LEWINGROUP®

CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 6 Evaluation and Monitoring Annual Report

Prepared for:

CMS

Submitted by:

The Lewin Group

With our partners:

Abt Associates, GDIT, and Telligen

June 2020

CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 6 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, Reed Cammarota, Rebecca Cherry, Andrea Chung, Syvart Dennen, Brady Durst, Gina Gerding, Jordan Kahn, Sharon Kim, Alex Lampert, William Sheahan, Katarina Swanson, Brittani Thomas, Peter Weidner, Dean Farley, 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

EXECUTIVE SUMMARY **1**

A. Introduction 1

B. Results 1

C. Discussion and Conclusions 4

I. INTRODUCTION **5**

A. BPCI Initiative 5

B. Research Questions 8

C. Methods 8

II. MODEL 2 RESULTS **12**

A. Impact of BPCI on Hospital-initiated Episodes 12

1. Key Findings *12*

2. Sample Characteristics *12*

3. Payment, Utilization, and Quality Outcomes *14*

B. Impact of BPCI on PGP-initiated Episodes 22

1. Key Findings *22*

2. Sample Characteristics *22*

3. Payment, Utilization, and Quality Outcomes *24*

C. Comparison of Impact of BPCI among Hospitals and PGPs 31

1. Key Findings *32*

2. Methods *32*

3. Payment, Utilization, and Quality Outcomes *34*

III. NET SAVINGS TO THE MEDICARE PROGRAM **42**

A. Key Findings 43

B. Methods 43

C. Results 44

1. Model 2 ..... *44*

2. Model 3 ..... *47*

IV. DISCUSSION AND CONCLUSION **49**

A. Discussion 49

B. Limitations 51

C. Conclusion 53

List of Appendices

Appendix A: Glossary of Terms & Acronym List..... A-1

Appendix B: BPCI Clinical Episodes and Medicare Severity Diagnosis Related Groups (MS-DRGs).....B-1

Appendix C: Methods C-1

Appendix D: Supplemental Participant Characteristics D-1

Appendix E: Impact of BPCI on Payment, Utilization, and Quality Measures, by Clinical Episode and Overall, Baseline to Intervention, Model 2 Hospitals.....E-1

Appendix F: Impact of BPCI on Allowed Payment, Utilization, and Quality Measures, by Clinical Episode and Overall, Baseline to Intervention, Model 2 PGPs F-1

Appendix G: Impact of BPCI on Payment, Utilization, and Quality Measures, by Surgical and Medical Clinical Episodes, Baseline to Intervention, Model 2 Hospital and PGPs G-1

Appendix H: Net Savings to the Medicare Program by Clinical Episode..... H-1

Appendix I: Comparison Group Standardized Difference Tables.....I-1

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 Centers for Medicare & Medicaid Services (CMS), through the Center for Medicare & Medicaid Innovation (CMMI), launched the risk-bearing phase of the BPCI initiative in 2013, under the authority of section 1115A of the Social Security Act.

BPCI Awardees, which may be hospitals, physician group practices (PGPs), post-acute care (PAC) providers, or other entities that convene health care organizations, entered into voluntary agreements with CMS to be held accountable for total Medicare episode payments. Those agreements also specified participants' choices among three payment Models, 48 clinical episodes, three episode lengths, and three risk tracks. Awardees also submitted BPCI implementation protocols that specified care redesign interventions and whether 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 several key design options implicitly recognizes the variability across health care markets, providers, and episodes of care. The resulting diversity in responses and impacts provides CMMI with information on the approaches that show the most promise in achieving payment reductions while maintaining or improving quality.

This report provides an update on Medicare claims-based impact estimates for key outcomes for the top 10 Model 2 hospital and PGP clinical episodes by volume through the fourth year of the BPCI initiative. Overall, our results are consistent with previous reports that demonstrated the initiative has resulted in reductions to episode payments, which were driven by declines in institutional PAC use and payments, while maintaining quality of care.[1] The report also includes an updated analysis of the estimate of net savings to the Medicare program. Although BPCI reduced episode payments, this reduction did not translate into overall savings to Medicare after taking into account reconciliation payments made to participants, a result consistent with the previous report. The final evaluation report will include impact estimates for Models 2, 3, and 4 and estimated savings to the Medicare program over the entire five years of the initiative.

B. Results

BPCI Model 2 accounted for nearly 90% of the more than one million episodes initiated in Models 2, 3, and 4 during the first four years of the initiative, spanning October 2013 (Q4 2013) through September 2017 (Q3 2017). The top 10 clinical episodes with the largest volume included in this report represent 78% of all episodes initiated by Model 2 hospitals and 74% of all episodes

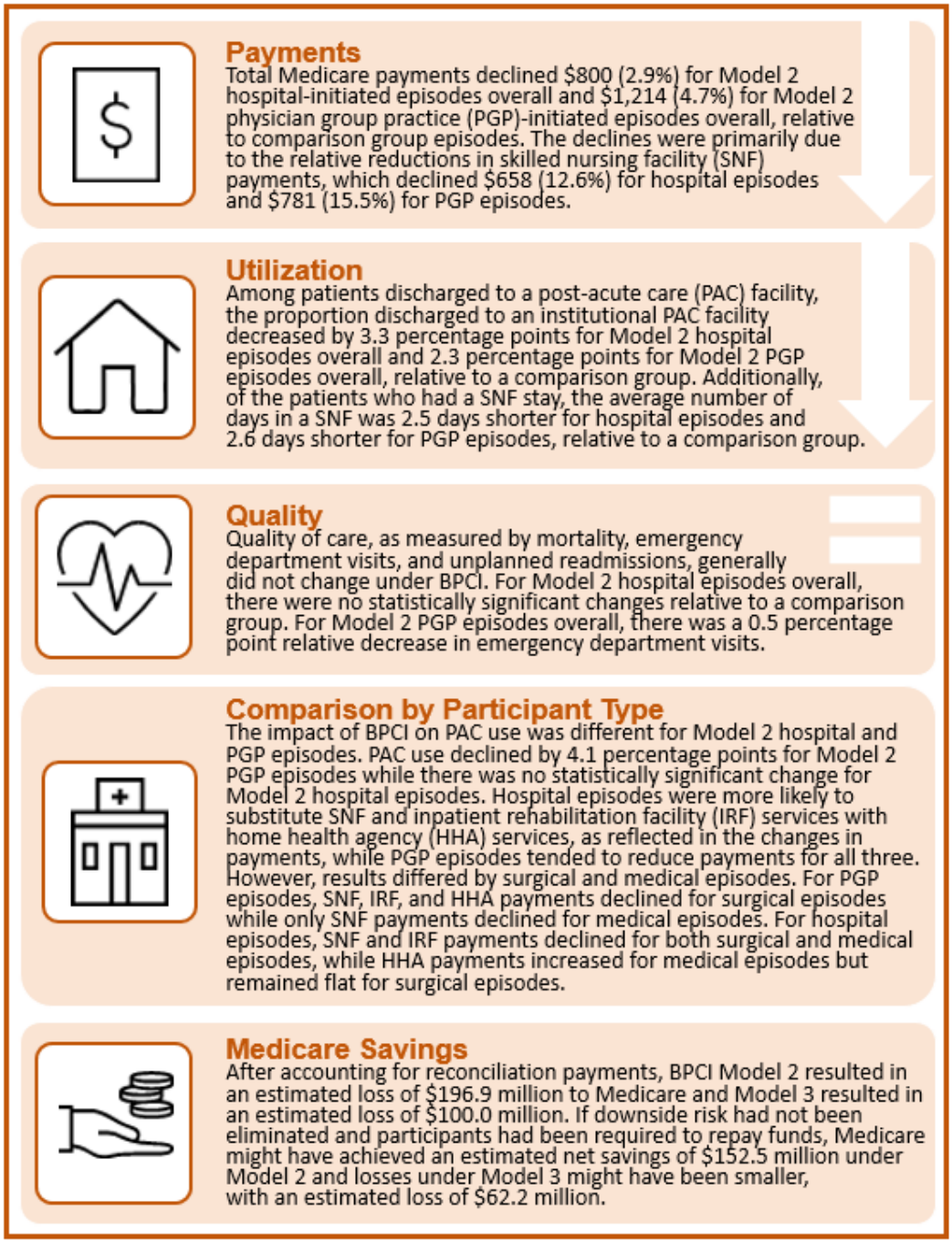
[1] See the Year 5 Evaluation & Monitoring Annual Report for previously reported results. The report is available for download at https://innovation.cms.gov/initiatives/Bundled-Payments/index.html. Model 1 began earlier than Models 2, 3, and 4 and was evaluated separately. The report is available for download from https://innovation.cms.gov/Files/reports/bpci-mdl1yr2annrpt.pdf.

initiated by Model 2 PGPs during the first four years of BPCI. Because participants were allowed to join BPCI over an extended period and to stop participating at any time with prior notice, the results in this report represent an average of eight quarters of participation for hospital and PGP participants. As of September 2017, approximately 29% of hospitals and 30% of PGPs that had participated withdrew completely[2] from the initiative. Exhibit ES-1 summarizes the key impact estimates for hospital and PGP participants through the first four years of the initiative. Overall, our results are consistent with previous reports that demonstrated the initiative has resulted in reductions to episode payments, which were driven by declines in institutional PAC use and payments, while maintaining quality of care.[3] However, Medicare continued to experience net financial losses under this model.

[2] Participants withdrew completely from BPCI when they no longer participated in any clinical episodes.

[3] See the Year 5 Evaluation & Monitoring Annual Report for previously reported results. The report is 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, Model 2 Hospitals and Model 2 Physician Group Practices, Q4 2013 – Q3 2017



Note: The estimates in this exhibit are the results of a difference-in-differences model. The results stated in the payment, utilization, and quality boxes are based on clinical episodes that had sufficient volume for risk-adjustment. Results in the comparison by participant type section are based on weighted averages of the clinical episodes that Model 2 hospitals and PGPs had in common, where the weights are based on the combined episode volume across both participant types; this common weight ensures that results are comparable given the different distribution of episodes across clinical episodes. The Medicare savings results include all 48 BPCI clinical episodes.

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 2017) 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 Model 2 as well as estimates of net savings to Medicare under Models 2 and 3. Consistent with previous reports, results indicate that Medicare payments declined under the BPCI initiative, which can largely be attributed to the continued pattern of reduced institutional PAC use, but this decline did not translate into overall savings to Medicare after taking into account reconciliation payments made to participants. Compared to the findings through December 2016 (Q4 2016) in the Year 5 Evaluation and Monitoring report, the declines in total payments and institutional PAC use in this report were generally larger through September 2017 (Q3 2017). While promising, this finding could be due in part to providers selectively withdrawing from the initiative or stopping participation in a clinical episode. New analysis shows there were differences in findings between episodes initiated by Model 2 hospitals and Model 2 PGPs, as well as between surgical and medical clinical episodes. Relative declines in total payments were larger for surgical episodes than for medical episodes for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes. This was due in part to relative declines in institutional PAC payments that were larger for surgical episodes than for medical episodes. In addition, home health agency (HHA) payments for surgical episodes decreased for Model 2 PGP episodes but did not change for Model 2 hospital episodes, while HHA payments for medical episodes increased for Model 2 hospital episodes and did not change for Model 2 PGP episodes. With respect to the effect of BPCI on quality of care as measured by readmissions, emergency department use, and mortality, there were no changes for Model 2 hospital-initiated episodes, and while there were a few indications for Model 2 PGP-initiated episodes, the direction was not consistent. 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. Medicare net losses under Model 2 through September 2017 in this report were slightly smaller compared to the estimated net losses through December 2016 in the Year 5 Evaluation and Monitoring Annual Report. However, net losses under Model 3 were larger through September 2017. The final BPCI report will include updated impact estimates for Model 2, 3 and 4 episodes and estimates of Medicare savings through the end of the BPCI initiative.

I. Introduction

A. BPCI Initiative

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 sixth report describes results for key outcomes under BPCI Model 2 for the first four years of the initiative, as well as an estimate of net savings to the Medicare program under Models 2 and 3. A summative evaluation of Models 2, 3, and 4 will be conducted in the seventh and final report under this contract.[5]

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) equal to the difference. 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 equal to 2–3% of what CMS estimated payments would have been absent the initiative (referred to as the benchmark price).

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. 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 services provided during the episode of care were bundled for payment purposes. The bundle definition, payment approach, and types of eligible episode-initiating (EI) providers varied by Model, as depicted in Exhibit 1. Throughout this report, we refer to Awardees and EIs collectively as "participants." See **Appendix A** for the

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

[5] Due to low participation, Model 4 results are not presented in this report. Please refer to the Year 2 annual report for details on Model 4. The report is available for download from https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

[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.

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.

Providers and other organizations that volunteered to participate could enter into the risk-bearing phase of the initiative during a two-year period 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 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.[7]

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

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

	Eligible Episode Initiator Types	Services Included in the Bundle[1]	Episode Triggering Events	Episode Length Options[2]	Payment Method	Potential Savings for Participants	Waiver Options
Model 2	✓ Hospital ✓ PGP	✓ Anchor hospitalization and concurrent professional services ✓ Post-discharge services, including hospital readmissions	✓ Admission to an episode-initiating hospital ✓ Attending or operating physician for the hospitalization is an episode-initiating PGP	✓ 30 days ✓ 60 days ✓ 90 days	Providers 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	✓ Services furnished after the anchor hospital discharge, including professional services and readmissions	✓ Admission to an episode-initiating SNF, HHA, IRF, or LTCH within 30 days of discharge from a hospital ✓ Admission to a PAC setting within 30 days of discharge from a hospitalization where the attending or operating physician is associated with a participating PGP	✓ 30 days ✓ 60 days ✓ 90 days	Providers 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	✓ Anchor hospitalization and concurrent professional services ✓ Hospital readmissions and associated professional services	✓ Admission to an episode-initiating 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 cost savings. For details on the various 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.

[1] Certain services, such as hospice, readmissions for certain MS-DRGs, and some Part B services are excluded.

[2] Post-hospital discharge

B. Research Questions

This sixth annual report provides an update of the impact of BPCI on payments, utilization of services, and quality of care for Medicare beneficiaries under Model 2 during the first four years of the initiative, from October 1, 2013 through September 30, 2017. The report also includes an updated analysis of net savings to the Medicare program under both 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 2.

Exhibit 2: BPCI Research Questions

Research Questions

1. What is the impact of the BPCI initiative on episode payments, utilization of post-acute care services, and the quality of care for Medicare beneficiaries?
2. How do the impacts of the BPCI initiative on Model 2 hospital episodes and Model 2 PGP episodes compare for the same set of clinical episodes?
3. What is the impact of the BPCI initiative on net savings to the Medicare program?

C. Methods

This report includes impact estimates for key outcomes for episodes initiated by Model 2 hospitals and physician group practices (PGPs). The outcomes and descriptions are listed in Exhibit 3 below. (See **Appendix C** for additional details.) During the first four years of the initiative, Model 2 hospital episodes accounted for 49% of the more than one million BPCI episodes initiated in Models 2, 3, and 4, while Model 2 PGP episodes accounted for 38%. At the nationwide level, Model 2 hospitals accounted for 2.8% of all eligible hospital discharges and Model 2 PGPs accounted for 2.5%.[8]

[8] All eligible hospital discharges include discharges that were eligible to be a BPCI episode from an eligible hospital. Eligible BPCI discharges are discharges in which the beneficiaries met the following criteria: 1) had a complete fee-for-service (FFS) enrollment history six months prior to anchor hospital admission; 2) maintained FFS enrollment in Medicare Parts A and B throughout the measurement period or until death; 3) had a measurement period that ended on or before December 31, 2017. Eligible hospitals include BPCI-participating hospitals and all other Inpatient Prospective Payment System (IPPS) hospitals excluding hospitals in Maryland. See **Appendix D** for the share of Model 2 hospital discharges among all eligible hospital discharges by clinical episode.

Exhibit 3: Claims-based Outcomes Definitions

Domain	Outcome Name	Description
Payment	Total Standardized Allowed Payment Amount	Average total Medicare Part A & B standardized allowed payment amount, during the inpatient stay + 90-day PDP
	Total Standardized Paid Amount	Average total Medicare Part A & B standardized paid amount, excluding beneficiary cost sharing, during the inpatient stay + 90-day PDP
	SNF Standardized Allowed Payment Amount, 90-day PDP	Average Medicare Part A standardized allowed payment amount for SNF, totaled within the 90-day PDP
	IRF Standardized Allowed Payment Amount, 90-day PDP	Average Medicare Part A standardized allowed payment amount for IRF, totaled within the 90-day PDP
Utilization	Number of Days in a SNF	Average number of SNF days of care during the 90-day PDP
	Discharged to Any PAC Setting	The proportion of episodes that were discharged from the anchor hospital to any PAC setting, including HHA
	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 among episodes that were discharged to any PAC setting (including HHA)
Quality	Unplanned Readmission Rate	Episodes with one or more unplanned, all-cause readmissions for any condition, 90 days after anchor discharge
	ED Use without Hospitalization	Episodes with one or more ED visits for which the beneficiary requires medical treatment but is not admitted to the hospital 90 days after discharge from an anchor hospital stay
	All-cause Mortality	Death from any cause during 90 days after discharge from the anchor hospital stay

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

We present three different sets of impact estimates in the report. First, in sections II.A and II.B, we present overall results for Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes, which include all clinical episodes with sufficient volume for risk adjustment. These clinical episodes represented 98% of episodes initiated by all Model 2 hospitals and 93% of episodes initiated by all Model 2 PGPs during the first four years of the initiative. (See **Appendix C** for a complete list of the 32 hospital-initiated clinical episodes and the 21 PGP-initiated clinical episodes included in the overall estimates.) Second, in those two sections, we also present the results for each of the top 10 clinical episodes with the largest volume based on the combined number of BPCI episodes for Model 2 hospitals and PGPs. The top 10 clinical episodes with the largest volume are listed in Exhibit 4 below. These clinical episodes represented 78% of all episodes initiated by Model 2 hospitals during the first four years of the initiative and 74% of all Model 2 PGP episodes. Third, in section II.C, we compare Model 2 hospital- and Model 2 PGP-initiated episodes based on a weighted average of the 21 clinical episodes included in the impact estimates for both participant types to allow a direct comparison. These 21 clinical episodes represent 91% of all Model 2 hospital-initiated episodes and 93% of all Model 2 PGP-initiated episodes during the first four years of the initiative.

Exhibit 4: Top 10 Clinical Episodes with the Largest Combined Episode Volume, Model 2 Hospitals and Model 2 PGPs, Q4 2013 – Q3 2017

Clinical Episode	Intervention Episodes: Model 2 Hospitals	Intervention Episodes: Model 2 PGPs	Total Intervention Episodes
Major joint replacement of the lower extremity	146,435	90,976	237,411
Sepsis	39,395	26,183	65,578
Congestive heart failure	49,761	12,659	62,420
Simple pneumonia and respiratory infections	31,917	13,300	45,217
Chronic obstructive pulmonary disease, bronchitis, asthma	28,672	11,849	40,521
Stroke	17,834	4,048	21,882
Urinary tract infection	12,384	7,461	19,845
Hip & femur procedures except major joint	11,173	7,397	18,570
Renal failure	10,756	6,970	17,726
Medical non-infectious orthopedic	9,818	3,963	13,781

***Note*:** Clinical episodes are ordered by the number of total intervention episodes.
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 2017) for BPCI providers.

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. 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 2017).[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 population. We report the highest level of statistical significance, $p<0.05$ or $p<0.10$, when we report a single estimate. When we report multiple estimates, we indicate the number that are statistically significant at the 10% level. See **Appendix C** for a detailed description of the DiD methodology, outcome definitions, methods for identifying comparison populations, statistical models, sensitivity analysis, and tests for parallel trends between BPCI and comparison episodes in the baseline period.

[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.

In addition to the payment, utilization, and quality of care impact estimates, we also estimate the net savings for the Medicare program for the first four years of BPCI under both Model 2 and Model 3. Net savings to Medicare was calculated by subtracting reconciliation payments from the change in aggregate non-standardized payments for all 48 episodes under Model 2 and Model 3, between the baseline period (October 2011 through September 2012) and the intervention period (October 2013 through September 2017). In this report, we also examine savings at the clinical episode level for the same 10 Model 2 hospital and Model 2 PGP clinical episodes listed in Exhibit 4 above. We present the highest level of statistical significance for the impact estimates, $p<0.05$ or $p<0.10$.

II. Model 2 Results

A. Impact of BPCI on Hospital-initiated Episodes

1. Key Findings

Impact on Payments

- During the first four years of the BPCI initiative, total Medicare-allowed payments during the inpatient stay plus 90 days post discharge declined an estimated $800 for Model 2 hospital-initiated episodes overall. There were declines for each of the 10 Model 2 hospital clinical episodes analyzed, and six of the 10 declines were statistically significant.
- SNF payments in the 90 days post discharge declined by an estimated relative $658 for hospital episodes overall, contributing to the reduction in total payments. There were declines in SNF payments for each of the 10 clinical episodes analyzed, and the decline was statistically significant for nine episodes.

Impact on Post-acute Care Utilization

- Among patients who were discharged to any PAC, there was an estimated relative decline of 3.3 percentage points in the proportion discharged to institutional PAC settings for Model 2 hospital-initiated episodes overall. The proportion declined for nine of the 10 clinical episodes analyzed, and five of the nine declines were statistically significant.
- Among patients with at least one day in the SNF, there was a statistically significant estimated relative decline of 2.5 SNF days for Model 2 hospital episodes overall. There were statistically significant declines in the number of SNF days for each of the 10 clinical episodes analyzed.

Impact on Quality

- In general, claims-based quality measures did not change under Model 2 for all hospital-initiated episodes or for the 10 clinical episodes analyzed separately.

2. Sample Characteristics

Of the 423 hospitals that voluntarily participated in BPCI, we analyzed the characteristics of 419 hospitals that received Medicare certification by 2011. These 419 hospitals were different from non-participating hospitals based on key characteristics. BPCI 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. Exhibits 5a-5b describe these and additional baseline characteristics. (See **Appendix C** for further details about these measures.) 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 E** of the Year 5 Evaluation & Monitoring Annual Report for more details and additional sample characteristics.)

Exhibit 5a-5b: 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 419 BPCI-participating hospitals, we were able to identify comparison hospitals for 406 hospitals in the analysis of the impact estimates. Exhibit 6 describes the sample of Model 2 hospitals included in the analysis, with characteristics for each of the top 10 clinical episodes with the largest volume presented first, followed by the characteristics of Model 2 hospitals in the analysis overall. The BPCI hospitals included in the impact analysis participated for an average of eight quarters. By the end of the fourth year of the initiative, 271 (67%) Model 2 hospitals in the analytic sample stopped participating in at least one clinical episode.[10] Twenty-eight percent of the episodes in the analytic sample were initiated by the hospitals that stopped participating in the clinical episode.

[10] As of September 2017, approximately 29% of hospitals that had ever participated in BPCI (regardless of whether they were in the analytic sample) withdrew completely from the initiative and no longer participated in any clinical episodes.

Exhibit 6: Characteristics of the Matched BPCI Providers Included in the BPCI Impact Estimates, Model 2 Hospitals, Q4 2013 – Q3 2017

Clinical Episode	BPCI Hospitals (N)	BPCI Episodes (N)	Average Length of Participation (Quarters)[1]	Hospitals that Stopped Participating in the Clinical Episode (N)	Proportion of Episodes from Hospitals that Stopped Participating (%)
Congestive heart failure	173	50,226	8	78	34.5%
Chronic obstructive pulmonary disease, bronchitis, asthma	133	29,058	8	49	26.6%
Hip & femur procedures except major joint	101	11,181	8	42	25.8%
Medical non-infectious orthopedic	94	9,901	7	46	28.4%
Major joint replacement of the lower extremity	303	146,492	9	91	17.9%
Renal failure	75	10,880	7	42	41.3%
Sepsis	119	39,823	7	59	32.4%
Simple pneumonia and respiratory infections	132	32,143	8	50	29.0%
Stroke	77	17,952	8	34	24.0%
Urinary tract infection	83	12,458	7	36	22.3%
Model 2 Hospitals Overall	**406**	**452,609**	**8**	**271**	**27.7%**

Note: Model 2 Hospitals Overall results represent 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 88% of all episodes across the 32 clinical episodes. The number of BPCI hospitals that stopped participating in the clinical episode represent 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.

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 2017) for BPCI providers.

[1] 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.

3. Payment, Utilization, and Quality Outcomes

This section presents the BPCI impact estimates for payments, utilization, and quality for hospital-initiated episodes for the first four years of the initiative. We present the results separately for the top 10 clinical episodes with the largest volume and for Model 2 hospitals overall. Detailed results are located in **Appendix E**.

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

In the four years since the implementation of BPCI, the total standardized allowed payment amount for the inpatient stay plus the 90-day PDP declined from the baseline to the intervention period relative to the comparison group for all 10 BPCI hospital-initiated clinical episodes analyzed. The decline was statistically significant for six clinical episodes (Exhibit 7). Total

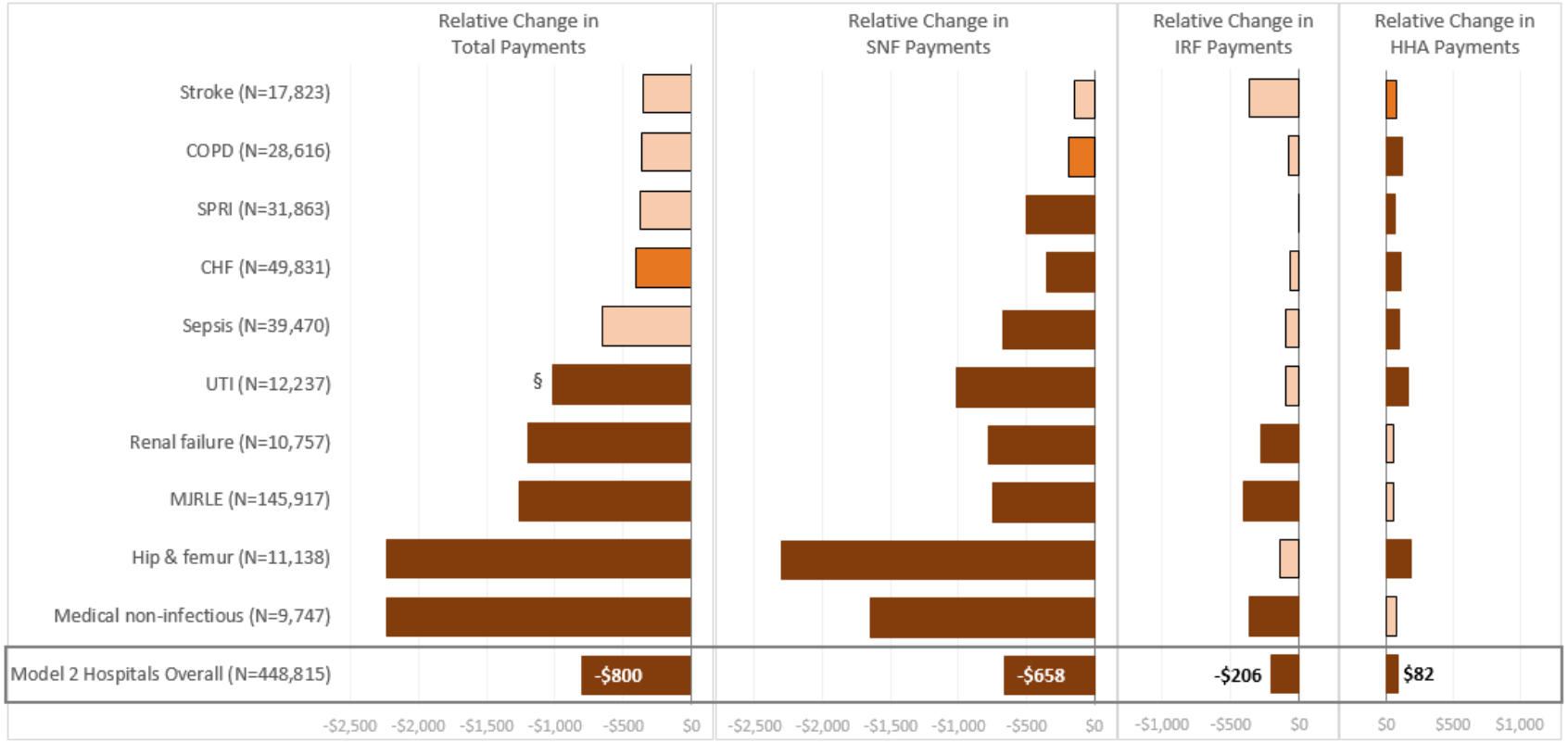
payments also declined an estimated $800 more for BPCI episodes relative to the comparison group for Model 2 hospital-initiated episodes overall ($p<0.05$). This translates to a percent change of 2.9% relative to what payments would have been absent the BPCI initiative (Exhibit 8).

The decline in total payments was driven by declines in skilled nursing facility (SNF) and inpatient rehabilitation facility (IRF) payments during the 90 days post discharge. SNF payments declined for all 10 clinical episodes, and the decline was statistically significant for nine of these episodes. Furthermore, IRF payments decreased for nine of the 10 clinical episodes; the decline in three clinical episodes was statistically significant ($p<0.05$). For Model 2 hospital-initiated episodes overall, SNF payments declined an estimated $658, or 12.6% ($p<0.05$), and IRF payments declined an estimated $206, or 16.8% ($p<0.05$).

While SNF and IRF payments declined, home health agency (HHA) payments increased for all 10 clinical episodes examined, and the increase was statistically significant for seven of these episodes (see Exhibit 7). For Model 2 hospital-initiated episodes overall, HHA payments increased an estimated $82, or 5.2%, more for BPCI episodes than for the comparison group ($p<0.05$).

These changes in service-level payments provide insights into how hospitals reduced total payments. Generally, average IRF payments are higher than average SNF payments, and both of these institutional PAC settings tend to have higher payments than HHA payments. The observed changes in service-level payments are consistent with the incentives of the BPCI program to reduce total payments by shifting service use from more expensive institutional PAC settings to less expensive HHA settings.

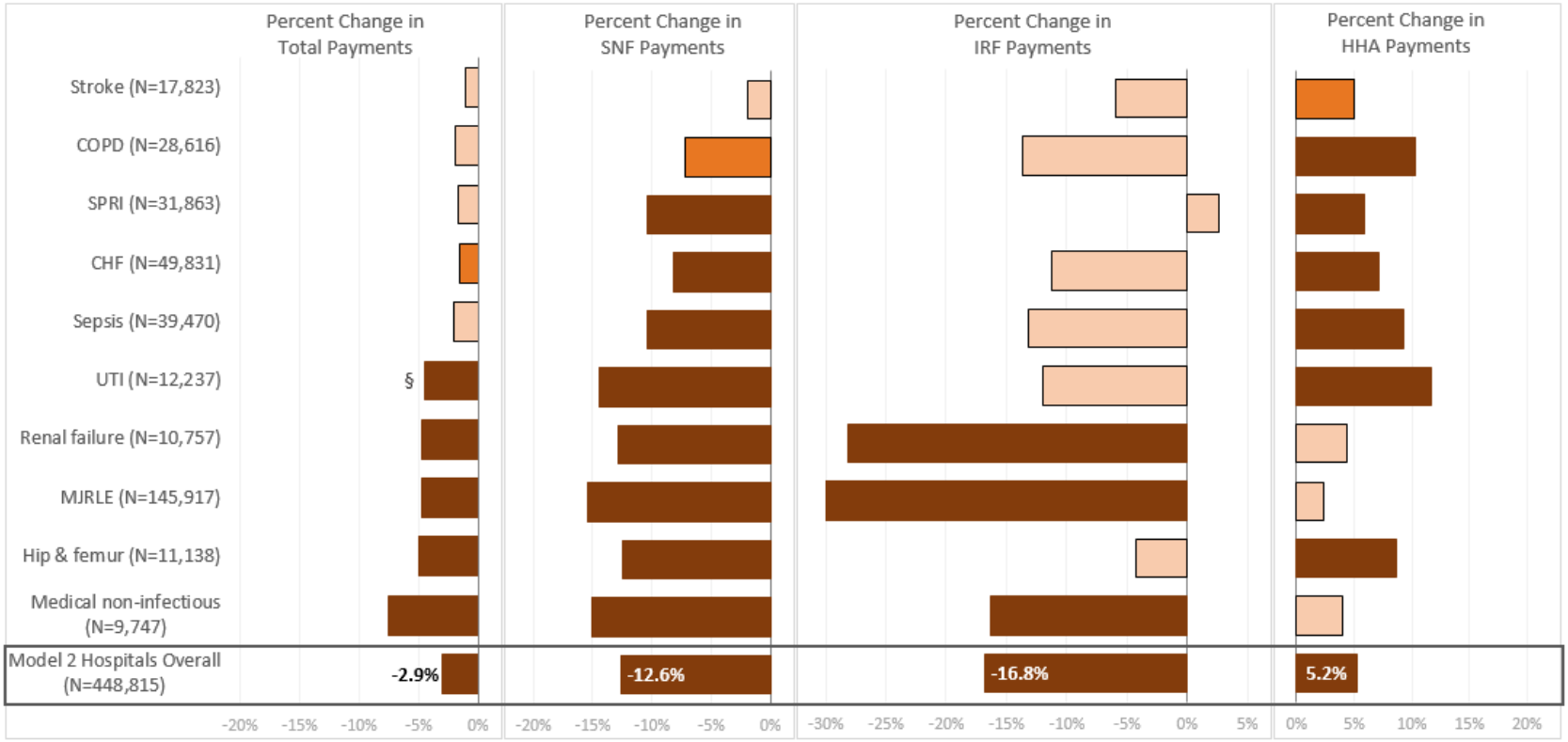
Exhibit 7: Impact of BPCI on Standardized Allowed Payment Amount Outcomes, by Clinical Episode and Overall, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. The 10 clinical episodes are ordered starting with the smallest relative decline in the total allowed payment amount during the inpatient stay through 90-day PDP. The Model 2 Hospitals Overall result includes the 32 clinical episodes that had sufficient volume to allow for risk-adjustment. These payment measures are not conditional upon the use of the service. The N for each clinical episode represents the number of BPCI episodes in the total payments calculation, while the Ns for the SNF, IRF, and HHA payment calculations may differ slightly. COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections. CHF=congestive heart failure. UTI=urinary tract infection. MJRLE=major joint replacement of the lower extremity. Hip & femur refers to hip & femur procedures except major joint. Medical non-infectious refers to medical non-infectious orthopedic. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level. § 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 2017) for BPCI and comparison providers.

Exhibit 8: Percent Change in Standardized Allowed Payment Amount Outcomes, by Clinical Episode and Overall, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The 10 clinical episodes are ordered starting with the smallest relative decline in the total payment measure. The Model 2 Hospitals Overall result includes the 32 clinical episodes that had sufficient volume to allow for risk-adjustment. These payment measures are not conditional upon the use of the service. The N for each clinical episode represents the number of BPCI episodes in the total payments calculation, while the Ns for the SNF, IRF, and HHA payment calculations may differ slightly. Episode payments absent BPCI, or the counterfactual, is the BPCI baseline payment amount plus the change in episode payment amount for the comparison group. The counterfactual can be expressed as: BPCI before + (Comparison after – Comparison before). The percent change can then be expressed as: (BPCI after – Counterfactual) / (Counterfactual). COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections. CHF=congestive heart failure. UTI=urinary tract infection. MJRLE=major joint replacement of the lower extremity. Hip & femur refers to hip & femur procedures except major joint. Medical non-infectious refers to medical non-infectious orthopedic. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level. § 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.

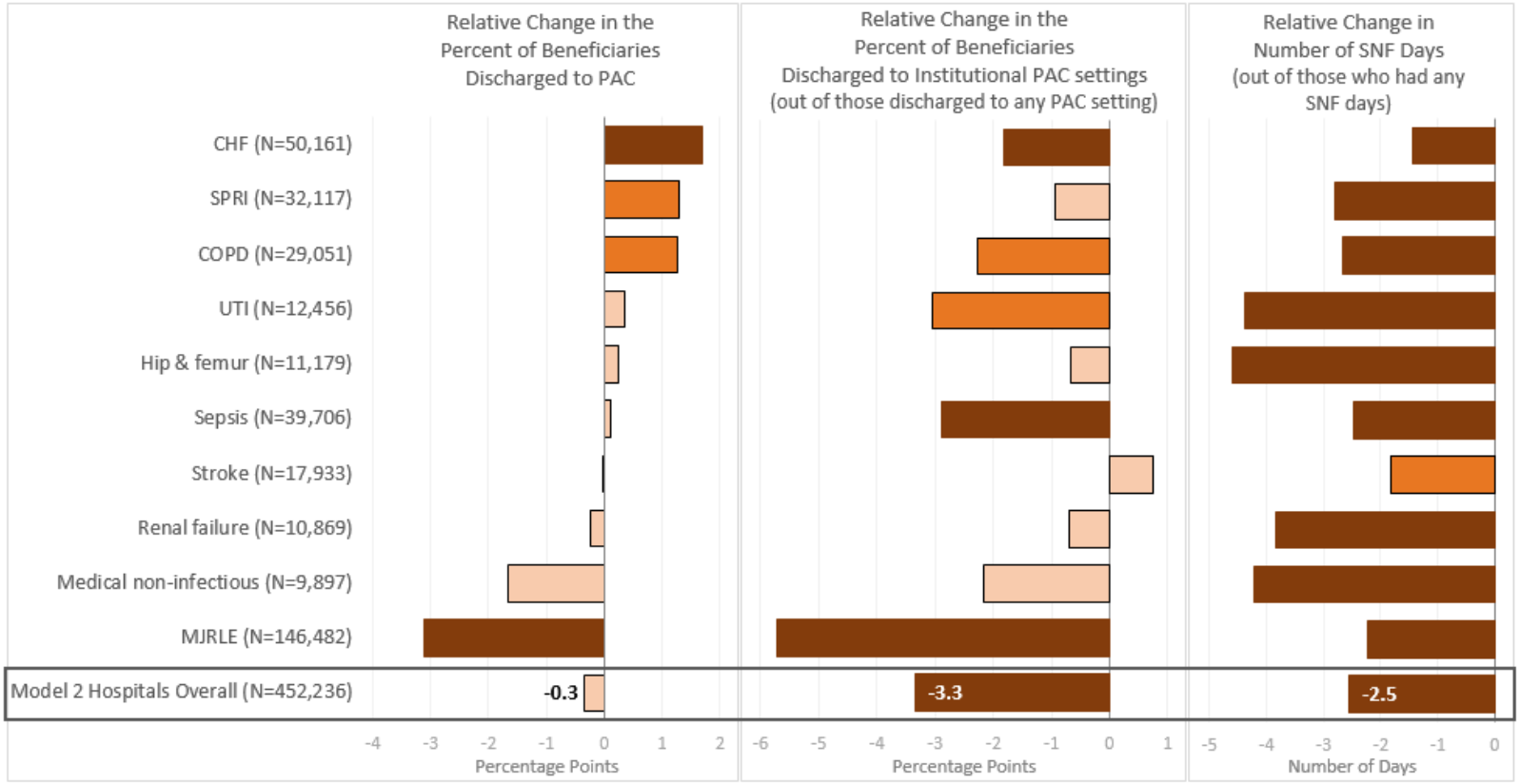
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2017) for BPCI and comparison providers.

b. How have the services changed under BPCI?

The changes in institutional PAC use were consistent with the changes in payments. Among patients who were discharged to any PAC setting, the proportion discharged to institutional PAC (SNF, IRF, or long-term care hospitals) declined in nine of the 10 clinical episodes, and the decline was statistically significant for five (see Exhibit 9). For Model 2 hospital-initiated episodes overall, the proportion of patients discharged to institutional PAC, of those who were discharged to any PAC setting, declined an estimated 3.3 percentage points more for BPCI episodes than for comparison group episodes ($p<0.10$). For BPCI episodes with at least one day in the SNF, the number of SNF days within the 90-day PDP declined relative to the comparison group in all 10 clinical episodes (see Exhibit 9). This result was also statistically significant for Model 2 hospital-initiated episodes overall, with an estimated 2.5 fewer SNF days during the 90-day PDP for BPCI episodes relative to the comparison group ($p<0.05$).

While there is evidence of lower use of institutional PAC services across clinical episodes, the change in the proportion of patients discharged to any PAC setting varied depending on whether the increase in the proportion of patients discharged to HHA offset the decrease in the proportion of patients discharged to institutional PAC settings. Across the 10 clinical episodes analyzed, the percentage of patients discharged to any PAC setting decreased from the baseline to the intervention period relative to the comparison group for three clinical episodes and was statistically significant for one: major joint replacement of the lower extremity (Exhibit 9). The percentage discharged to any PAC setting increased for six clinical episodes and was statistically significant for three (Exhibit 9). These three episodes also tended to have smaller relative declines in SNF payments, IRF payments, and the number of SNF days compared to other clinical episodes. For one clinical episode, stroke, there was no change compared to the comparison group. For Model 2 hospital-initiated episodes overall, there was a small decrease in the proportion of patients discharged to PAC relative to the comparison group, which was not statistically significant.

Exhibit 9: Impact of BPCI on PAC Utilization, by Clinical Episode and Overall, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2017



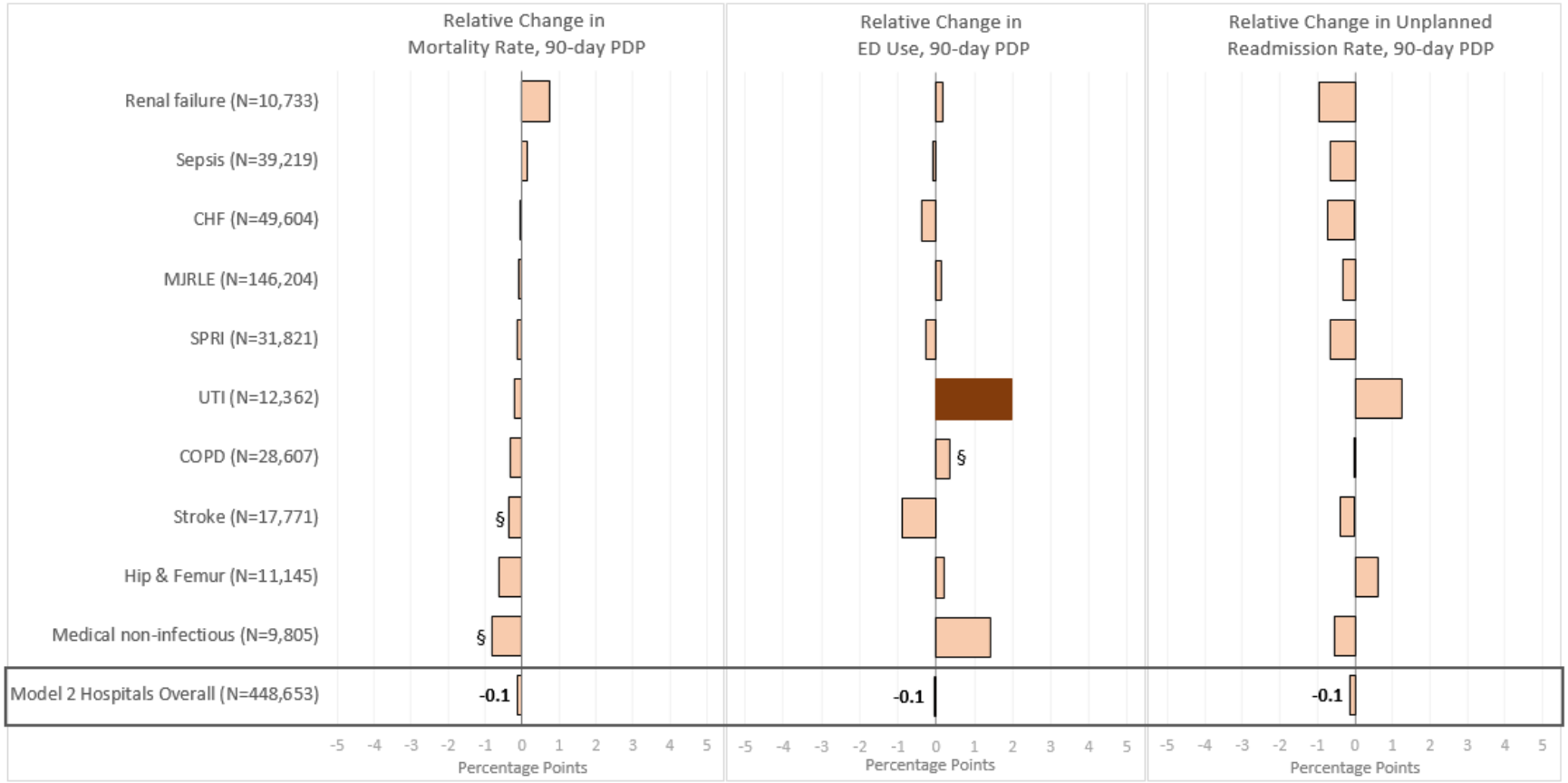
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and ordered starting with the smallest relative decline in the percent of beneficiaries discharged to PAC. The Model 2 Hospitals Overall result includes the 32 clinical episodes that had sufficient volume to allow for risk-adjustment. The percent of beneficiaries discharged to institutional PAC settings is conditional on being discharged to any PAC setting, and the number of SNF days is conditional on having any SNF use. Therefore, the N for each clinical episode represents the number of BPCI episodes in the percent discharged to any PAC setting calculation, while the Ns for the percent discharged to institutional PAC settings and the number of SNF days differ. See **Appendix E** for sample size by outcome by clinical episode. CHF=congestive heart failure. SPRI=simple pneumonia and respiratory infections. COPD=chronic obstructive pulmonary disease. UTI=urinary tract infection. Hip & femur refers to hip & femur procedures except major joint. Medical non-infectious refers to medical non-infectious orthopedic. MJRLE=major joint replacement of the lower extremity. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

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 2017) 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 quality of care. Moreover, reconciliation payments are based on reducing payments below a target price and are not linked to the quality of care provided. We therefore examine three key claims-based quality measures to assess changes in quality. Consistent with previous reports, the results of this analysis suggest that the quality of care generally did not change under BPCI for Model 2 hospital-initiated episodes in the clinical episodes analyzed or for Model 2 hospital-initiated episodes overall. Across the three measures and 10 clinical episodes analyzed, there was only one statistically significant relative change: an increase in emergency department (ED) use during the 90-day PDP for urinary tract infection episodes ($p<0.05$). (See Exhibit 10.) The consistent lack of harm across multiple comparisons strongly affirms our conclusion that BPCI Model 2 was not associated with a worsening of care for beneficiaries.

Exhibit 10: Impact of BPCI on Claims-based Quality Outcomes, by Clinical Episode and Overall, Model 2 Hospitals, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this table are the results of a difference-in-differences (DiD) model. The N for each clinical episode represents the number of BPCI episodes in the ED use and unplanned readmission calculations. ED=emergency department. CHF=congestive heart failure. MJRLE=major joint replacement of the lower extremity. SPRI=simple pneumonia and respiratory infections. UTI=urinary tract infection. COPD=chronic obstructive pulmonary disease. Hip & femur refers to hip & femur procedures except major joint. Medical non-infectious refers to medical non-infectious orthopedic. Dark orange cells indicate DiD estimates that are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level. § 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 2017) for BPCI and comparison providers.

B. Impact of BPCI on PGP-initiated Episodes

1. Key Findings

Impact on Payments

- During the first four years of the BPCI initiative, total Medicare-allowed payments during the inpatient stay plus 90 days post discharge declined an estimated $1,214 for Model 2 PGP-initiated episodes overall relative to a comparison group. There were declines for seven of the 10 Model 2 PGP clinical episodes analyzed, and four of the declines were statistically significant.
- SNF payments in the 90 days post discharge declined by an estimated relative amount of $781 for PGP episodes overall, contributing to the reduction in total payments. There were declines in SNF payments for eight of the 10 clinical episodes analyzed, and the decline was statistically significant for five episodes.

Impact on Post-acute Care Utilization

- Among patients who were discharged to any PAC, there was an estimated relative decline of 2.3 percentage points in the proportion discharged to institutional PAC settings for Model 2 PGP-initiated episodes overall. The proportion declined for eight of the 10 clinical episodes analyzed, and two of the eight declines were statistically significant.
- Among patients with at least one day in the SNF, there was a statistically significant estimated relative decline of 2.6 SNF days for PGP episodes overall. There were declines for eight of the 10 clinical episodes analyzed, and the decline was statistically significant for five clinical episodes.

Impact on Quality

- There were a few indications of changes in quality for the Model 2 PGP-initiated episodes, but the direction was not consistent.

2. Sample Characteristics

We analyzed characteristics of BPCI-participating physician group practices (PGPs), such as the composition of physician specialties and volume.[11] Of the 272 PGPs participating in Model 2, 245 were included in the analysis of PGP characteristics.[12] Of these 245 PGPs, 20% were classified as hospitalist practices, while 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

[11] 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.

[12] 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.

48 BPCI clinical episodes. See **Appendix C** for further details about participant characteristics, including physician specialty categories.

Of the 245 BPCI-participating PGPs, we were able to identify comparison groups for 189 PGPs in the analysis of the impact estimates. Exhibit 11 describes the sample of Model 2 PGPs included in the analysis, with characteristics for each of the top 10 clinical episodes with the largest volume presented first, followed by Model 2 PGP overall. Providers in this analysis participated for an average of eight quarters. By the end of the fourth year of the initiative, 109 of the 189 (58%) Model 2 PGPs stopped participating in at least one clinical episode.[13] Twenty-nine percent of the episodes in the analytic sample were initiated by the PGPs that stopped participating in the clinical episode.

Exhibit 11: Characteristics of the Matched BPCI Providers Included in the BPCI Impact Estimates, Model 2 PGPs, Q4 2013 – Q3 2017

Clinical Episode	BPCI PGPs (N)	BPCI Episodes (N)	Average Length of Participation (Quarters)[1]	PGPs that Stopped Participating in the Clinical Episode (N)	Proportion of Episodes from PGPs that Stopped Participating in the Clinical Episode (%)
Congestive heart failure	47	12,761	7	18	30.7%
Chronic obstructive pulmonary disease, bronchitis, asthma	57	12,010	7	29	39.5%
Hip & femur procedures except major joint	63	7,401	7	28	49.6%
Medical non-infectious orthopedic	42	4,007	7	20	46.9%
Major joint replacement of the lower extremity	112	90,992	8	31	17.2%
Renal failure	43	7,053	7	21	46.9%
Sepsis	61	26,472	7	31	36.8%
Simple pneumonia and respiratory infections	63	13,415	7	30	35.3%
Stroke	36	4,073	8	12	26.2%
Urinary tract infection	51	7,507	7	22	52.1%
Model 2 PGPs Overall	**189**	**227,246**	**8**	**109**	**28.5%**

Note: Model 2 PGPs Overall results represent 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 61% of all episodes across the 21 clinical episodes. The number of BPCI PGPs that stopped participating in the clinical episode represent 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.

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 2017) for BPCI providers.

[1] 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.

[13] As of September 2017, approximately 30% of PGPs that had ever participated in BPCI (regardless of whether they were in the analytic sample) withdrew completely from the initiative and no longer participated in any clinical episodes

3. Payment, Utilization, and Quality Outcomes

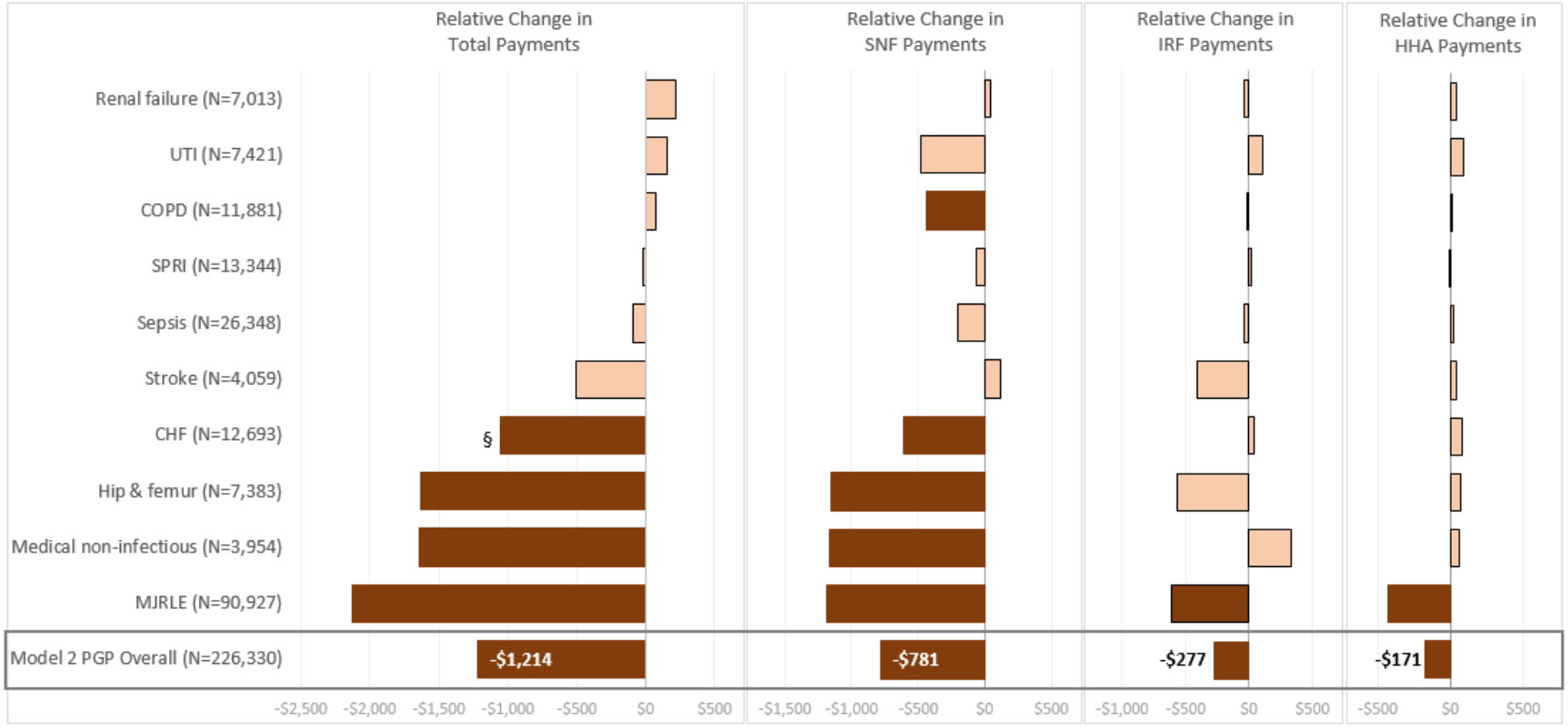
This section presents the BPCI impact estimates for payments, utilization, and quality for PGP-initiated episodes for the first four years of the initiative. We present the results separately for the top 10 clinical episodes with the largest volume and for Model 2 PGP episodes overall. Detailed results are located in **Appendix F**.

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

In the four years since the implementation of BPCI, the total standardized allowed payment amount for the inpatient stay plus the 90-day PDP declined from the baseline to the intervention period relative to the comparison group for seven of the 10 BPCI PGP-initiated clinical episodes analyzed. The decline was statistically significant for four clinical episodes ($p<0.05$). (See Exhibit 12.) For Model 2 PGP-initiated episodes overall, total payments declined an estimated $1,214 more for BPCI episodes relative to the comparison group ($p<0.05$). This translates to a percent change of 4.7% relative to what payments would have been absent the BPCI initiative (Exhibit 13).

Declines in PAC payments during the 90 days post discharge contributed to the decline in total payments. SNF payments declined for eight of the 10 clinical episodes, and the decline was statistically significant for five clinical episodes ($p<0.05$). (See Exhibit 12.) Additionally, IRF payments decreased for five of the 10 clinical episodes; although the decline was statistically significant for only one clinical episode ($p<0.05$). Finally, while HHA payments increased slightly for eight clinical episodes, none of these increases were statistically significant. HHA payments decreased for the remaining two clinical episodes, and the decline was statistically significant for major joint replacement of the lower extremity ($p<0.05$). (See Exhibit 12.) For Model 2 PGP-initiated episodes overall, SNF payments declined an estimated $781 (15.5%), IRF payments declined an estimated $277 (30.4%), and HHA payments declined an estimated $171 (11.0%) more for BPCI episodes relative to the comparison group ($p<0.05$).

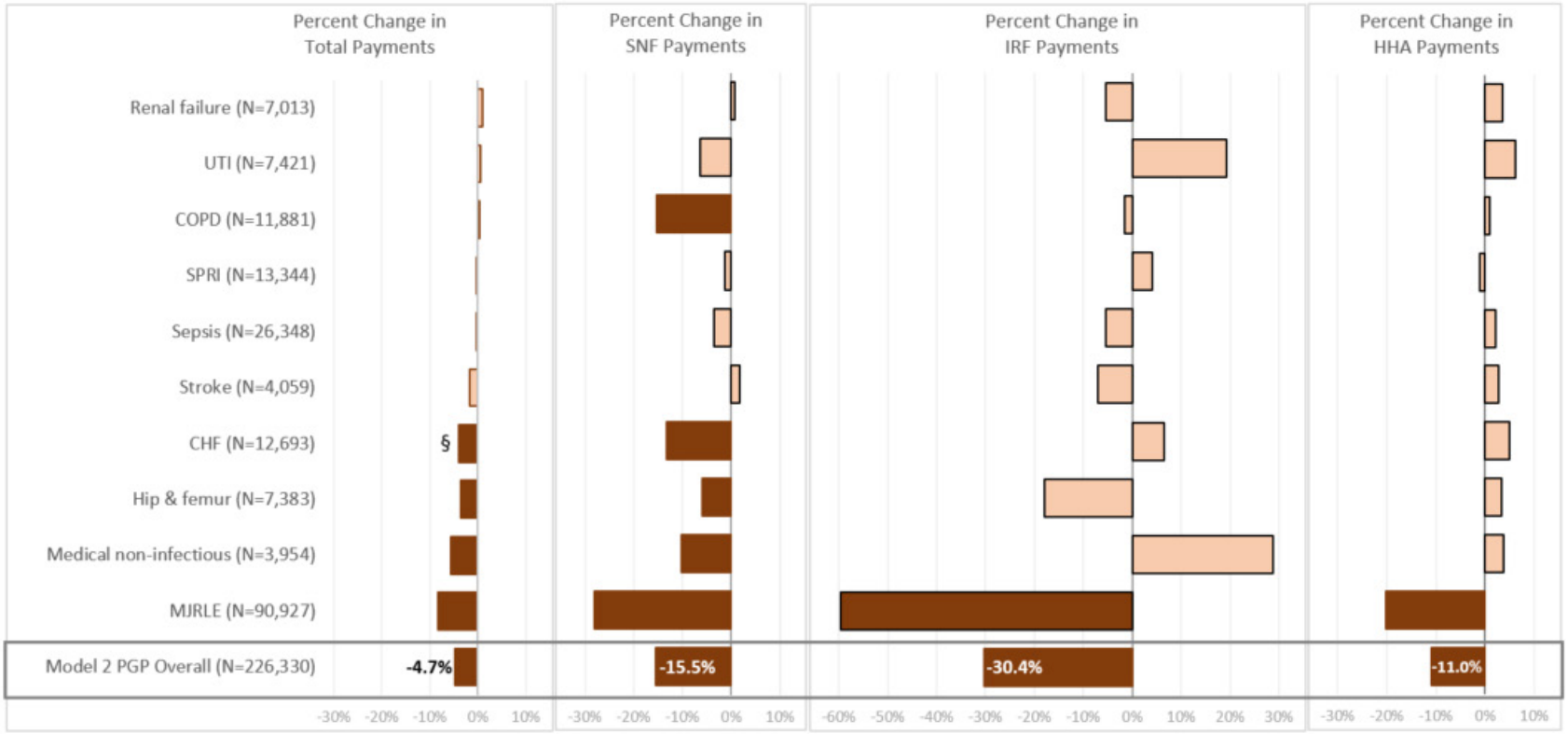
Exhibit 12: Impact of BPCI on Standardized Allowed Payment Amount Outcomes, by Clinical Episode and Overall, Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and ordered starting with the smallest relative decline in total payments. The Model 2 PGP Overall result includes the 21 clinical episodes that had sufficient volume to allow for risk-adjustment. These payment measures are not conditional upon the use of the service. The N for each clinical episode represents the number of BPCI episodes in the total payments calculation, while the Ns for the SNF and HHA payment calculations may differ slightly. UTI=urinary tract infection. COPD=chronic obstructive pulmonary disease. SPRI = simple pneumonia and respiratory infections. CHF=congestive heart failure. Hip & femur refers to hip & femur procedures except major joint. Medical non-infectious refers to medical non-infectious orthopedic. MJRLE=major joint replacement of the lower extremity. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level. § 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 2017) for BPCI and comparison providers.

Exhibit 13: Percent Change in Standardized Allowed Payment Amount Outcomes, by Clinical Episode and Overall, Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The 10 clinical episodes are ordered starting with the smallest relative decline in total payments. The Model 2 PGPs Overall result includes the 21 clinical episodes that had sufficient volume to allow for risk-adjustment. These payment measures are not conditional upon the use of the service. The N for each clinical episode represents the number of BPCI episodes in the total payments calculation, while the Ns for the SNF, IRF, and HHA payment calculations may differ slightly. Episode payments absent BPCI, or the counterfactual, is the BPCI baseline payment amount plus the change in episode payment amount for the comparison group. The counterfactual can be expressed as: BPCI before + (Comparison after – Comparison before). The percent change can then be expressed as: (BPCI after – Counterfactual) / (Counterfactual). COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections. CHF=congestive heart failure. UTI=urinary tract infection. Hip & femur refers to hip & femur procedures except major joint. Medical non-infectious refers to medical non-infectious orthopedic. MJRLE=major joint replacement of the lower extremity. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level. § 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 2017) for BPCI and comparison providers.

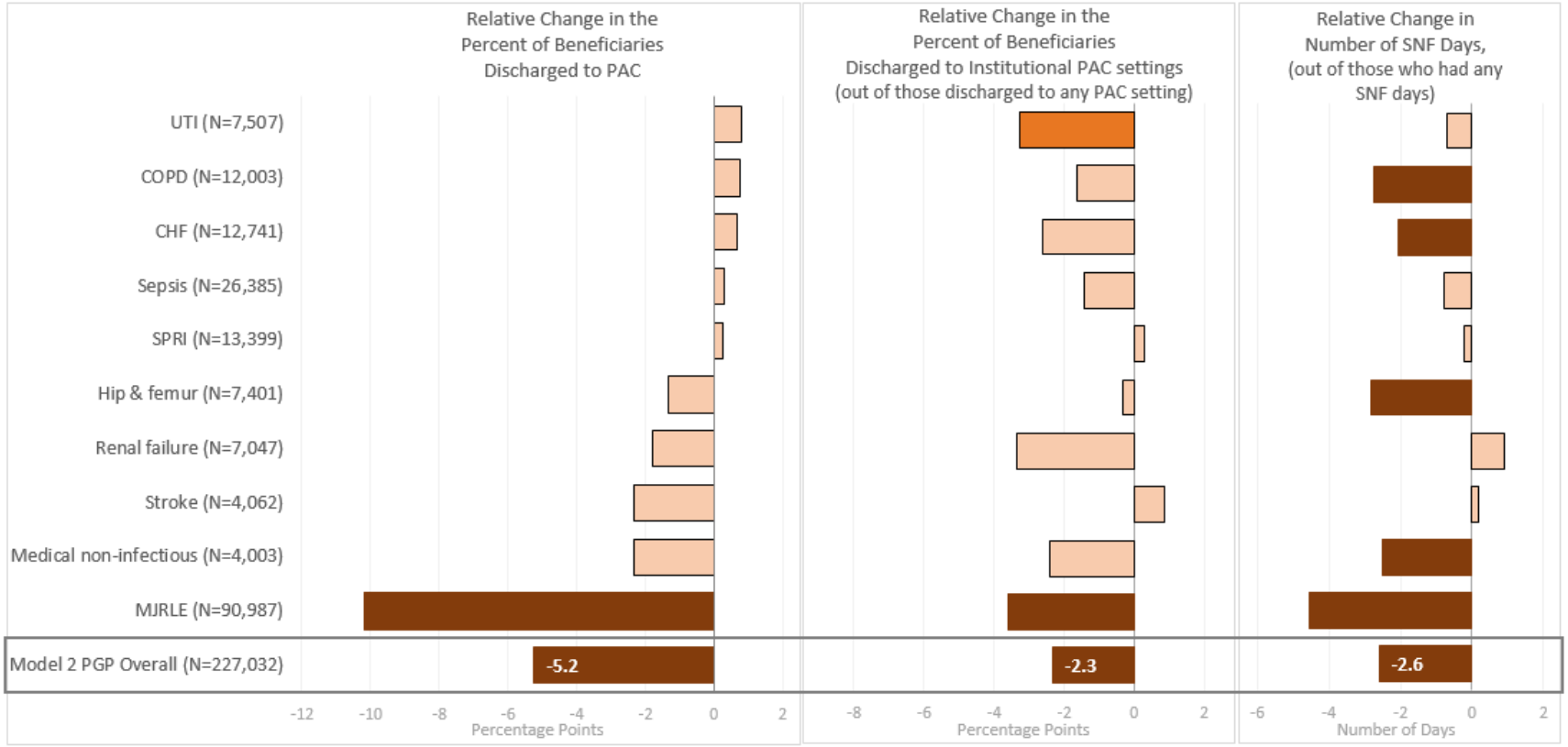
b. How have the services changed under BPCI?

The changes in institutional PAC use were consistent with the changes in payments. Among patients who were discharged to any PAC setting, the proportion discharged to institutional PAC settings declined in eight of the 10 clinical episodes, and the decline was statistically significant for two (see Exhibit 14). For Model 2 PGP-initiated episodes overall, the proportion of patients discharged to institutional PAC settings also declined by an estimated 2.3 percentage points more for BPCI episodes relative to the comparison group ($p<0.05$). For BPCI episodes with at least one day in SNF, the number of SNF days within the 90-day PDP declined relative to the comparison group in eight of the 10 clinical episodes, and the decline was statistically significant for five ($p<0.05$). (See Exhibit 14.) This result was also statistically significant for Model 2 PGP-initiated episodes overall, with an estimated 2.6 fewer SNF days during the 90-day PDP for BPCI episodes relative to the comparison group ($p<0.05$).

While there is evidence of lower institutional PAC use, the evidence was not as strong for lower use of any PAC services. Across the 10 PGP clinical episodes analyzed, the percentage of patients discharged to PAC settings decreased from the baseline to the intervention period relative to the comparison group for five clinical episodes (Exhibit 14). Major joint replacement of the lower extremity was the only clinical episode with a statistically significant change, with an estimated decrease of 10.2 percentage points more for BPCI episodes relative to the comparison group ($p<0.05$). For Model 2 PGP-initiated episodes overall, the proportion of BPCI patients discharged to PAC facilities decreased an estimated 5.2 percentage points more relative to the comparison group ($p<0.05$).

Similar to Model 2 hospital-initiated episodes, Model 2 PGP-initiated episodes had reductions in total payments due to shifting services from more expensive SNF and IRF settings to less expensive HHA and due to reduced SNF days for those with any SNF use.

Exhibit 14: Impact of BPCI on PAC Utilization, by Clinical Episode and Overall, Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model and ordered starting with the smallest relative decline in the percent of beneficiaries discharged to PAC. The Model 2 PGP Overall result includes the 21 clinical episodes that had sufficient volume to allow for risk-adjustment. The percent of beneficiaries discharged to institutional PAC settings is conditional on being discharged to any PAC setting, and the number of SNF days is conditional on having any SNF use. Therefore, the N for each clinical episode represents the number of BPCI episodes in the percent discharged to any PAC calculation, while the Ns for the percent discharged to institutional PAC settings and the number of SNF days differ. See **Appendix F** for sample size by outcome by clinical episode. UTI=urinary tract infection. COPD=chronic obstructive pulmonary disease. CHF=congestive heart failure. SPRI=simple pneumonia and respiratory infections. Hip & femur refers to hip & femur procedures except major joint. Medical non-infectious refers to medical non-infectious orthopedic. MJRLE=major joint replacement of the lower extremity. Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

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 2017) for BPCI and comparison providers.

c. How has quality of care changed under BPCI?

Results of the claims-based quality measures suggest that there is no clear pattern of change in the quality of care under BPCI for Model 2 PGP-initiated episodes overall or in the top 10 clinical episodes by volume (Exhibit 15). For Model 2 PGP-initiated episodes overall, there was no statistically significant change in the mortality rate or the unplanned readmission rate during the 90-day PDP, and there was a statistically significant relative decrease in ED use ($p<0.10$).

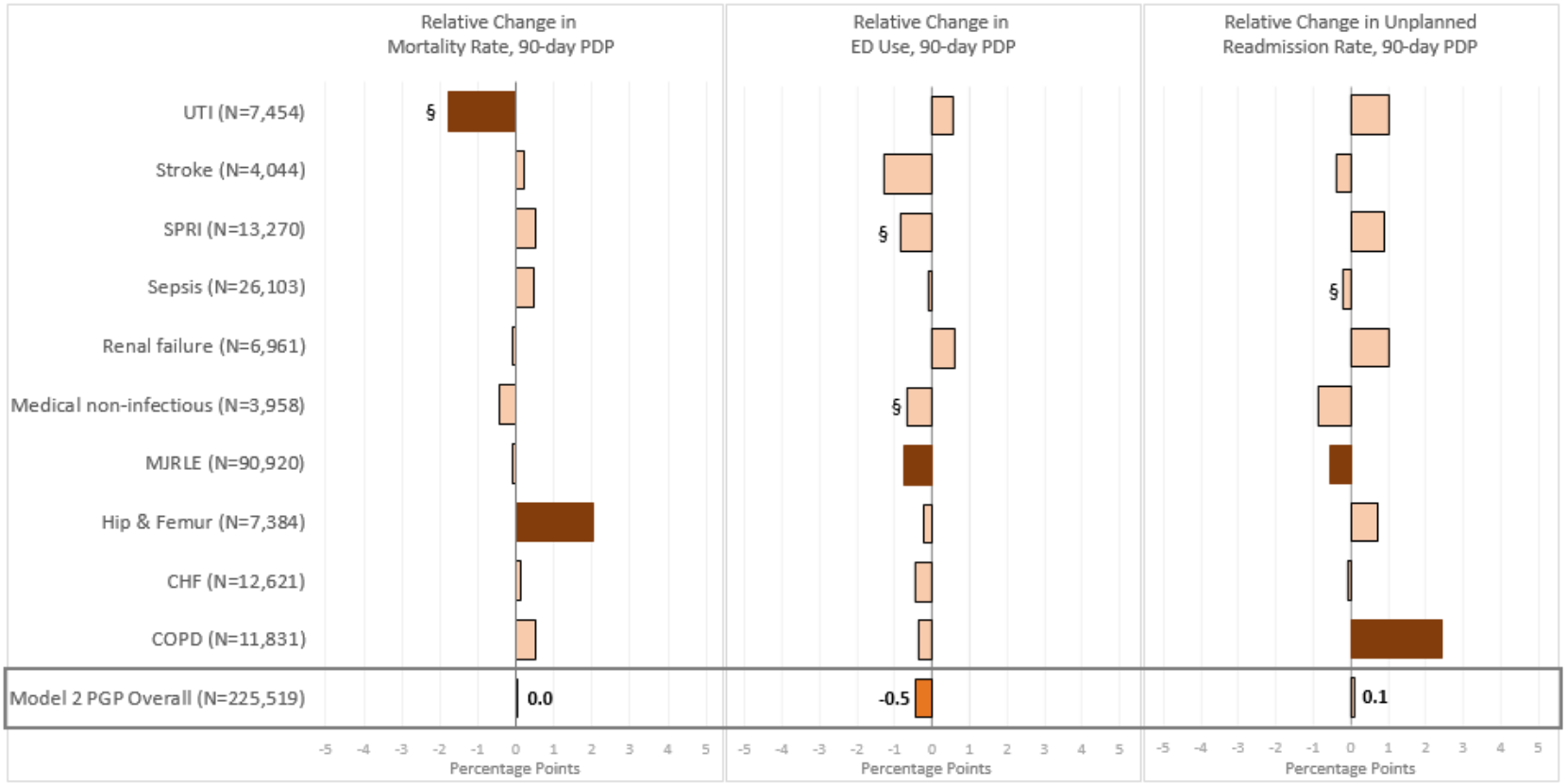
Across the three outcomes and 10 clinical episodes analyzed, there were five statistically significant results. For major joint replacement of the lower extremity episodes, there were statistically significant relative declines in ED use and unplanned readmission rates during the 90-day PDP ($p<0.05$). For chronic obstructive pulmonary disease episodes, there was a statistically significant relative increase in readmission rates ($p<0.05$).

In addition, there were two statistically significant results for the mortality rate, but one of these results did not meet the assumption of parallel trends and the second was not robust.[14] There was a decline in mortality for urinary tract infection episodes ($p<0.05$), but data from the baseline period shows that urinary tract infection episodes associated with BPCI PGPs and matched hospitals were not on parallel trends for mortality rates, which is required for an unbiased estimate. There was a statistically significant relative increase in the mortality rate for hip and femur procedures except major joint ($p<0.05$). However, this result was not robust to the inclusion of four additional quarters in the baseline period.[15]

[14] We conducted two additional analyses for each impact estimate. First, we tested whether the episodes associated with BPCI participants and matched hospitals were on parallel trends during the baseline period for the outcome, which is required for an unbiased estimate. Second, we tested the robustness of the impact estimates using a longer baseline period, including four additional quarters (Q4 2010 through Q3 2012), and using the same comparison group of providers.

[15] The two-year baseline time period (Q4 2010 through Q3 2012) resulted in an estimate that was smaller in magnitude and was not statistically significant.

Exhibit 15: Impact of BPCI on Claims-based Quality Outcomes, by Clinical Episode and Overall, Model 2 PGPs, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this table are the results of a difference-in-differences (DiD) model. The N for each clinical episode represents the number of BPCI episodes in the ED use and unplanned readmission calculations. PDP=post-discharge period. pp=percentage points. ED=emergency department. UTI=urinary tract infection. SPRI=simple pneumonia and respiratory infections. Medical non-infectious refers to medical non-infectious orthopedic. MJRLE=major joint replacement of the lower extremity. CHF=congestive heart failure. COPD=chronic obstructive pulmonary disease. Hip & femur refers to hip & femur procedures except major joint. Dark orange cells indicate DiD estimates that are statistically significant at the 5% level. Light orange cells indicate DiD estimates that are statistically significant at the 10% level.

§ 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 2017) for BPCI and comparison providers.

C. Comparison of Impact of BPCI among Hospitals and PGPs

To understand whether Model 2 participants differed in their approach to reducing episode payments, we compared Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes during the first four years of the initiative across the 21 clinical episodes in common. To ensure that the results were comparable between the two provider types, we used the same weights for both Model 2 hospitals and Model 2 PGP to aggregate clinical episode level estimates as described in the method section below.

1. Key Findings

Impact on Payments

- During the first four years of the BPCI initiative, total Medicare-allowed payments during the inpatient stay plus 90 days post discharge declined an estimated $891 ($p<0.05$) for Model 2 hospital-initiated episodes and an estimated $1,095 ($p<0.05$) for Model 2 PGP-initiated episodes, relative to their respective comparison groups, based on weighted averages of the 21 clinical episodes in common.
- Among surgical clinical episodes, the decline in total payments was larger for PGP episodes than for hospital episodes ($p<0.05$), while for medical clinical episodes, the decline was larger for hospital episodes than for PGP episodes, although this difference was not statistically significant.
- For both types of clinical episodes, the decline in total payments was driven mostly by reductions in SNF and IRF payments. For surgical clinical episodes, there was also a decline in HHA payments for PGP episodes ($p<0.05$), but not for hospital episodes. For medical clinical episodes, there was an increase in HHA payments for hospital episodes ($p<0.05$), but there was no statistically significant change for PGP episodes.

Impact on Post-acute Care Utilization

- The proportion of patients discharged to any PAC setting declined more for PGP episodes (4.1 percentage points, $p<0.05$) than hospital episodes (0.9 percentage points), and this difference between participant types was statistically significant.
- There was a slightly larger decline for hospital episodes than PGP episodes in the proportion of patients discharged to institutional PAC settings among those discharged to any PAC setting (3.3 percentage points and 2.3 percentage points, respectively, both $p<0.05$), but this difference between participant types was not statistically significant.
- Both participant types had similar declines in the number of SNF days among patients with at least one day in the SNF (2.5 percentage points and 2.4 percentage points for hospital and PGP episodes, respectively, both $p<0.05$). For surgical clinical episodes, PGPs had a greater reduction in the number of SNF days than hospitals (4.3 vs 2.3, respectively, both $p<0.05$), but for medical clinical episodes, hospitals had a greater reduction in SNF days than PGPs (2.6 vs 1.0 respectively, both $p<0.05$). The difference between participant types was statistically significant for surgical clinical episodes but it was not statistically significant for medical clinical episodes.

Impact on Quality

- In general, claims-based quality measures did not change under Model 2 for hospital- or PGP-initiated episodes.
- When examined by surgical and medical episodes, there was only one statistically significant result: the relative decline in ED use for surgical PGP episodes (0.7 percentage points, $p<0.05$).

2. Methods

To compare hospital-initiated and PGP-initiated episodes, we calculated volume-weighted participant-level impact estimates using the clinical episode estimates from the 21 clinical

episodes included in sections II.A and II.B. Because the distribution of episodes across clinical episodes differed for hospitals and PGPs, we used the same weight for both provider types (based on the combined episode volume for both types) and applied it to the clinical-episode level DiD estimates. (See Exhibit 16 below.) The weight is calculated as the number of combined intervention episodes from both hospitals and PGPs for each clinical episode divided by the total number of episodes across the 21 clinical episodes for both hospitals and PGPs. We also estimated the impact of BPCI on surgical and medical clinical episodes for hospital and PGP episodes using the same weighting methodology.

Exhibit 16: Distribution of Episodes across Clinical Episodes, Q4 2013 through Q3 2017

	Clinical Episode	Percent of Model 2 Hospital and Model 2 PGP Episodes as a Share of All Clinical Episodes	Percent of Model 2 Hospital and Model 2 PGP Episodes as a Share of Surgical Clinical Episodes or Medical Clinical Episodes
Surgical Clinical Episodes	Major joint replacement of the lower extremity	36%	84%
	Hip & femur procedures except major joint	3%	7%
	Percutaneous coronary intervention	2%	4%
	Spinal fusion (non-cervical)	1%	3%
	Major joint replacement of the upper extremity	1%	2%
Medical Clinical Episodes	Sepsis	10%	18%
	Congestive heart failure	10%	17%
	Simple pneumonia and respiratory infections	7%	12%
	Chronic obstructive pulmonary disease, bronchitis, asthma	6%	11%
	Stroke	3%	6%
	Urinary tract infection	3%	5%
	Renal failure	3%	5%
	Medical non-infectious orthopedic	2%	4%
	Acute myocardial infarction	2%	3%
	Other respiratory	2%	3%
	Cardiac arrhythmia	2%	3%
	Cellulitis	2%	3%
	Esophagitis, gastroenteritis and other digestive disorders	2%	3%
	Gastrointestinal hemorrhage	1%	2%
	Nutritional and metabolic disorders	1%	2%
	Gastrointestinal obstruction	1%	1%

Note: The clinical episodes in this exhibit are restricted to the 21 clinical episodes in common between the Model 2 hospital and Model 2 PGP impact analyses. Percentages are calculated as a share of all 21 clinical episodes listed and as a share of the five surgical clinical episodes or 16 medical clinical episodes listed. Clinical episodes within the surgical and medical categories are ordered by the number of total intervention episodes.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q3 2017 for BPCI providers.

3. Payment, Utilization, and Quality Outcomes

This section presents the BPCI impact estimates for payments, utilization, and quality during the first four years of the initiative for the clinical episodes that Model 2 hospitals and Model 2 PGPs have in common. We also present the results by surgical and medical clinical episodes. Detailed results are located in **Appendix G**.

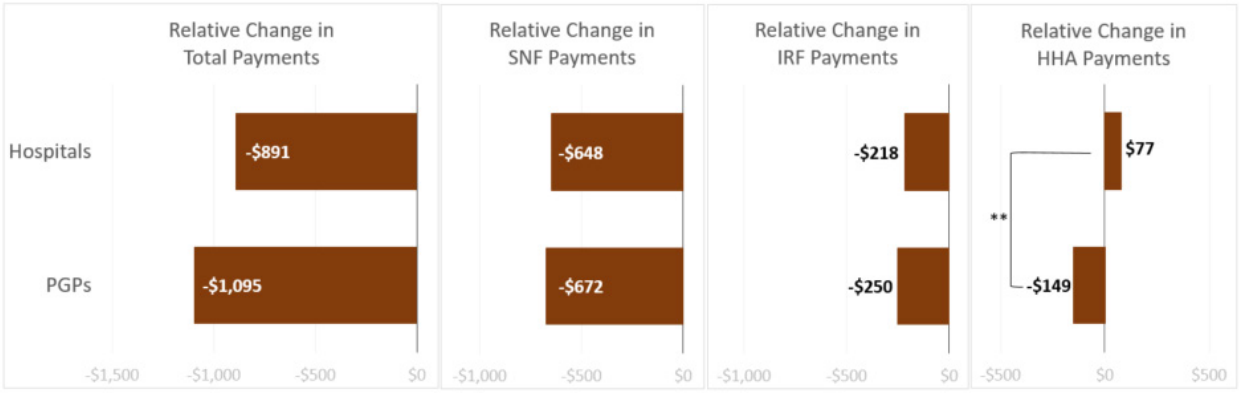
a. How have the average standardized payments changed under BPCI for Model 2 Hospitals and PGPs?

During the first four years of BPCI, the total standardized allowed payment amount for the inpatient stay plus the 90-day PDP declined from baseline to the intervention period relative to the comparison group for both hospital-initiated episodes and PGP-initiated episodes. Total payments declined an estimated $891 more for hospital episodes than for comparison episodes ($p<0.05$) and an estimated $1,095 more for PGP episodes ($p<0.05$). (See Exhibit 17.) This translates to a percent decline of 3.3% for hospital episodes and 4.1% for PGP episodes relative to what payments would have been absent the BPCI initiative (Exhibit 18).

For both Model 2 participant types, the decline in SNF and IRF payments during the 90-days post discharge contributed to the decline in total payments. SNF payments declined an estimated $648, or 12.0%, for hospital episodes ($p<0.05$) and an estimated $672, or 14.9%, for PGP episodes ($p<0.05$), relative to episodes in their respective comparison groups (Exhibits 17 and 18). IRF payments declined by an estimated relative $218, or 16.7%, for hospital episodes ($p<0.05$) and by an estimated relative $250, or 20.5%, for PGP episodes ($p<0.05$).

While the declines in SNF and IRF payments were comparable for both participant types, the relative change in HHA payments were different ($p<0.05$). HHA payments increased by an estimated relative $77, or 5.6%, for hospital episodes ($p<0.05$), while they declined by an estimated relative $149, or 7.0%, for PGP episodes ($p<0.05$). (See Exhibits 17 and 18.)

Exhibit 17: Impact of BPCI on Standardized Allowed Payment Amount Outcomes, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



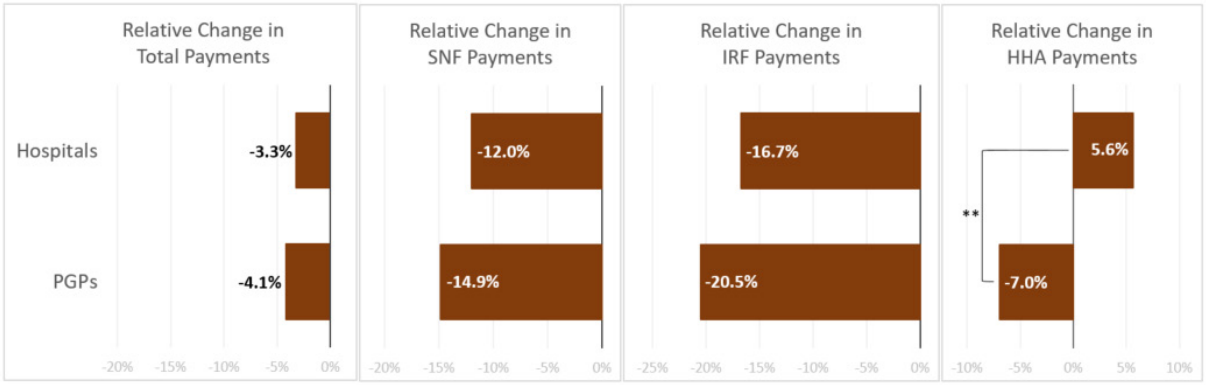
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. They are calculated as a volume-weighted average of the 21 clinical episodes included in the impact estimates for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes in sections II.A and II.B. (See Appendix C for the list of clinical episodes in common.) Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

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 2017) for BPCI and comparison providers.

Exhibit 18: Percent Change in Standardized Allowed Payment Amount Outcomes, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



Note: Episode payments absent BPCI, or the counterfactual, is the BPCI baseline payment amount plus the change in episode payment amount for the comparison group. The counterfactual can be expressed as: BPCI before + (Comparison after – Comparison before). The percent change can then be expressed as: (BPCI after – Counterfactual) / (Counterfactual). Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

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 2017) for BPCI and comparison providers.

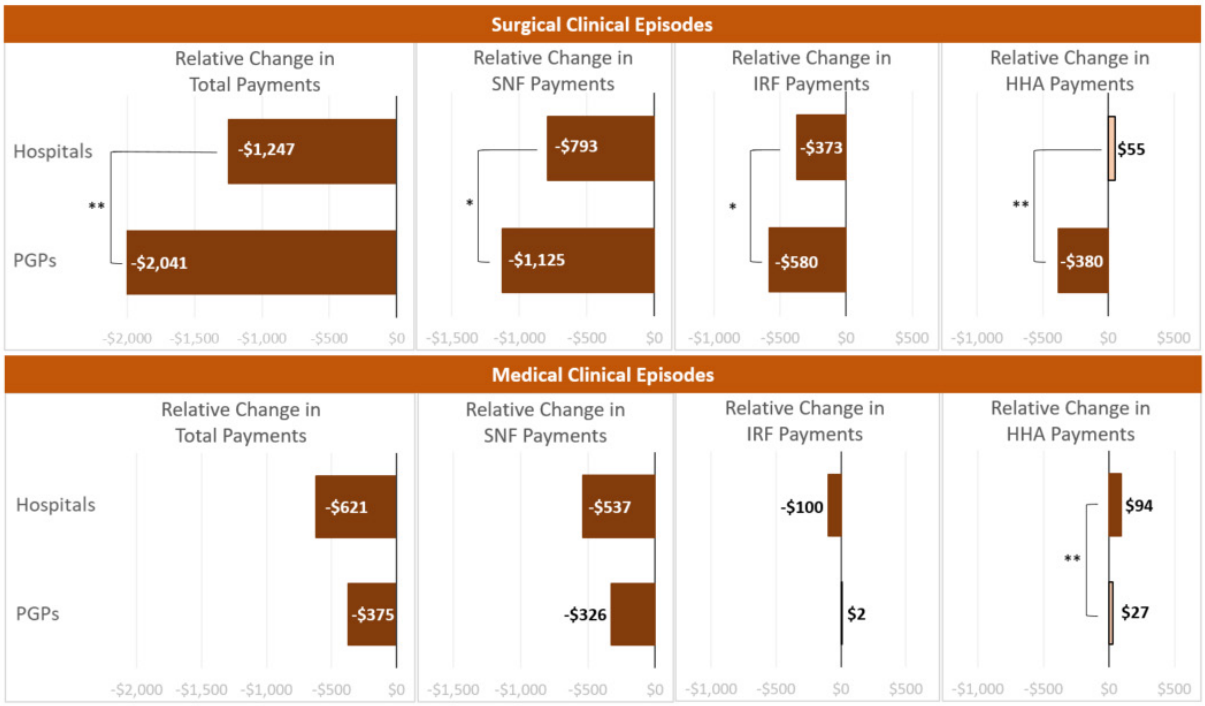
When the analysis was grouped by surgical and medical clinical episodes, total payments for hospital episodes and PGP episodes declined for both participant types, but the magnitudes of these declines differed. For surgical clinical episodes, the relative decline in total payments was larger for PGP episodes than for hospital episodes ($p<0.05$). For medical clinical episodes, the relative decline was larger for hospital episodes than PGP episodes, although the difference was not statistically significant. For surgical episodes, total payments declined an estimated $2,041 for PGP episodes ($p<0.05$) and an estimated $1,247 for hospital episodes ($p<0.05$) relative to their comparison groups (Exhibit 19). This translates to a percent decline of 7.7% for PGP surgical episodes and 4.3% for hospital surgical episodes relative to what payments would have been absent the BPCI initiative (Exhibit 20). For medical episodes, total payments declined an estimated $621, or 2.4%, for hospital episodes ($p<0.05$) and an estimated $375, or 1.5%, for PGP episodes relative to their comparison groups ($p<0.05$).

Similar to total payments, the impact of BPCI on SNF, IRF, and HHA payments in the 90-day post discharge period differed in magnitude by surgical and medical clinical episodes, but the direction of the impact was generally similar. The relative changes in SNF and IRF payments were larger for PGP surgical episodes than hospital surgical episodes, while the relative changes were not statistically significantly different for hospital and PGP medical episodes. For surgical clinical episodes, SNF and IRF payments declined more for PGP episodes than for hospital episodes, and these differences were statistically significant ($p<0.10$); HHA payments also declined for PGP

episodes but did not change for hospital episodes (p<0.05). (See Exhibit 19.) SNF payments declined an estimated $1,125, or 25.3%, for PGP surgical episodes (p<0.05) and an estimated $793, or 14.2%, for hospital surgical episodes (p<0.05), relative to their comparison groups (Exhibits 19 and 20). IRF payments also declined by an estimated $580, or 52.8%, for PGP surgical episodes (p<0.05) and by an estimated $373, or 26.3%, for hospital surgical episodes (p<0.05), relative to their comparison groups. HHA payments declined an estimated $380, or 18.8%, for PGP surgical episodes (p<0.05), while there was no statistically significant change for hospital surgical episodes relative to their comparison groups (Exhibits 19 and 20).

For medical clinical episodes, SNF and IRF payments declined and HHA payments increased for both participant types. These changes were larger for hospital episodes than PGP episodes, but the difference between participant types was only statistically significant for the HHA payment measure. SNF payments declined an estimated $537, or 10.4%, for hospital medical episodes (p<0.05) and an estimated $326, or 7.0%, for PGP medical episodes (p<0.05), relative to their comparison groups (Exhibits 19 and 20). IRF payments declined an estimated $100, or 9.4%, for hospital medical episodes (p<0.05), while there was no statistically significant change for PGP medical episodes. HHA payments increased an estimated $94, or 7.8%, for hospital medical episodes (p<0.05), while there was no statistically significant change for PGP medical episodes (Exhibits 19 and 20).

Exhibit 19: Impact of BPCI on Standardized Allowed Payment Amount Outcomes for Surgical and Medical Clinical Episodes, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



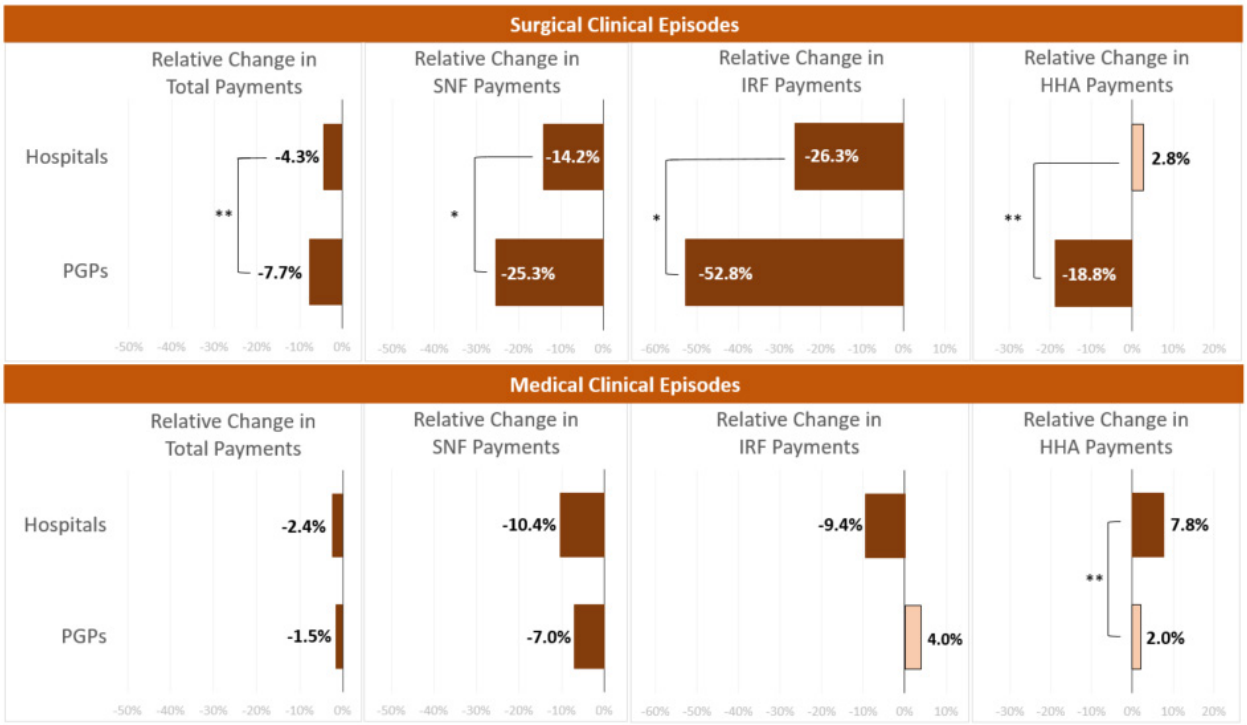
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. They are calculated as a volume-weighted average of the 21 clinical episodes included in the impact estimates for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes in sections II.A and II.B. (See Appendix C for the list of clinical episodes in common.) Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

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 2017) for BPCI and comparison providers.

Exhibit 20: Percent Change in Standardized Allowed Payment Amount Outcomes for Surgical and Medical Clinical Episodes, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



Note: Episode payments absent BPCI, or the counterfactual, is the BPCI baseline payment amount plus the change in episode payment amount for the comparison group. The counterfactual can be expressed as: BPCI before + (Comparison after – Comparison before). The percent change can then be expressed as: (BPCI after – Counterfactual) / (Counterfactual). Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

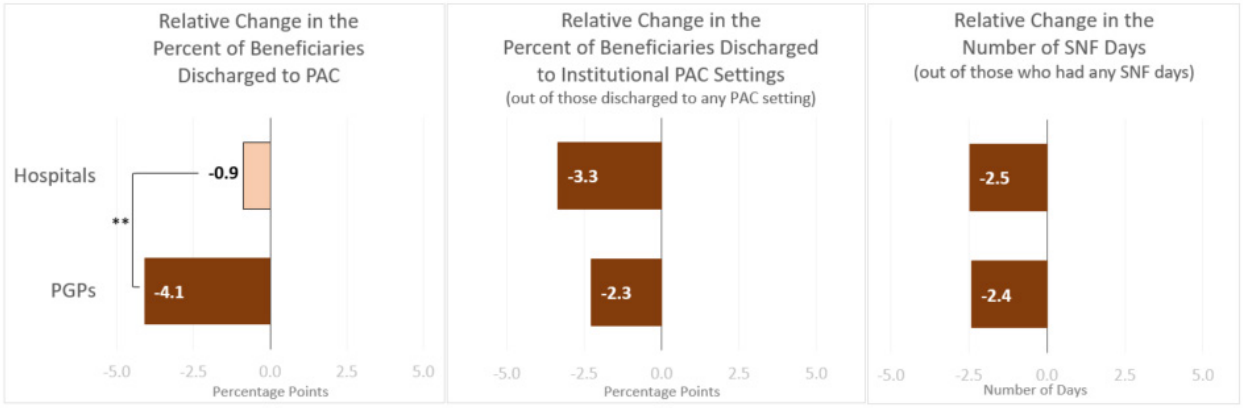
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2017) for BPCI and comparison providers.

b. How have the services changed under BPCI for Model 2 Hospitals and PGPs?

In general, there was a relative decline in service use for both hospital-initiated and PGP-initiated episodes and the changes were consistent with the changes in PAC payments. The relative decline in the proportion of patients discharged to any PAC setting (SNF, IRF, long term care hospital, and HHA) was larger for PGP episodes than for hospital episodes ($p<0.05$). For PGP episodes, the percent of beneficiaries discharged to any PAC setting declined by an estimated 4.1 percentage points ($p<0.05$), while the decline for hospital episodes was not statistically significant (Exhibit 21). The proportion of patients discharged to institutional PAC settings, among those discharged to any PAC setting, declined for both hospital and PGP episodes. While the decline was larger for

hospital episodes, the difference between the two participant types was not statistically significant. For hospital episodes, the proportion of patients discharged to institutional PAC settings declined an estimated 3.3 percentage points ($p<0.05$), and the proportion declined an estimated 2.3 percentage points for PGP episodes ($p<0.05$). (See Exhibit 21.) The number of SNF days within the 90-day PDP, for episodes with at least one day in the SNF, declined for both hospital and PGP episodes, and the decline was similar for both participant types (2.5 days and 2.4 days, respectively, both $p<0.05$).

Exhibit 21: Impact of BPCI on Post-acute Care Utilization, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. They are calculated as a volume-weighted average of the 21 clinical episodes included in the impact estimates for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes in sections II.A and II.B. (See Appendix C for the list of clinical episodes in common.) Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

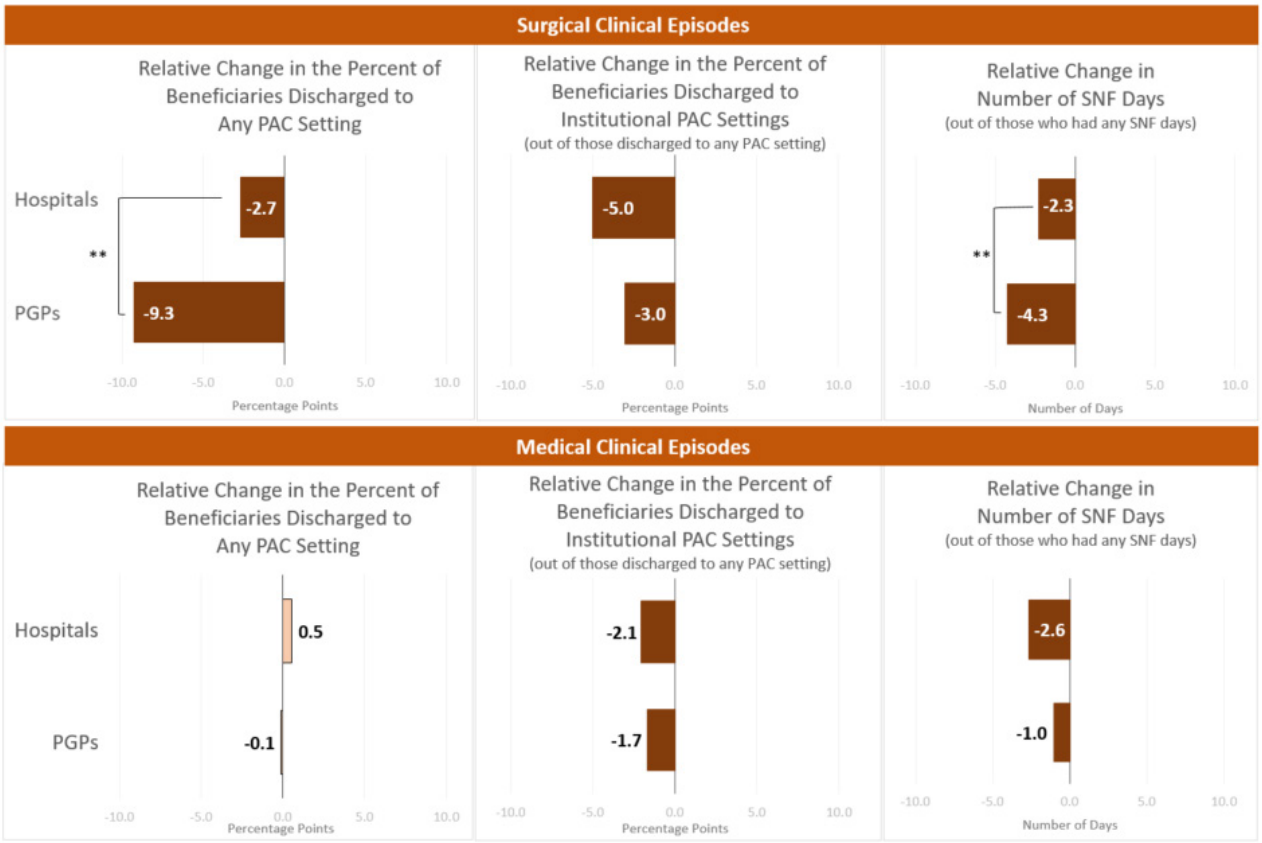
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2017) for BPCI and comparison providers.

There was a decline in PAC use for surgical episodes for both hospitals and PGPs, but the magnitude of the decline differed between the participant types. The relative decline in the proportion of patients discharged to any PAC setting was greater for PGP surgical episodes than for hospital surgical episodes ($p<0.05$). For PGP surgical episodes, the proportion declined an estimated 9.3 percentage points ($p<0.05$), while for hospital surgical episodes, the proportion declined an estimated 2.7 percentage points ($p<0.05$). The relative decline in the proportion of patients discharged to institutional PAC settings, of those discharged to any PAC setting, was greater for hospital surgical episodes than PGP surgical episodes (5.0 percentage points and 3.0 percentage points respectively, both $p<0.05$). (See Exhibit 22.) However, the difference between the two participant types was not statistically significant. The number of SNF days within the 90-day PDP, for episodes with at least one day in the SNF, declined for both hospital and PGP surgical episodes, but the decline was larger for PGP surgical episodes ($p<0.05$). For PGP surgical episodes, there was an estimated decline of 4.3 days ($p<0.05$) while for hospital surgical episodes there was an estimated decline of 2.3 days ($p<0.05$).

The impact of BPCI on PAC use for medical episodes was similar for hospital and PGP episodes, with no statistically significant differences between the two participant types across outcomes.

Unlike surgical episodes, for medical episodes there was no statistically significant change in the proportion of patients discharged to any PAC setting for either participant type. However, the proportion of patients discharged to institutional PAC settings, of those discharged to any PAC setting, declined for both hospital medical episodes and PGP medical episodes by similar amounts (2.1 percentage points for hospital medical episodes and 1.7 percentage points for PGP medical episodes, both $p<0.05$). (See Exhibit 22.) The number of SNF days within the 90-day PDP, for episodes with at least one day in the SNF, also declined for both hospital and PGP medical episodes (2.6 days and 1.0 days, respectively, both $p<0.05$).

Exhibit 22: Impact of BPCI on Post-acute Care Utilization for Surgical and Medical Clinical Episodes, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. They are calculated as a volume-weighted average of the 21 clinical episodes included in the impact estimates for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes in sections II.A and II.B. (See Appendix C for the list of clinical episodes in common.) Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

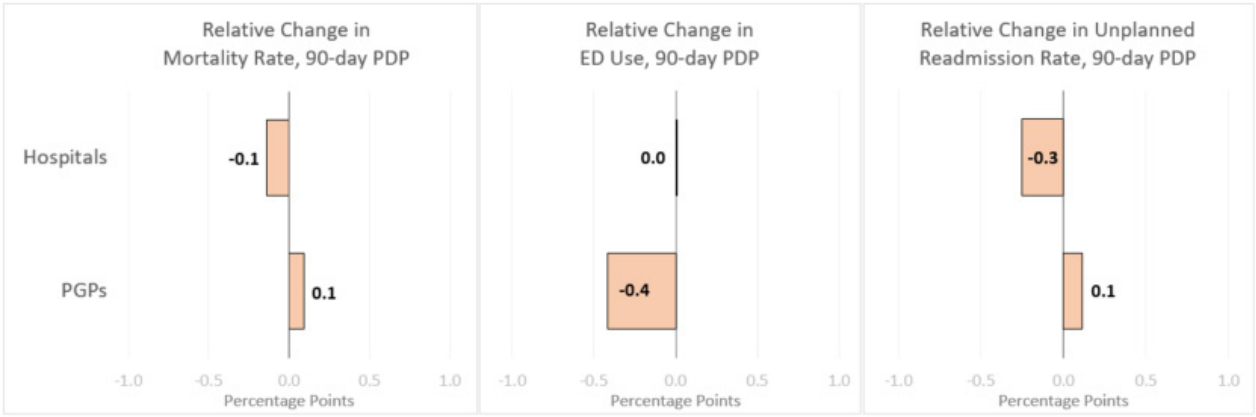
Source: Lewin analysis of Medicare claims and enrollment data for the baseline period (Q4 2011 through Q3 2012) and the intervention period (Q4 2013 through Q3 2017) for BPCI and comparison providers.

c. How has quality of care changed under BPCI for Model 2 Hospitals and PGPs?

Generally, there was a reduction in total payments, PAC payments, and PAC utilization for both hospitals and PGPs. Despite the differences in magnitude of the results between the two participant types in total payments, PAC payments, and PAC utilization, results of the claims-based quality measures suggest that the quality of care generally did not change under BPCI for either participant type. There was no statistically significant change for hospital-initiated or PGP-initiated episodes relative to their comparison groups for mortality rate, ED use, or unplanned readmission rate in the 90-day PDP (Exhibit 23). Similarly, there was no statistically significant difference between the two participant types for any of the three quality outcomes.

Furthermore, there were minimal differences in claims-based quality measures when examined by surgical and medical clinical episodes. There was only one statistically significant difference between hospital surgical episodes and PGP surgical episodes. For PGP surgical episodes, ED use declined by an estimated 0.7 percentage points more for BPCI relative to the comparison group ($p<0.05$), while there was no statistically significant change for hospital surgical episodes relative to the comparison group (Exhibit 24).

Exhibit 23: Impact of BPCI on Claims-based Quality Outcomes, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



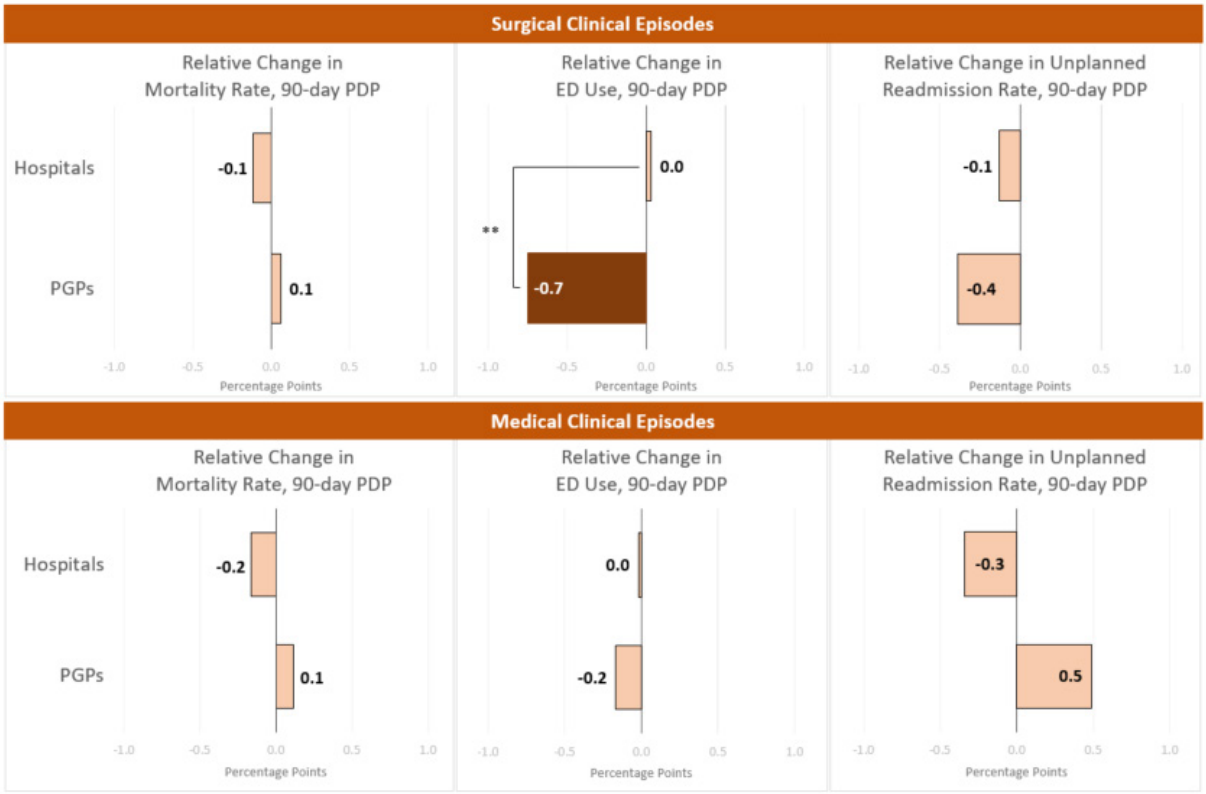
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. They are calculated as a volume-weighted average of the 21 clinical episodes included in the impact estimates for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes in sections II.A and II.B. (See Appendix C for the list of clinical episodes in common.) Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

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 2017) for BPCI and comparison providers.

Exhibit 24: Impact of BPCI on Claims-based Quality Outcomes for Surgical and Medical Clinical Episodes, Model 2 Hospitals and Model 2 PGPs, Clinical Episodes in Common, Baseline to Intervention, Q4 2013 – Q3 2017



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. They are calculated as a volume-weighted average of the 21 clinical episodes included in the impact estimates for both Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes in sections II.A and II.B. (See Appendix C for the list of clinical episodes in common.) Dark orange bars indicate DiD estimates are statistically significant at the 5% level. Light orange bars indicate DiD estimates are statistically significant at the 10% level.

* Indicates the difference between the hospital and PGP estimates is statistically significant at the 10% level.

** Indicates the difference between the hospital and PGP estimates is statistically significant at the 5% level.

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 2017) for BPCI and comparison providers.

III. Net Savings to the Medicare Program

BPCI Model 2 and Model 3 participants reduced Medicare payments across a variety of clinical episodes.[16] However, payment reductions may not translate into savings to Medicare because they do not account for the reconciliation payments that Medicare paid to or recovered from participants. The reconciliation process, in which actual spending was compared with a discounted benchmark or target price, was the mechanism for providing incentives to participants to lower episode spending while achieving savings for Medicare. For each clinical episode, CMS created a participant-specific benchmark by updating historical episode payments to the intervention year by national spending trends. This benchmark represented what payments would have been absent BPCI. The benchmark was then discounted by 2–3% to become the target price. Participants with actual episode payments below their target price received the difference as reconciliation payments. Conversely, participants with actual episode payments above their target price repaid the difference to CMS. This reconciliation process was intended to ensure that CMS achieved savings equal to the discount applied to the benchmark. However, CMS eliminated downside risk for some episodes over the course of the BPCI initiative and did not require participants to make repayments when payments were above the target price.[17]

This section presents two sets of estimates of Medicare savings based on the change in aggregate payments accounting for reconciliation payments. One set of estimates uses actual reconciliation payments paid to and received from participants in which CMS eliminated repayment responsibility ("with downside risk eliminated" or "as implemented"). The other set of estimates uses the hypothetical reconciliation payments that would have been paid to and received from participants if CMS had not retrospectively eliminated repayment responsibility ("with downside risk not eliminated" or "as originally designed").

[16] See the Model 2 Results chapter for payment estimates for the top 10 clinical episodes with the largest volume under Model 2 through Q3 2017. See the Year 5 annual report for payment estimates for all remaining Model 2 clinical episodes and for Model 3 clinical episodes through Q4 2016. The report is available for download from https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

[17] In November 2014, CMS eliminated downside risk for episodes initiated between Q4 2013 and Q4 2014 due to inaccurate target prices. In July 2016, CMS also eliminated downside risk for any episode of care initiated in 2015 as a result of the episode attribution issues caused by incorrect PGP Reassignment Lists, and, in December 2017, CMS extended this elimination of downside risk for all such episodes initiated through Q3 2016.

A. Key Findings

Model 2

- After accounting for reconciliation payments through Q3 2017, BPCI Model 2 resulted in an estimated loss of 0.8% of the benchmark ($197 million) to Medicare, ranging from a loss of 0.1% to 1.5% ($14 million to $380 million, 90% confidence interval).
- Had CMS not eliminated downside risk and had required participants to return funds when payments were above the target, reconciliation payments would have been lower, and Medicare would have achieved an estimated savings of 0.6% of the benchmark ($152 million) for Model 2, ranging from a loss of 0.1% ($31 million) to a savings of 1.3% ($335 million, 90% confidence interval).

Model 3

- After accounting for reconciliation payments through Q3 2017, BPCI Model 3 resulted in an estimated loss of 3.6% of the benchmark ($100 million) to Medicare, ranging from a loss of 2.0% to 5.2% ($56 million to $144 million, 90% confidence interval).
- Had CMS not eliminated downside risk, reconciliation payments would have been lower, and the estimated loss to Medicare would have been 2.2% of the benchmark ($62 million), ranging from a loss of 3.8% to 0.7% ($19 million to $106 million, 90% confidence interval).

B. Methods

Net savings to Medicare for both Model 2 and Model 3 were defined as the difference between non-standardized paid amounts[18] and reconciliation payments made to or received from BPCI participants following the formula below:

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

The change in aggregate non-standardized payments is approximated by multiplying the estimates from the difference-in-difference (DiD) model, which estimates the change in per-episode standardized Medicare paid amounts during the inpatient stay and 90-day PDP, by a standardized to non-standardized conversion factor. For each model, the per-episode change in standardized payments was multiplied by the total number of BPCI episodes in the first four years of the initiative and then converted to non-standardized payments.[19] We also present the estimated savings to Medicare, the change in non-standardized payments, and reconciliation payments as a

[18] 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 actual 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.

[19] 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.

percentage of the benchmark. 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 savings to Medicare in two ways. We present results for Model 2 and Model 3 overall as the program was implemented, in which CMS waived repayment responsibility. We also calculate the savings to Medicare as the program was initially designed, in which CMS had not retrospectively eliminated repayment responsibility. We present these results at the Model 2 and Model 3 overall levels, as well as at the clinical episode level for the top 10 Model 2 hospital and Model 2 PGP clinical episodes. We present the highest level of statistical significance for the impact estimates, $p<0.05$ or $p<0.10$.

C. Results

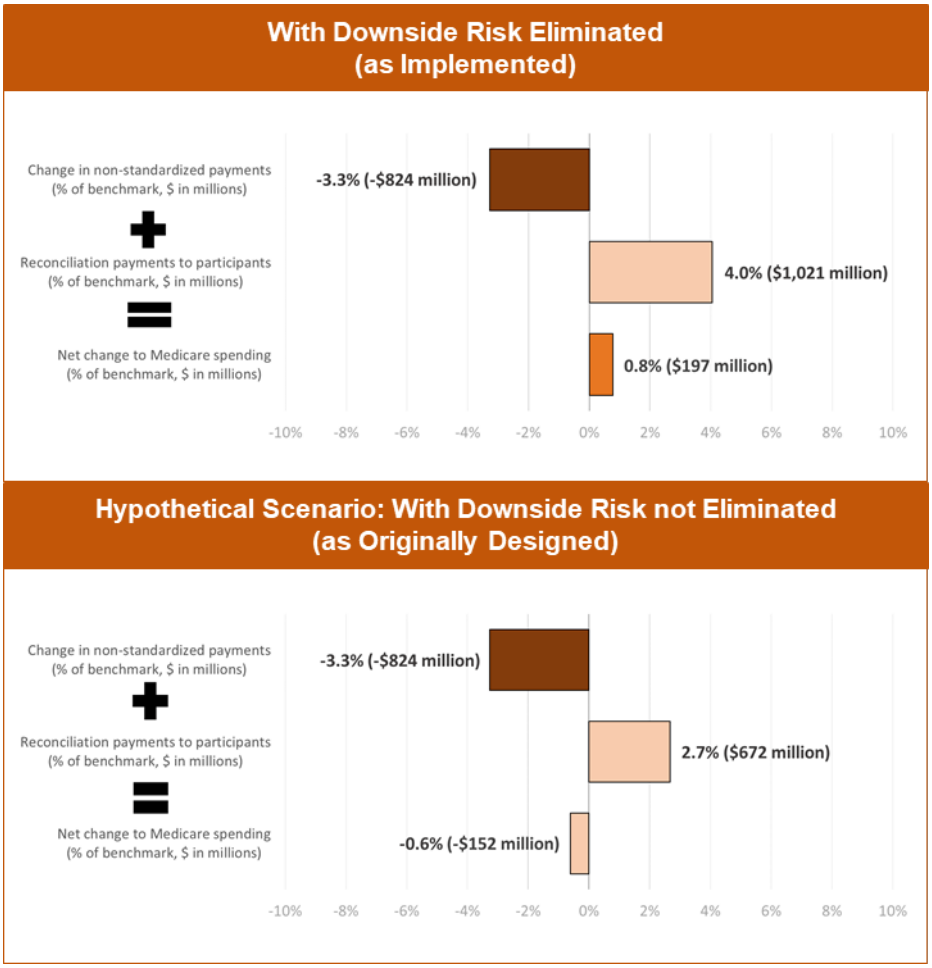
1. Model 2

As described above, BPCI was designed to achieve savings of 2-3% of the benchmark. However, during the first four years of the initiative, Model 2 was associated with an estimated loss to Medicare of 0.8% of the benchmark ($196,937,863), ranging from a loss of 1.5% to 0.1% ($13,960,440 to $379,915,286, $p<0.10$) (see Exhibit 25). The net loss is equivalent to a per-episode loss to Medicare of $199. While aggregate non-standardized payments declined an estimated 3.3% of the benchmark ($823,968,101) across all 990,488 BPCI Model 2 episodes,[20] reconciliation payments for Model 2 through Q3 2017 were 4.0% ($1,020,905,964), which outweighed the reductions achieved in aggregate payments.

If CMS had not eliminated downside risk for some episodes, Medicare likely would have realized savings. In this scenario, reconciliation payments would have fallen from 4.0% to 2.7% of the benchmark ($671,500,685). Subtracting this amount from the reduction in aggregate non-standardized payments results in an estimated savings to Medicare of 0.6% ($152,467,416, Exhibit 25), which is equivalent to a per-episode savings to Medicare of $154. This estimate is not statistically significant at the 10% level, since the confidence interval ranges from a loss of 0.1% ($30,510,007) to a savings of 1.3% ($335,444,840).

[20] Model 2 hospitals initiated 522,936 episodes through September 30, 2017, and Model 2 PGPs initiated 467,552.

Exhibit 25: Estimated Change in Medicare Spending with and without Downside Risk Eliminated, Model 2, Q4 2013 – Q3 2017



***Note*:** The estimates of the change in aggregate non-standardized payments are from difference-in-differences (DiD) models of standardized Medicare paid amounts during the anchor stay and 90-day PDP. "With downside risk eliminated" depicts estimates of 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 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. Dark orange indicates estimates that are statistically significant at the 5% level. Light orange cells indicate estimates that are statistically significant at the 10% level.

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 2017) for BPCI and comparison providers and CMS data on reconciliation payments.

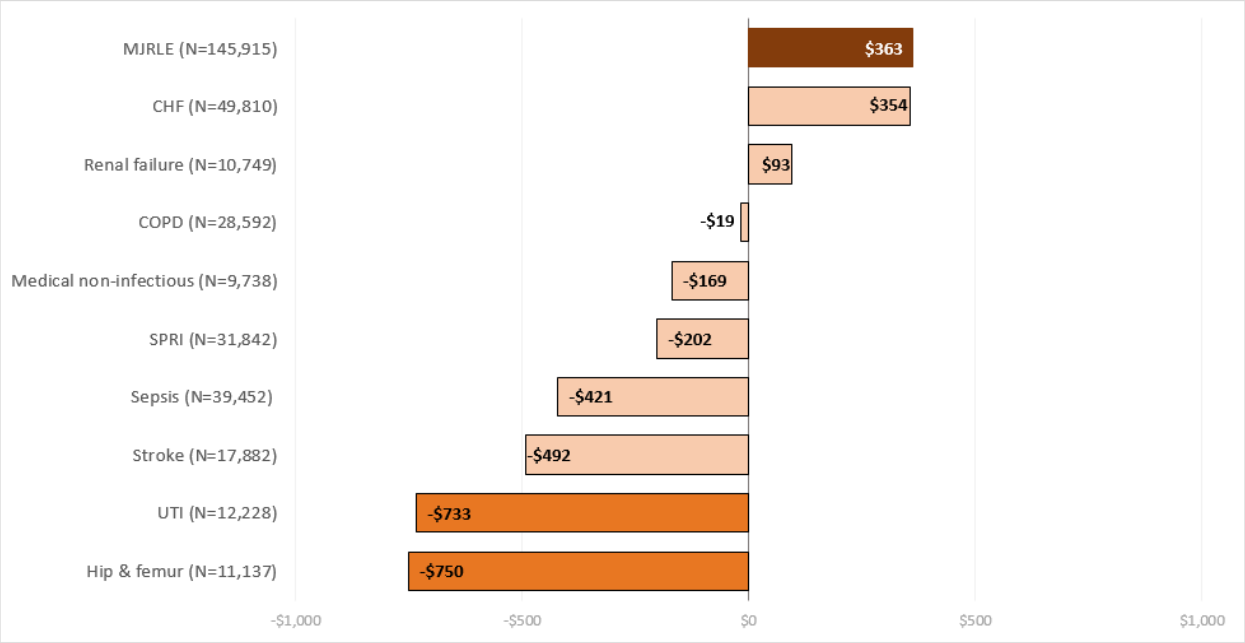
a. Hospitals

Across the 10 Model 2 hospital clinical episodes examined, net savings to Medicare ranged from a loss of $750 to a savings of $363 per episode under the hypothetical scenario that downside risk was not eliminated (Exhibit 26).[21] There was a per-episode loss to Medicare among seven of the 10 clinical episodes and the loss was statistically significant for two episodes. Among the three

[21] The range estimating the loss to Medicare is based on a 90% confidence interval.

clinical episodes with per-episode savings to Medicare, the estimate was statistically significant for one: major joint replacement of the lower extremity ($p<0.05$). Detailed results are located in **Appendix H**.

Exhibit 26: Estimated Per-Episode Savings to Medicare in the Hypothetical Scenario with Downside Risk not Eliminated (as Originally Designed), by Clinical Episode, Model 2 Hospitals, Q4 2013 – Q3 2017



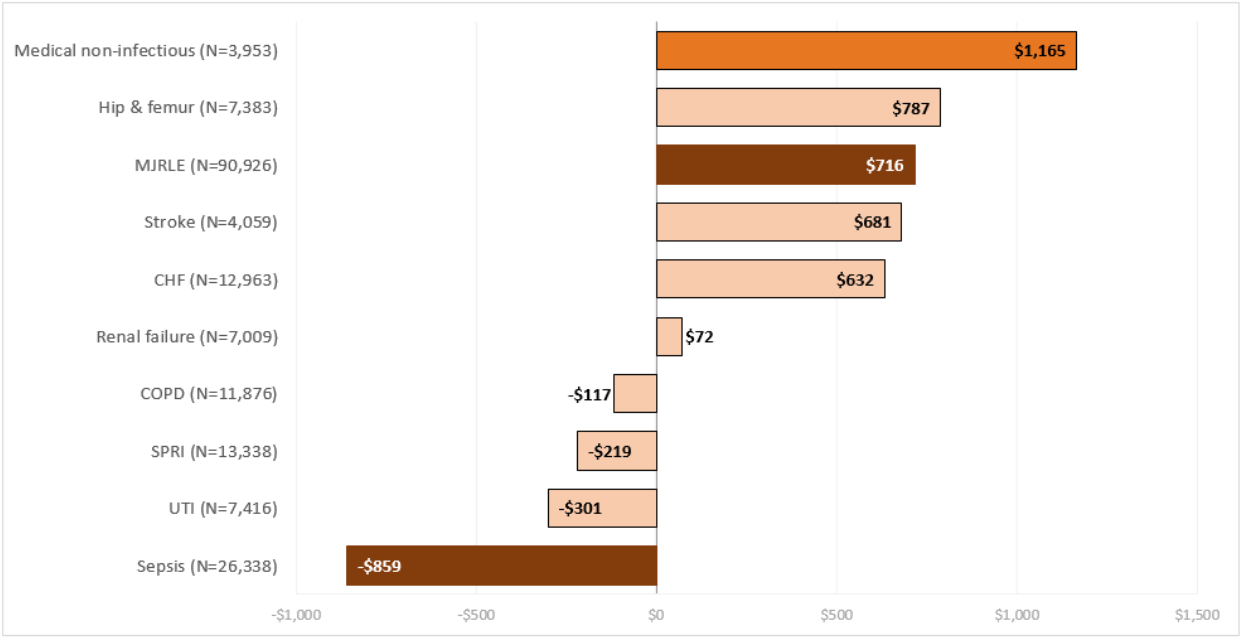
***Note*:** The estimates depict savings to Medicare in the hypothetical scenario where the model was implemented as designed, i.e., repayment responsibility was not eliminated. Net savings to Medicare is the difference between the change in aggregate non-standardized payments based on a difference-in-differences (DiD) model and reconciliation payments. CHF=congestive heart failure. MJRLE=major joint replacement of the lower extremity. COPD=chronic obstructive pulmonary disease. Medical non-infectious refers to medical non-infectious orthopedic. SPRI=simple pneumonia and respiratory infections. UTI=urinary tract infection. Hip & femur refers to hip & femur procedures except major joint. Dark orange indicates estimates that are statistically significant at the 5% level. Light orange indicates estimates that are statistically significant at the 10% level.

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 2017) for BPCI and comparison providers and CMS data on reconciliation payments.

b. PGPs

Across the 10 Model 2 PGP clinical episodes examined, the net savings to Medicare ranged from a loss of $859 to a savings of $1,165 per episode under the hypothetical scenario that downside risk was not eliminated (Exhibit 27). There was a per-episode loss to Medicare among four of the 10 clinical episodes. The loss was statistically significant for sepsis only ($p<0.05$). Among the six clinical episodes with savings to Medicare, the estimates were statistically significant for two clinical episodes: major joint replacement of the lower extremity ($p<0.05$) and medical non-infectious orthopedic ($p<0.10$). Detailed results are located in **Appendix H**.

Exhibit 27: Estimated Per-Episode Savings to Medicare in the Hypothetical Scenario with Downside Risk not Eliminated (as Originally Designed), by Clinical Episode, Model 2 PGPs, Q4 2013 – Q3 2017



***Note*:** The estimates depict savings to Medicare in the hypothetical scenario where the model was implemented as designed, i.e., repayment responsibility was not eliminated. Net savings to Medicare is the difference between the change in aggregate non-standardized payments based on a difference-in-differences (DiD) model and reconciliation payments. Medical non-infectious refers to medical non-infectious orthopedic. Hip & femur refers to hip & femur procedures except major joint. MJRLE=major joint replacement of the lower extremity. CHF=congestive heart failure. COPD=chronic obstructive pulmonary disease. SPRI=simple pneumonia and respiratory infections. UTI=urinary tract infection. Dark orange indicates estimates that are statistically significant at the 5% level. Light orange indicates estimates that are statistically significant at the 10% level.

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 2017) for BPCI and comparison providers and CMS data on reconciliation payments.

2. Model 3

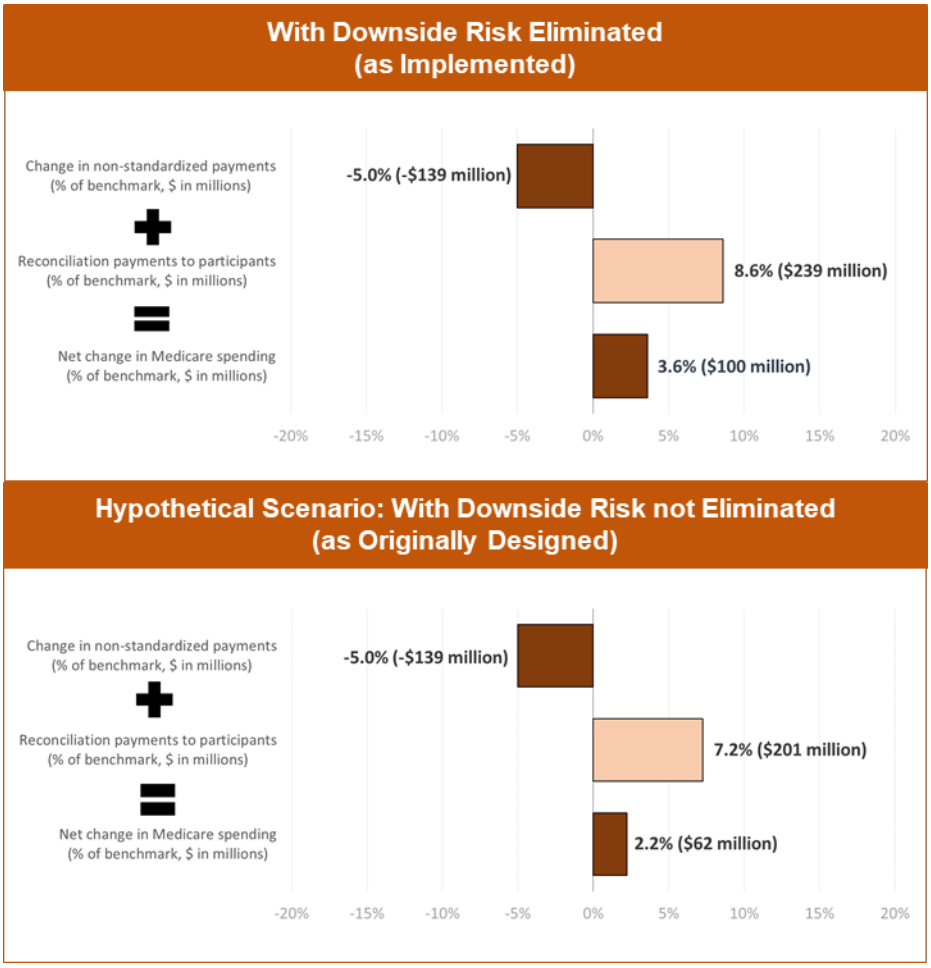
During the first four years of BPCI, Model 3 was associated with an estimated loss to Medicare of 3.6% of the benchmark ($100,033,848), ranging from a loss of 5.2% to 2.0% ($56,355,402 to $143,712,294, $p<0.05$). (See Exhibit 28.)[22] The net loss is equivalent to a per-episode loss to Medicare of $821. While the estimated decline in aggregate non-standardized payments was 5.0% ($138,664,780) across all 121,786 Model 3 episodes, reconciliation payments through Q3 2017 were 8.6% ($238,698,628), which far outweighed the estimated declines in Medicare payments.

If CMS had not eliminated downside risk, losses to Medicare would have been smaller. In this hypothetical scenario, reconciliation payments would have fallen from 8.6% to 7.2% of the benchmark ($200,849,996). Subtracting this amount from the reduction in aggregate non-standardized payments results in a net estimated loss to Medicare of 2.2% ($62,185,216),

[22] The range estimating the loss to Medicare is based on a 90% confidence interval.

ranging from 3.8% to 0.7% ($18,506,770 to $105,863,662, p<0.05). (See Exhibit 28.)[23] In this scenario, the net loss is equivalent to a per-episode loss to Medicare of $511.

Exhibit 28: Estimated Change in Medicare Spending with and without Downside Risk Eliminated, Model 3, Q4 2013 – Q3 2017



Note: The estimates of the change in aggregate non-standardized payments are from a difference-in-differences (DiD) model of standardized Medicare paid amounts during the qualifying inpatient stay and 90-day PDP. "With downside risk eliminated" depicts estimates of 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 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 Dark orange cells indicate estimates that are statistically significant at the 5% level. Light orange indicates estimates that are statistically significant at the 10% level.

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 2017) for BPCI and comparison providers. CMS data on reconciliation payments.

[23] The range estimating the loss to Medicare is based on a 90% confidence interval.

IV. Discussion and Conclusion

A. Discussion

This sixth annual BPCI evaluation report presents updated results using data for episodes initiated through the first four years of the model, which reflect an average of eight quarters of participation experience. For Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes, we estimated the impact of BPCI on 10 outcomes for the top 10 clinical episodes with the largest volume and for each participant type overall. With additional experience under the initiative, results remain consistent that 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. There are few indications in claims-based results that BPCI affected quality of care. Compared to the results for episodes initiated through December 2016 in the Year 5 Evaluation and Monitoring Annual Report, the declines in total payments and institutional PAC use in this report were generally larger across the top 10 clinical episodes with the largest volume initiated through September 2017. While this is a promising finding, it could be due in part to providers selectively withdrawing from the initiative or stopping participation in a clinical episode.

A new analysis for this report examined differences in outcomes for hospital and PGP episodes. For a common set of 21 clinical 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, results differed for surgical versus medical clinical episodes. For surgical clinical episodes, PGP episodes had a greater reduction in total payments than hospital episodes while for medical clinical episodes, hospital episodes had a greater reduction in total payments than PGP episodes.

These differences may stem from the different levers available to redesign care and financial motivations of the two participant types. 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.

There are other plausible explanations for the differences in results for hospital and PGP episodes. For instance, hospitals may be more risk averse and attempt to avoid hospital readmissions by sending patients to home health. Hospitals may also own or have close relationships with home health providers, which might provide hospitals with incentives to send their patients there. PGPs may also have been more likely to use other services to manage patients during the episode, which would not be captured through claims data or impact Medicare payments. Further analysis would be required to confirm what drives the observed differences in strategy.

In this report, we also estimated savings to the Medicare program for Model 2 and Model 3 for all 48 clinical episodes. Consistent with earlier estimates, even though BPCI resulted in reductions in episode payments, Medicare experienced 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 through September 2017 in this report were slightly smaller compared to the estimated net losses through December 2016 in the Year 5 Evaluation and Monitoring Annual Report. However, net losses under Model 3 were larger through September 2017. At the clinical episode level, major joint replacement of the lower extremity would have had a statistically significant estimate of net savings through September 2017 for episodes initiated under both Model 2 hospitals and Model 2 PGPs in the hypothetical scenario that downside risk was not temporarily eliminated. Under the same hypothetical scenario, Model 2 PGP episodes would also have had net savings for medical non-infectious orthopedic episodes through Q3 2017.[24]

As described above, during the reconciliation process, actual 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 2–3% to create a target price. Medicare savings, therefore, depended on benchmarks accurately reflecting what episode payments would have been absent BPCI. National trends, however, may not have been accurately capturing the payment changes for BPCI participants that would have happened in the absence of BPCI. This, in return, 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.[25] The evaluation, in contrast, uses a comparison group instead of trended historical payments to represent 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 incorporates risk adjustment to account for changes in patient mix from baseline to intervention period as well as any potential cost-shifting to services not covered by the episode while the CMS benchmark does not account for these factors.

Other features of the initiative also contributed to the lack of Medicare savings. CMS eliminated downside risk during periods of the initiative to accommodate start-up challenges experienced by the Innovation Center 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.

[24] In the Year 4 Evaluation and Monitoring Annual Report, the most recent report with clinical episode level estimates of savings, congestive heart failure also had statistically significant net savings to Medicare through September 2016 under the hypothetical scenario of no elimination of downside risk. While not statistically significant, the point estimates in this report, through September 2017, were similar. Model 2 PGP episodes were not analyzed in the Year 4 Evaluation and Monitoring Annual Report. The report is available for download at https://innovation.cms.gov/initiatives/Bundled-Payments/index.html.

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

The design of BPCI Advanced aims to incorporate lessons learned from the BPCI initiative. Entry and exit opportunities are scaled back under BPCI Advanced. BPCI Advanced also reduced the number of 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 projection. 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 other factors that will affect the payments for an episode of care, outside of the model, 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.

It remains to be seen whether BPCI Advanced will achieve net savings to Medicare and whether it will improve quality of care for Medicare beneficiaries. CMS has a separate contract in place for the evaluation of BPCI Advanced which will assess whether and the extent to which these goals are achieved.

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 9 of 80 results, or 11% ($p<0.10$). 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, 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.[26]

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 62% of Model 2 PGP participants, on average by clinical episode, in the analysis. Second, over 70% 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. This exclusion eliminated between 31% and 52% of potential hospitals from inclusion in the comparison group.

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, but as with all regression models, 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 Medicare program savings required several assumptions. First, we assumed the analysis sample was representative of all clinical episodes. We extrapolated the impact of BPCI on payments in the analysis sample of clinical episodes with sufficient volume for risk adjustment to

[26] 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 in the absence of BPCI, as the estimated average in the BPCI intervention sample converges to the long-term average rate.

all clinical episodes, including some for which we did not produce BPCI impact estimates.[27] 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

There is still more to learn from the BPCI initiative. There are four more quarters of claims data to evaluate. The final report will include impact estimates through the end of the initiative for Model 2 hospital-initiated episodes and Model 2 PGP-initiated episodes, as well as episodes initiated under Model 3 SNFs, Model 3 HHAs, and Model 4 hospitals. It will also include estimates of Medicare savings through the end of the BPCI initiative.

[27] We produced BPCI impact estimates for Model 2 hospital participants in 32 clinical episodes (98% of all Model 2 hospital episodes), for Model 2 PGP participants in 21 clinical episodes (93% of all Model 2 PGP episodes), for Model 3 SNF participants in 11 clinical episodes (74% of all Model 3 SNF episodes), and for Model 3 HHA participants in 3 clinical episodes (61% of all Model 3 HHA episodes).