target price (i.e., those that would have had to make repayments) were much more likely to exit, whereas participants with average payments below the target price (i.e., those likely to receive reconciliation payments) were much more likely to continue participation. This resulted in CMS paying out higher reconciliation payments than anticipated.

CMS began the BPCI Advanced initiative when the BPCI initiative ended in September 2018, and the design of the BPCI Advanced model incorporated lessons learned from the BPCI initiative. Entry and exit opportunities were scaled back under BPCI Advanced. BPCI Advanced also includes fewer clinical episodes and focuses on episodes with less payment variation and sufficient episode volume. BPCI Advanced uses a participant-specific target price by updating historical episode payments to the intervention year with spending trends of a peer group instead of national spending trends. In addition, target prices are risk-adjusted to reflect patient mix during the performance period. However, BPCI Advanced target prices are constructed with projected peer group spending trends rather than actual (i.e., retrospective) peer group spending trends. This allows for target prices to be calculated and provided to participants in advance of model deadlines so that participants know approximately what CMS intends to pay for episodes before assuming financial risk. While this gives more certainty to participants, its success hinges on reasonably accurate future trends projections. There would be financial risks to participants and CMS depending on the direction and magnitude of projection errors. Additionally, prospectively providing target prices might increase self-selection into models with voluntary participation, such as BPCI Advanced.

The target price under BPCI Advanced is intended to represent Medicare payments absent the model, after allowing for pre-determined intended savings to the Medicare program. Determining the appropriate price requires accounting for changes in medical care delivery, payment and coverage changes, and non-model factors that will affect the payments for an episode of care, which might be particularly challenging with prospective target pricing. If target prices are set too high, providers will receive too much in reconciliation payments and Medicare will not benefit from changes in care delivery as intended under BPCI Advanced. Conversely, if target prices are set too low, providers are unlikely to choose to participate or may exit BPCI Advanced when they

can, which would limit the benefits of the initiative to Medicare. Early results for BPCI Advanced indicate that retrospective target prices in the BPCI Advanced model would have lowered target prices by more than $600 million during the first two years of the model.[35] Moreover, small changes in coding practices can also make setting accurate target prices difficult. In fiscal year 2017, the coding guidelines of the International Classification of Diseases, 10th revision, included small coding changes for both congestive heart failure and pneumonia, two of the highest-volume clinical episodes in the BPCI Advanced model. As a result of these coding changes, similar patients would probably be classified as having more serious congestive heart failure and pneumonia diagnoses in the treatment period than in the baseline period. Because target prices are appropriately based on the seriousness of a patient's diagnosis, target prices increased in the BPCI Advanced model, which resulted in excess payments to participants. As a result, the model is currently on pace to lose more than $2 billion. CMS corrected the BPCI Advanced benchmarks in January 2021 to minimize losses in future years.

B. Limitations

The primary analytic approach for this evaluation is dependent on how well the comparison group represents what would have happened absent the BPCI initiative. An unbiased DiD estimate requires a matched comparison group that is similar to BPCI providers on key factors expected to influence their decision to participate in BPCI. In addition, because the DiD estimate attributes differences in trends between BPCI and the comparison group during the intervention period to the BPCI initiative, it is essential that the two groups have parallel trends for a given outcome during the baseline period. With these goals in mind, we matched providers and episodes on several factors, including payment and quality outcomes. In most combinations of Model, episode initiator type, and clinical episodes, the comparison group represented a close match to the BPCI providers on these outcomes. For some combinations, however, the comparison episodes were not as close a match as we would like, even after multiple attempts to improve the match. In some cases, we rejected the null hypothesis that there were parallel trends for key quality and total payment impact estimates tested; we rejected 73 out of 561 results, or 13% ($p<0.10$). Thus, for these estimates, the underlying assumptions of the DiD method were violated, which may bias our results for these few individual estimates. In some instances, even when we failed to reject the parallel trend hypothesis, participants initiated a small number of episodes and there were large differences in baseline outcome levels, which raises questions about whether the BPCI and matched comparison group had the same underlying trend in that outcome.[36]

Researchers have noted that matching on outcome levels in the pre-intervention period may mitigate or exacerbate bias depending on whether treatment and potential comparison providers

[35] Smith, Brad (2021). CMS Innovation Center at 10 Years — Progress and Lessons Learned. *New England Journal of Medicine*, 384(8), 759–764.

[36] For example, a high readmission rate among BPCI episodes in the baseline sample due to an extreme value could lead to a large difference in average baseline readmission rates between BPCI and comparison episodes. In this example, we would expect the differences in readmission rates to narrow during the intervention period, even absent BPCI, as the estimated average in the BPCI intervention sample converges to the long-term average rate.

are drawn from the same distribution or different distributions.[37,38,39] If treatment and potential comparison providers are drawn from the same distribution and differences in outcome levels in the pre-intervention period are due to the treatment assignment mechanism, then matching on outcome levels in the pre-intervention period would mitigate bias.[40,41] However, if treatment and potential comparison providers are drawn from different distributions, then the matched comparison group would likely revert to its mean in the intervention period, creating a biased DiD estimate.

The evidence indicates that BPCI participants and non-participants were drawn from the same distribution (see **Appendix C**, Exhibit C.11) and that differences in outcome levels in the baseline are due to the self-selection of providers into BPCI. First, BPCI was a nation-wide initiative with a large number of participants that spanned a wide range of geographies and provider types. Second, matches were found for nearly all BPCI participants within the specified calipers (see **Appendix C**, Exhibit C.12). Finally, through matching, the BPCI participants and the matched comparison providers would similarly experience reversion to the mean, making the matched comparison providers the appropriate counterfactual for BPCI participants.

BPCI hospitals were matched to non-BPCI hospitals. For PGPs, we did not have reliable data on physician affiliation to create non-BPCI PGPs, so we used a hospital-level matching approach to create a comparison group for Model 2 PGPs. Under this approach, hospitals with Model 2 PGP-initiated episodes were matched to similar non-BPCI hospitals using the same methods used to construct a comparison group for Model 2 hospitals. For Model 2 PGPs, however, this approach presented additional challenges that limited the percentage of participants included in the analysis and reduced the pool of non-participating hospitals eligible for inclusion in the comparison group. First, the Model 2 PGP analysis is restricted to episodes initiated at hospitals with enough baseline and intervention episodes to be included in the hospital-level matching for PGPs. This restriction resulted in the inclusion of 83% of Model 2 PGP participants and 57% of Model 2 PGP episodes in the clinical episodes with sufficient volume for analysis. Second, 77% of non-participating hospitals otherwise eligible for inclusion in the comparison group were exposed to BPCI through episodes initiated by Model 2 and Model 3 PGPs. To provide a large pool of eligible comparison hospitals for the Model 2 PGPs while also limiting the comparison pool's exposure to BPCI, hospitals were excluded from the PGP comparison pool if more than one percent of their patient discharges in the same clinical community were treated by physicians in BPCI PGPs (see **Appendix C** for the definition of clinical community and additional information). This exclusion eliminated between 34% and 59% of potential hospitals from inclusion in the comparison group depending on the clinical episode.

[37] Daw, J. R., & Hatfield, L. A. (2018). Matching and Regression to the Mean in Difference-in-Differences Analysis. *Health services research,* 53(6), 4138–4156.

[38] Daw, J. R., & Hatfield, L. A. (2018). Matching in Difference-in-Differences: between a Rock and a Hard Place. *Health services research*, 53(6), 4111–4117.

[39] Ryan, A. M. (2018). Well-Balanced or too Matchy-Matchy? The Controversy over Matching in Difference-in-Differences. *Health services research*, 53(6), 4111–4117.

[40] Daw, J. R., & Hatfield, L. A. (2018). Matching in Difference-in-Differences: between a Rock and a Hard Place. *Health services research*, 53(6), 4111–4117.

[41] Ryan, A. M. (2018). Well-Balanced or too Matchy-Matchy? The Controversy over Matching in Difference-in-Differences. *Health services research*, 53(6), 4111–4117.

Despite the limitations of our Model 2 PGP comparison group methodology, we believe it is the best approach given the data constraints. This belief is supported by the fact that no alternative methodologies have been presented in the peer-reviewed literature. Furthermore, BPCI Advanced benchmark pricing is based on a similar hospital-level approach.

The majority of the analyses in this report are risk-adjusted to account for differences in provider and market characteristics, as well as patient mix that is measurable with claims data. As with all regression models, however, it is possible that we did not control for all characteristics that may affect the outcomes.

As a result of the limitations summarized above, our results for some individual outcomes among specific Model, episode initiator type, and clinical episode combinations may be biased. However, our overall conclusion that BPCI has reduced episode payments while maintaining quality of care remains due to the consistency over time, across outcomes, clinical episodes, and robustness checks.

The estimate of net Medicare savings required several assumptions. First, we assumed the analytical sample was representative of all BPCI episodes. Recall that the analytical sample contains only those clinical episodes with sufficient provider and episode volume for analysis. To obtain an estimate of net Medicare savings for all BPCI episodes, we extrapolated the impact of BPCI on payments estimated using the analytical sample to all BPCI episodes. The analytical sample included 88% of BPCI Model 2 hospital-initiated episodes, 44% of Model 2 PGP-initiated episodes, 49% of Model 3 SNF-initiated episodes, and 55% of Model 3 HHA-initiated episodes. Second, we used BPCI episodes from the analytical sample to calculate a conversion rate from standardized Medicare payments to non-standardized payments, which we assumed was similar for the baseline and intervention period, as well as for BPCI and comparison episodes. Third, we assumed no change in episode volume due to BPCI.

C. Conclusion

This final annual BPCI evaluation report strengthens the evidence that bundled payments can reduce payments for multiple clinical episodes without compromising quality of care. However, after considering reconciliation payments that CMS paid to or recovered from participants under the BPCI design, Medicare spending was higher under the initiative relative to what would have been spent even under the hypothetical scenario in which downside risk had not been eliminated. These results informed the design of BPCI Advanced. BPCI Advanced features modified target prices that incorporate risk adjustment and reflect peer performance and a higher discount. Changes to the target prices were intended to encourage both high and low cost providers to participate, which would lessen the self-selection we have seen in BPCI. Some BPCI clinical episodes were not included in BPCI Advanced due to high clinical heterogeneity or small volume. In addition, the participant entry and exit opportunities were scaled back under BPCI Advanced.

Early results indicate that the BPCI Advanced Model was quick to achieve reductions in episode payments for several hospital-initiated clinical episodes without any decline in quality of care.[42] Despite this promising beginning, however, BPCI Advanced resulted in net losses to Medicare.

[42] The CMS Bundled Payments for Care Improvement Advanced Model: Year 2 Evaluation Annual Report is available for download from https://innovation.cms.gov/innovation-models/bpci-advanced.

Beginning in Model Year 4, CMS implemented significant changes to the target pricing methodology and clinical episode definition, which are intended to correct target prices that are too high and better account for selection into the model. Future BPCI Advanced evaluation reports will explore how recent changes in clinical episode definitions and target pricing methodology affect the ability of the model to achieve Medicare savings.