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

Prepared for:

CMS

Prepared by:

The Lewin Group

With our partners:

Abt Associates, GDIT, Telligen, and Optum

June 2018

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

The Lewin Group

Authors:

Laura Dummit, Grecia Marrufo, Jaclyn Marshall, Tristan Ackerman, Sarah Bergman, Aylin Bradley, Rebecca Cherry, Inna Cintina, Daniel Cooper, Syvart Dennen, Philip Dishuck, Ayah Fannoun, Megan Hyland, Gina Gerding, Dan Gregory, Catherine Griffin, Ashley Johnson, Susan Joy, Jordan Kahn, Jon Kelly, Alex Lampert, Karla Lopez de Nava, Brandon Maughan, Court Melin, Sebastian Negrusa, Julie Somers, Dan Strubler, Katarina Swanson, Ellen Tan, Kyi-Sin Than, Peter Weidner, Ayanna Williams, Grace Yang, David Zhang, Dean Farley, Ian Breunig, Andrea Hassol, Qian Gu, Sean McClellan, Matt Trombley, Christine LaRocca, Becky Blystone, Colleen Kummet

Lewin's address:

3130 Fairview Park Dr, Suite 500, 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-500-2011-00001I Task Order HHSM-500-T0007.

Table of Contents

List of Appendices

Executive Summary

The Centers for Medicare & Medicaid Services (CMS) implemented the risk-bearing phase of Models 2, 3, and 4 of the Bundled Payments for Care Improvement (BPCI) initiative under the authority of the Center for Medicare & Medicaid Innovation (CMMI) in October 2013.[1] The BPCI initiative tests four Models for linking provider payments for a clinical episode of care to determine whether bundled payments can reduce Medicare payments while maintaining or improving quality of care. The voluntary initiative is designed to allow participants to choose among several key options, such as payment approach, type of clinical episode, and episode definitions. This design implicitly recognizes the variability across health care markets, providers, and episodes of care. The resulting diversity in responses and impacts will provide CMMI with information on the approaches that show the most promise in achieving payment reductions while maintaining or improving quality.

This annual report uses payment, utilization, and quality outcomes to describe the experience of BPCI Models 2 and 3 during the first three years of the initiative, from Q4 2013 through Q3 2016.[2] Because organizations were able to join and add clinical episodes over an extended period, the data in this report are based on an average of five quarters of participation. Our results are consistent with previous reports that indicate that BPCI participants are responding to the initiative's incentives by reducing Medicare payments. The next annual report will be a summative evaluation of BPCI that will incorporate all analyses conducted during the five year contract.

A. Structure of the Initiative

The BPCI initiative rewards participants financially for reducing Medicare payments for a clinical episode of care relative to a target price. BPCI Awardees, which can be health care providers or other entities that convene health care organizations, entered into agreements with CMS to be held accountable for total Medicare episode payments. Awardees' agreements with CMS specified their Model choice as well as choices among 48 clinical episodes, other episode characteristics, and multiple options for program rule waivers and financial arrangements with other parties. The clinical episodes are defined by the Medicare Severity Diagnosis Related Group (MS-DRG) of the anchor hospitalization. Providers and other organizations voluntarily participate in BPCI. They could enter into the risk-bearing phase of the initiative during a 2-year period, through September 2015, and enter additional clinical episodes into the risk-bearing phase through December 2015. Providers can stop participating in a given clinical episode on a quarterly basis. Awardees can terminate their participation in the initiative at any time.

Almost all services provided during the clinical episode are bundled for payment purposes. Hospice and certain services unrelated to the anchor hospitalization are excluded from the bundle, such as readmissions for certain MS-DRGs and some Part B services. The two BPCI Models evaluated in this report are:

- **Model 2** – This Model has the most comprehensive bundle, which includes the anchor inpatient hospital stay and all concurrent professional services and other Medicare Part A- and Part B-covered services (with certain exclusions) furnished within the chosen episode

[1] Model 1 began earlier than Models 2, 3, and 4 and was evaluated separately.
[2] Model 4 was not included because of limited participation in the Model.

length of 30, 60, or 90 days post discharge. Individual providers continue to be paid Medicare fee-for-service amounts, and aggregated episode payments are reconciled retrospectively against a target price, which CMS determined based on discounted, historical Medicare payments. When Awardees' episode payments are less than the target price, Awardees may receive the difference, termed the net payment reconciliation amount (NPRA), which they can keep or share with their partnering providers. When Awardees' episode payments are greater than the target price, they may have to pay amounts to CMS. The episode initiator (EI), that is, the provider associated with the start of the episode, can be a hospital or a physician group practice (PGP).

- **Model 3** – The episode starts when a beneficiary is admitted to a participating skilled nursing facility (SNF), home health agency (HHA), inpatient rehabilitation facility (IRF), or long-term care hospital (LTCH) within 30 days of a hospital discharge for an MS-DRG in the participant's chosen clinical episode. Alternatively, the episode starts when a beneficiary is admitted to a SNF, HHA, IRF, or LTCH within 30 days of a hospitalization for which the attending or operating physician was a member of a Model 3 PGP that was participating in the clinical episode that contained the MS-DRG of the beneficiary's hospitalization. The bundle includes all services (with certain exclusions) within the chosen episode length of 30, 60, or 90 days. Individual providers continue to be paid Medicare fee-for-service amounts, and aggregated episode payments are reconciled retrospectively against the target price. When Awardees' episode payments are less than the target price, Awardees may receive NPRA, which they can keep or share with their partnering providers. When Awardees' episode payments are greater than the target price, they may have to pay amounts to CMS. The EI can be a SNF, HHA, IRF, LTCH, or PGP.

B. Evaluation Design

The BPCI evaluation is based on a mixed methods approach that incorporates multiple data sources, including Medicare claims, patient assessments, beneficiary surveys, site visits, and participant interviews. This annual report, however, includes only claim-, patient assessment- and survey-based outcomes. The claim analyses use a difference-in-differences (DiD) design to estimate the differential change in payment, quality, and utilization 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 claims and patient assessment data were also used to address the question of whether the mix of patients of BPCI participants changed during the intervention. Survey analyses use a cross-sectional design to estimate the differences in patient-reported outcomes between respondents who received services from BPCI providers and respondents who received services from a comparison group of non-BPCI providers. Outcomes are risk adjusted to account for differences in patient mix.

Because the BPCI initiative includes multiple start dates for participants and various combinations of Models, EIs, and clinical episodes, the results are first differentiated by Model and EI provider type. Results are then stratified by clinical episode because of their different underlying cost and utilization patterns. The sample sizes are not sufficient to examine every Model, EI, and clinical episode combination. Additionally, this report does not include analyses of episodes initiated by PGP EIs, which are expected in the next annual report.

C. Results

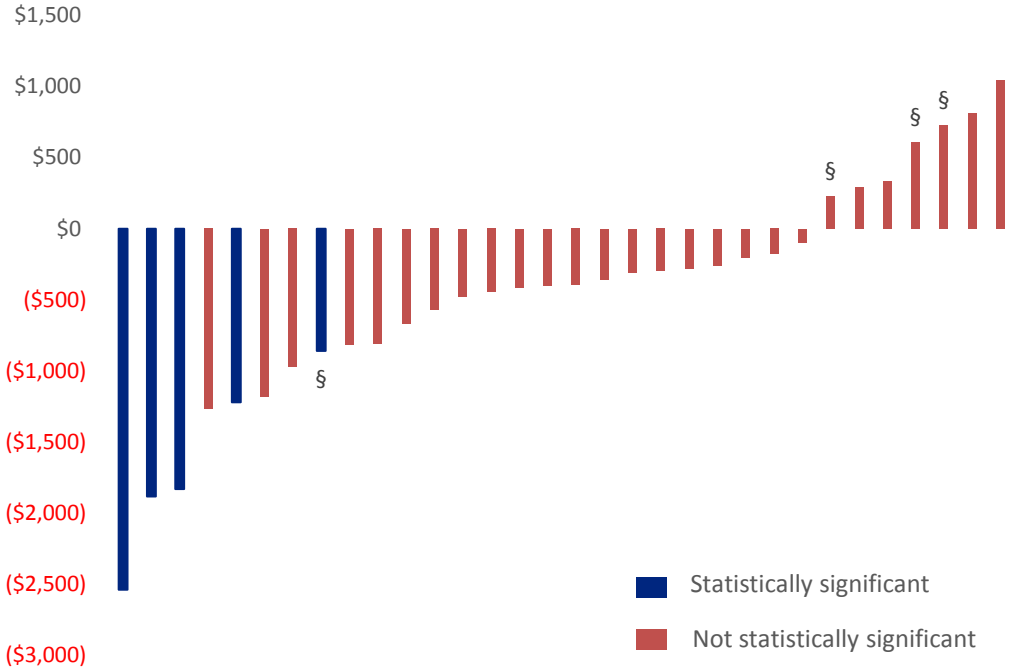
For Model 2, we had sufficient sample size to evaluate 32 hospital-initiated clinical episodes in the claims analyses. The number of hospital EIs in the evaluation ranged from 26 to 303 across these clinical episodes, and they initiated between 1,089 and 97,922 episodes over the first 12 quarters of the initiative. Because providers were allowed to join BPCI over an extended period and stop participating on a quarterly basis, these data represent an average of five quarters of participation. The beneficiary survey covered 21 hospital-initiated Model 2 clinical episodes from May 2015 through November 2016. A total of 20,319 BPCI patients responded to the survey. These responses were weighted to represent all 73,000 episodes that occurred during the period covered by the sample and were pooled across all clinical episodes.

For Model 3, we had sufficient sample size to analyze 11 SNF-initiated clinical episodes. The number of SNF EIs in the evaluation ranged from 78 to 236 across these clinical episodes, and they initiated between 676 and 5,711 episodes over the first twelve quarters of the initiative. These data represent an average of five quarters of participation. We were able to evaluate the impact of BPCI on three HHA-initiated clinical episodes. These episode data also represent five quarters of participation under BPCI. We do not report Model 3 beneficiary survey results because there was insufficient volume to create a representative sample of clinical episodes.

- **The majority of clinical episodes under Model 2 had relative declines in total Medicare payments.**

  The total standardized allowed payment amount for the inpatient stay plus 90-day post-discharge period declined in 25 of the 32 clinical episodes (Exhibit ES-1). This relative decline was statistically significant ($p<0.05$) for five clinical episodes: transient ischemia, major joint replacement of the lower extremity (MJRLE), medical non-infectious orthopedic, hip and femur procedures except major joint, and urinary tract infection.

Exhibit ES-1: Impact of BPCI on Total Standardized Allowed Payment Amount for the Inpatient Stay Plus 90-day Post-discharge Period, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013 - Q3 2016



Note: The estimates in this exhibit are the result of a difference-in-differences (DiD) model. These amounts combine the Medicare program payments with the patient coinsurance and copayment amounts and then adjust for Medicare payment policies to ensure that any differences across time and providers reflect real differences in resource use rather than Medicare payment policies (e.g., teaching payments or differential payment updates). ACH = acute care hospital.

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

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

- **Lower use of institutional post-acute care led to reduced Medicare payments under Model 2.**

  The standardized allowed amount for SNF care went down for 28 of the 32 clinical episodes, and the relative decline was statistically significant for 13 ($p<0.10$). The standardized allowed amounts for IRF generally went down as well. Payments for HHA services increased for all but two clinical episodes, and the relative change was statistically significant for 12. The increased HHA payments were not high enough to offset the reduced payments for SNF and IRF.

  The overall pattern of reduced institutional post-acute care (PAC) payments under BPCI is substantiated by the utilization data. The proportion of patients discharged to a PAC provider did not change much for most clinical episodes, although there were three clinical episodes with a statistically significant decline in the proportion of patients discharged to PAC and two with a statistically significant increase ($p<0.10$). For 28 of the clinical episodes, among the patients who received PAC, a lower proportion were discharged to institutional PAC (SNF, IRF, or LTCH). This was a statistically significant reduction for eight clinical episodes. There was a statistically significant

increase for one clinical episode in the proportion of patients who were discharged to institutional PAC, among all patients who received PAC services.

Among patients receiving SNF care, the average number of SNF days in the 90 days following the inpatient qualifying hospitalization declined for 27 clinical episodes. This decline was statistically significant for 18. Among patients receiving any home health care, the average number of visits went up for the vast majority of clinical episodes, although the magnitudes tended to be small.

- **There were few indications in the claim-based measures that BPCI affected quality of care under Model 2.**

  The three claim-based measures of quality of care, emergency department visits, mortality, and unplanned readmissions, did not signal any widespread effect of BPCI on quality. Across these measures for the 32 clinical episodes, there were few statistically significant differences in the change between BPCI and the comparison group and the point estimates of the differential change tended to be small and in both directions. For all-cause mortality, there was a statistically significant relative decline for three clinical episodes. However because mortality is a relatively rare event, these changes may reflect typical fluctuation, so corroborating evidence is needed before attributing the decline to BPCI. There was a statistically significant decrease in emergency department use for one clinical episode. For one clinical episode, there was a statistically significant increase in unplanned readmissions, although BPCI and comparison provider readmission rates for this clinical episode were not on parallel trends during the baseline period, so this result may be biased.

- **Beneficiary surveys indicated that BPCI did not affect self-reported changes in functional status and had small negative effects on care experiences and satisfaction.**

  Differences between BPCI and comparison respondents were small and not statistically significant for six out of seven measures of change in self-reported functional status across all clinical episodes. Relative to the comparison group, a smaller proportion of BPCI respondents reported favorable care experiences for six of nine measures ($p<.10$). BPCI respondents were also less likely than comparison respondents to report the highest levels of satisfaction with their overall recovery since leaving the hospital ($p<.10$).

- **Across most clinical episodes, the average resource intensity of patients did not change relative to the comparison group under Model 2.**

  We examined patient characteristics that are associated with resource use because changes in the mix of patients in a clinical episode from historical levels could affect participants' ability to reduce episode payments. If the patient mix was less resource intensive during the intervention period, it could result in unwarranted NPRA payments to BPCI participants. There were few indications, however, that there were systematic changes in patient mix. Two exceptions were non-fracture MJRLE and spinal fusion clinical episodes. For both of these elective surgeries, there were indications that the patients of BPCI participants were less resource intensive during the intervention period than in the baseline, relative to the comparison group. This is particularly notable

because participants have the ability to identify these patients prior to admission for these elective surgeries and, therefore, could select less resource intensive patients to improve their ability to achieve positive NPRA. Qualitative data suggest another reason for the decline in patient resource intensity for these elective surgeries. Some participants said they postponed surgeries for higher risk patients until certain risk factors could be addressed, thus improving patient outcomes.

- **EIs that stopped participating in a given clinical episode under Model 2 contributed a large share of episodes across many clinical episodes.**

  Providers are able to stop participating in a clinical episode on a quarterly basis, although the episodes they initiated during their time in the initiative are included in the BPCI impact estimates. During the first three years of the initiative (Q4 2013 through Q3 2016), 58 hospital EIs stopped participating in the MJRLE clinical episode and these EIs accounted for 14% of the MJRLE episode volume. A smaller number of hospital EIs stopped participating in other clinical episodes, but because those that stopped were a larger share of EIs that ever participated in that clinical episode, the impact of the EIs leaving the clinical episode on the outcomes presented in this report may be considerably larger. For four clinical episodes, hospital EIs that stopped participating comprised 50% or more of the episodes during the intervention period. An additional seven clinical episodes had hospital EIs that stopped participating that contributed over 40% of the episodes.

- **Episode payments declined for most of the clinical episodes under Model 3 participating SNF EIs.**

  In 9 of the 11 clinical episodes, we observed statistically significant declines in the SNF standardized allowed amount over the 90-day post-discharge period. All nine clinical episodes had accompanying increases in HHA amounts in the period, but not enough to offset the reduced SNF payments, and the differential increase in HHA payments was statistically significant in five of these instances ($p<0.10$). For eight of the clinical episodes, increased hospital readmission payments offset some of the reduced SNF payments, although the differential increase in readmission payments was statistically significant in only three of these instances ($p<0.10$).

  Mirroring the payment data, there was a statistically significant relative decline in SNF days for the nine clinical episodes with reduced SNF payments. Home health visits went up in 10 clinical episodes, but the increase was statistically significant for only two.

- **There were a few indications of a relative decline in quality of care under Model 3 for SNF-initiated episodes, although further analysis suggests these estimates are not robust.**

  There were a few indicators in the claim-based measures that BPCI reduced quality of care under Model 3. Most changes in quality measures were not statistically significant. For two clinical episodes, however, the measures raised some concern. Chronic obstructive pulmonary disease (COPD) episodes had a statistically significant increase in mortality. This measure can be volatile over time, but the absolute level of the increase was relatively high. Although not statistically significant, there were

substantial relative increases in unplanned readmission rates and emergency department use for COPD episodes as well. For stroke episodes, both the increases in unplanned readmissions and emergency department use were statistically significant; although mortality declined, the reduction in mortality was not statistically significant. There were large differences in baseline values for many of the claim-based quality measures under Model 3, despite our attempts to find a good comparison group, which likely contributed to these findings. Sensitivity analyses with repeated samples of comparison episodes also indicated that the statistical significance of some results may have been due to the chance selection of particular comparison episodes, although the direction of the impact appears robust.

- **BPCI-participating SNFs may have treated less resource intensive patients in four clinical episode strata under Model 3.**

  Relative to Model 2, there may be more opportunities for Model 3 participants to change their mix of patients because PAC providers can evaluate patients while they are in the hospital to determine whether to admit them. For SNF EIs, four of the 12 Model 3 clinical episode strata had indications that BPCI patients were less resource intensive during the intervention period than the baseline, relative to the comparison group. Patients in MJRLE fracture, congestive heart failure, medical non-infectious orthopedic, and stroke clinical episodes were statistically significantly less resource-intensive across several measures. COPD episodes appeared to have more resource intensive patients relative to the comparison group, and the remaining clinical episode strata had no consistent change in resource intensity. In addition, across all clinical episodes, the four assessment-based functional status measures (moving in bed, transferring, walking in room, and toileting) indicated that BPCI-participating SNFs treated patients who required less assistance after joining BPCI relative to the change for the comparison group. A less resource intensive patient mix could result in unwarranted NPRA payments to BPCI participants, because the target price was based on the historical patient mix.

- **Over one-third of Model 3 SNF episodes were initiated by SNF EIs that ultimately stopped participating in the clinical episode.**

  Over the first 12 quarters of the initiative, 36% of episodes across the clinical episodes we could evaluate were initiated by SNF EIs that ultimately stopped participating in that clinical episode. For three clinical episodes, over 60% of the episodes were initiated in SNFs that withdrew from the model, and an additional three had 50% or more of their episodes start in SNFs that stopped participating in that clinical episode.

- **There were few indications of any statistically significant impacts of BPCI on HHA-initiated episodes under Model 3.**

  Over half of HHAs stopped participating in the episode for two of the clinical episodes analyzed, accounting for 21% and 58% of the episodes. Total episode payments declined in two out of the three clinical episodes, but the relative payment reductions were not statistically significant. They appeared to be due to reduced HHA, SNF, and readmission payments. Although emergency department use increased for HHA-initiated congestive heart failure (CHF) episodes, they declined for simple pneumonia

episodes. None of the other changes in quality measures were statistically significant or large absolute amounts.

- **Across the 46 Model 2 and Model 3 clinical episodes examined in this report, after considering the net payment reconciliation amount (NPRA) paid to participants, it is unlikely that that the Medicare program achieved savings under BPCI except for MJRLE and CHF clinical episodes under Model 2 ACH.**

  Although there were statistically significant declines in Medicare allowed payments for five Model 2 ACH clinical episodes and six Model 3 clinical episodes, the Medicare program did not achieve savings for 9 of these clinical episodes after accounting for the NPRA paid by Medicare to BPCI participants. MJRLE and CHF clinical episodes under Model 2 ACH were the only exceptions; the estimated decline in payments was statistically significantly larger than the NPRA paid to participants. However, it should be noted that the NPRA used in this analysis does not account for the fact that participants were not required to repay NPRA to Medicare for a portion of the initiative, which results in an overestimate of Medicare program savings.

D. Discussion and Conclusion

This fourth annual BPCI evaluation report presents results based on an average of five quarters of experience for both Model 2 and Model 3 participants. We estimated the impact of BPCI on select outcomes for 32 clinical episodes under Model 2 and 14 clinical episodes under Model 3. Our results are consistent with previous reports that indicated that BPCI participants responded to the initiative's incentives by reducing Medicare payments. We continue to see general patterns of reduced intensity of PAC, with shifts from institutional care to home health care. After considering the NPRA paid to participants that reduced their episode payments below their target amount, however, the Medicare program likely achieved savings on only two clinical episodes. In addition, there are few indications in claims data that BPCI affected quality of care, either positively or negatively. The beneficiary survey found no consistent adverse impacts of BPCI on changes in self-reported functional status. Model 2 BPCI patients, however, reported worse care experiences than comparison patients, although the differences were small. The next annual report will be more comprehensive by considering data across multiple sources.

This report documents the significant number of episodes that were initiated by participants that ultimately stopped participating in that clinical episode. This could have biased our results in two ways. Participants that were having difficulty reducing their episode payments below the target amount may have been likelier to exit. At the same time, the changing composition of the participants may have led to an unbalanced comparison group. However, it is unclear whether the changing composition of participants affected the estimated impact on the reported outcomes or the direction or magnitude of any possible effect. To estimate the effect of any bias, the next annual report will include estimates based on intent-to-treat methods. It will include all episodes of BPCI participants – including episodes from participants after they stopped participating in the clinical episode. We will also conduct analyses that will help in understanding the participant characteristics that are correlated with ending participation in a clinical episode.

There are additional limitations with this report. We do not include results from all data sources used in our evaluation. This report also does not include PGP-initiated episodes. Further, our

primary analytic approach is dependent on how well the comparison group represents what would have happened absent the BPCI initiative. For Model 3 participants, the comparison group and BPCI participants differed on key baseline characteristics, particularly quality measures, which may have contributed to the concerning quality outcomes for two clinical episodes. Sensitivity analyses with repeated samples of comparison episodes also indicated that the statistical significance of some results may have been due to the chance selection of particular comparison episodes. Further, this report does not measure the variation in impact of BPCI across providers. In addition, because we are measuring multiple outcomes across the range of Model, participant, and clinical episode combinations, by chance alone some results will appear statistically significant even though they are not true effects of the initiative. The estimates of Medicare program savings may be biased because they are based on standardized allowed payments, not actual Medicare program payments. The NPRA we used in this analysis does not account for the fact that participants were not required to repay NPRA for a portion of the initiative, which results in an overestimate of Medicare program savings. Finally, Medicare savings estimates do not incorporate the possible effect of BPCI on episode volume. If BPCI causes an unwarranted increase in episode volume, net savings to Medicare will be less. Nevertheless, we do not anticipate these limitations to substantively change this report's conclusions about net Medicare savings.

The next summative evaluation report will incorporate results and conclusions based on the many analyses conducted over the five-year contract. We will examine the impact of additional time under the initiative on payment, quality, and utilization outcomes. To date, we have estimated the impact of BPCI on the average episode. In the next report, we will also estimate the impact of BPCI on quality of care and beneficiary satisfaction among vulnerable beneficiaries. One of the most important advances during the next year will be analyzing the impact of BPCI on the BPCI-participating PGPs.[3] We will also expand our examination of the factors that contribute to whether a participant can reduce episode payments below its target price under BPCI to additional Model 2 hospital and PGP clinical episodes. We will refine our methodology to estimate the impact of BPCI on net savings to Medicare. Finally, for MJRLE, which is an elective surgery and is the highest volume episode in BPCI, we will assess whether BPCI caused an increase in episode volume, which may affect our conclusions about net Medicare savings.

[3] The lists of BPCI-participating physicians by PGP from Q1 2016 onward were corrected in Q1 2017. The evaluation team implemented and tested the revised methodology in July 2017.

I. Introduction

The Bundled Payments for Care Improvement (BPCI) initiative is designed to test whether linking payments for all providers involved in furnishing 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) implemented the risk-bearing phase of Models 2, 3, and 4 of the BPCI initiative in October of 2013 under the authority of the Center for Medicare and Medicaid Innovation (CMMI).[4] BPCI Awardees, which may be hospitals, physician groups, post-acute care (PAC) providers, or other entities that convene health care organizations, entered into agreements with CMS to be held accountable for total Medicare episode payments. Those agreements also specified Awardees' choices among three payment Models, which differ in the services included in the episode bundle and in payment method; 48 clinical episodes; three episode lengths; and three risk tracks. Awardees also submitted BPCI implementation protocols that specified whether they would use available program rule waivers, beneficiary engagement incentives, or financial arrangements that could be protected under specific waivers of fraud and abuse laws.

This report describes the evaluation of BPCI Models 2 and 3; Model 4 was not included in this annual report due to small sample sizes (see Year 3 annual report for the most recent Model 4 results). Model 1 was evaluated separately. Awardees in Models 2 or 3 are rewarded for reducing Medicare payments for the bundle of services in the episode relative to a target price. The target price is determined by CMS and is generally based on historical payments attributed to the episode initiating provider for the same type of clinical episode. When aggregate Medicare episode payments are less than the target price, Awardees may receive net payment reconciliation amounts (NPRA) that reflect this difference, which they can keep or share with their partnering providers. When aggregate episode payments are higher than the target price, Awardees may have to pay amounts to CMS. Thus, to obtain positive NPRA, Awardees have incentives to reduce aggregate episode payments.

The Lewin Group, with our partners Abt Associates, Inc., GDIT, Telligen, and Optum, is under contract to CMS to evaluate and monitor the impact of BPCI Models 2, 3, and 4. This is the fourth of five annual reports under this contract that synthesizes the findings from various evaluation and monitoring activities.

The objective of this annual report is to provide a timely update of the impact of BPCI Models 2 and 3 on payment, utilization, and quality outcomes from the first three years of BPCI experience, from Q4 2013 through Q3 2016.[5] The next annual report, will be a summative evaluation of BPCI that will incorporate all analyses conducted during the five year contract.

A. BPCI Initiative

The BPCI initiative incorporates multiple approaches to aligning incentives for providers involved in an episode of care. Under each BPCI Model, an episode of care is triggered by a hospitalization

[4] Model 1 began earlier than Models 2, 3, and 4.
[5] **Appendix A** includes an acronym list and glossary for common terms used in this report.

for a Medicare Severity Diagnosis Related Group (MS-DRG) contained in one of 48 clinical episodes (see **Appendix B** for a list of the 48 clinical episodes and associated MS-DRGs).

The services provided during the clinical episode are bundled for payment purposes. Hospice and specific services unrelated to the triggering hospitalization are excluded from the bundle, such as readmissions for certain MS-DRGs and some Part B services. The bundle and payment approach vary by Model as follows:

- **Model 2** has the most comprehensive bundle, which includes the triggering hospital stay (i.e., the anchor hospitalization) and all professional items and services (with certain exclusions) furnished within the chosen episode length of 30, 60, or 90 days post-discharge. The episode starts when a beneficiary is admitted to an episode initiating acute care hospital (ACH or hospital) or when the attending or operating physician for the beneficiary's hospitalization is in an episode initiating physician group practice (PGP). Individual providers are paid regular Medicare fee-for-service amounts throughout the episode and aggregate episode payments are reconciled retrospectively against the target price.
- The **Model 3** bundle includes items and services furnished after the anchor hospital discharge, within the chosen episode length of 30, 60, or 90 days. The episode starts when a beneficiary is admitted to an episode initiating skilled nursing facility (SNF), home health agency (HHA), inpatient rehabilitation facility (IRF), or long-term care hospital (LTCH) within 30 days of discharge from a hospitalization for a chosen clinical episode. In the case of PGP episode initiators (EIs), the episode starts when a beneficiary is admitted to a post-acute care (PAC) setting within 30 days of discharge from a hospitalization where the attending or operating physician for the beneficiary's hospitalization is in a participating PGP. Individual providers are paid regular Medicare fee-for-service amounts throughout the episode and aggregate episode payments are reconciled retrospectively against the target price.

There are 336 possible unique combinations of Model, clinical episode, and EI provider type in BPCI across Models 2 and 3.[6] During the first three years of the initiative, episodes were initiated in 258 of the possible combinations. (See **Appendix C** for the count of episodes and episode initiators by Model and clinical episode during the first three years of the initiative.) Of these combinations, only 46 had sufficient participation and volume to support a regression-based difference-in-differences (DID) analysis using a matched comparison group.[7,8]

[6] In addition, Awardees may select one of three options for bundle length and risk track. Risk track refers to the level of winsorization, that is, the outliers that are excluded from the reconciliation payment calculation (Risk track A includes episodes whose payments fall between the 1st and 99th percentile of national payments for that MS-DRG; risk track B includes the 5th to 95th percentile; and risk track C includes the 5th to 75th percentile).

[7] We did not conduct any impact analysis on Model 2 or 3 PGP EIs because the methodology for identifying BPCI PGP episodes was being finalized during the time the analyses were being conducted for this report.

[8] The results were stratified for major joint replacement for lower extremity (MJRLE) clinical episodes into fractures and non-fractures for both Model 2 hospital episode initiators and Model 3 SNF episode initiators. The results for coronary artery bypass graft (CABG) clinical episodes were stratified into emergent and non-emergent for Model 2 hospital episode initiators. The results for MJRLE fractures, MJRLE non-fractures, CABG emergent, and CABG non-emergent are included in the appendices.

CMS implemented BPCI in a phased approach. Participants could apply for Phase 1, the preparation phase, and then transition into Phase 2, the risk bearing phase, over an extended period. The first participants began Phase 2 for at least some of their clinical episodes by October 1, 2013. By October 1, 2015, all participants had to transition their clinical episodes to Phase 2.

Please refer to the Year 3 annual report for additional detail on the BPCI initiative.[9]

B. Research Questions

The fourth annual report updates the impact of BPCI Models 2 and 3 on payment, utilization, and quality outcomes. We include the following domains:

- **Payment and utilization –** Payments are based on Medicare standardized allowed amounts, which we risk adjust to account for differences in patient and provider characteristics.[10] Utilization measures include inpatient lengths of stay for Model 2 and PAC use, such as number of home health visits and the number of SNF days for Models 2 and 3.
- **Quality of care –** Claim-based quality measures are mortality, emergency department use, and unplanned readmissions. Survey-based quality measures include self-reported changes in functional status, care experiences, and overall satisfaction with recovery. All quality measures were risk adjusted to account for patient and provider differences.
- **Patient mix** – Demographic characteristics, prior health conditions, and prior health care utilization are used to assess patient mix. For Model 3 episodes, we also examined patient assessment data.

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

[10] Medicare allowed amounts are the Medicare paid plus beneficiary cost-sharing amounts, which are standardized to remove Medicare policy adjustments to ensure that any payment differences across time and providers reflect real differences in resource use rather than Medicare payment policies (e.g., teaching payments or differential payment updates).

II. Methods

A. Data Sources

We used Medicare administrative data to identify EIs and comparison providers as well as describe episodes of care (see Exhibit 1). We used provider-level data sources to identify and describe BPCI-participating EIs and select comparison providers. Medicare claims and enrollment data were used to construct episodes of care for patients at BPCI EIs (BPCI population) and matched comparison providers. We also used claims and survey data to create outcome measures and risk adjustment variables.

Exhibit 1: Data Sources used in BPCI Evaluation Year 4 Report

	Dataset Name	Date Range	Dataset Contents	Use
Provider-level data sources	CMS's BPCI database - BPCI Participant and Episode Reports	2013-2016	Information compiled by CMS on BPCI participants and future participants and their clinical episodes, including participant name, CMS Certification Number, location, type (ACH, SNF, etc.), BPCI "role," Model, clinical episode(s) and length(s), BPCI participation start and end dates, and contact information.	Used to identify Quarter 4 2013 through Quarter 3 2016 BPCI EIs and clinical episodes. Identified Model 1 participants to exclude from comparison group.
	Medicare Provider Enrollment, Chain, and Ownership System (PECOS)	2011-2014	Information on Medicare providers, including ownership and chain relationships among providers.	Used to identify ownership of BPCI EIs and potential comparison providers and to create an indicator of whether the provider was part of a chain. Both of these characteristics were used in the creation of the comparison groups.
	Provider of Services (POS) file	2011-2015	Information on Medicare-approved institutional providers, including provider number, size, and staffing.	Used in descriptive analysis of BPCI and non-BPCI participants. Used as predictors in provider propensity model on participation in BPCI or characteristics for Mahalonobis matching.
	Area Health Resource File (AHRF)	2011	County-level data on population, environment, geography, health care facilities, and health care professionals.	Used as predictors in provider propensity model on participation in BPCI or characteristics for Mahalonobis matching.
	Master Data Management (MDM)	2013-2016	Provider- and beneficiary-level information on participation in CMMI payment demonstration programs.	Used to identify providers that are involved in an Accountable Care Organization (ACO) or other Medicare Shared Savings programs.
	Episode files from Reconciliation contractor	2013-2014	Final episode SAS research datasets used by the Reconciliation contractor.	Used to validate our implementation of the BPCI episode construction methodology.

	Dataset Name	Date Range	Dataset Contents	Use
Transaction-level data sources	Medicare fee-for-service (FFS) claims	Jan 2010-Dec 2016	Medicare Part A and B claims.	Used to create episodes of care and outcome measures such as readmissions, emergency department (ED) visits, number of days in each care setting (e.g., skilled nursing facility). Also used to create risk factors including hierarchical condition categories (HCCs) and health care utilization prior to anchor or qualifying hospitalization.
	Medicare standardized payments	Jan 2011-Dec 2016	Medicare standardized payments for 100% of Part A and B claims received via the Integrated Data Repository (IDR) from another CMS contractor.	Used to create Medicare standardized payment amounts (Part A and B) and allowed standardized payment amounts (including beneficiary out-of-pocket amounts).
	The Master Beneficiary Summary File (MBSF)	Jan 2010-Dec 2016	Beneficiary and enrollment information, including beneficiary unique identifier, address, date of birth/death, sex, race, age, and Medicare enrollment status.	Used to identify eligibility for episodes of care, beneficiary demographic characteristics, and beneficiary eligibility for inclusion in the denominator for each of the outcome measures.
	Minimum Data Set (MDS) patient assessments	2011-2016	Comprehensive post-acute patient assessments completed by clinicians. Required for residents of Medicare-certified skilled nursing facilities (SNFs). Administered at admission, at discharge, days 14, 30, 60, 90, and quarterly thereafter.	Provided conditions and functional status upon admission to SNF in Model 3.
	Outcome and Assessment Information Set (OASIS) patient assessments	2011-2016	Comprehensive post-acute patient assessments completed by clinicians. Required for Medicare-paid home health agency (HHA) patients. Completed at the start of care and at discharge, and when care resumes following a hospitalization.	Provided conditions and functional status upon admission to HHA in Model 3.
	Beneficiary survey	2015-2016	Surveys completed by Medicare beneficiaries or their proxies. Received approximately 90 days after hospital discharge.	Used to create outcomes measures such as self-reported change in functional status, care experience, and overall satisfaction with recovery.

B. Claim-based Analyses

1. Study Population

The BPCI study population includes all episodes initiated by BPCI EIs that had Phase 2 episodes between Q4 2013 and Q3 2016. If an EI terminated during this period, we include the episodes that it initiated up until its withdrawal date.

The quantitative analysis uses a difference-in-differences (DiD) design to estimate the differential change in cost, quality, and utilization outcomes between the baseline and an intervention period for beneficiaries who received services from BPCI EIs relative to beneficiaries who received services from a comparison group of non-BPCI providers. This comparison group needs to be similar to BPCI EIs with respect to baseline characteristics that could affect their decision to participate and could be related to their performance under BPCI. Such characteristics include market-level and provider-specific attributes. Because providers voluntarily enroll in BPCI, EIs were likely to be different than non-participants in ways that could bias our results. For example, BPCI EIs may have had less efficient care in the pre-intervention period and consequently had more room for improvement relative to non-participants.

We constructed comparison groups for 46 Model, provider type, and clinical episodes from the universe of Medicare providers that had not entered Phase 2 of BPCI. For this report, we examined clinical episodes initiated by Model 2 ACH, Model 3 SNF, and Model 3 HHA BPCI EIs.[11] Each unique Model-EI type-clinical episode group was considered to have a sufficient sample size for meaningful analysis if there were 20 EIs with 1,000 clinically relevant episodes.[12] However, a few groups with somewhat lower sample sizes were included if they had unique policy importance.

Comparison providers and episodes were selected in four steps. First, providers were identified as potential comparison providers if they: (i) shared certain key characteristics with BPCI EIs, (ii) were eligible to participate in the BPCI initiative, (iii) were not located in markets where BPCI EIs had over half of the discharges associated with any of the 48 BPCI clinical episodes, (iv) were not participating in BPCI, and (v) were not affiliated with BPCI participants. Second, each BPCI EI was matched with up to 15 comparison providers using a statistical matching technique to minimize the differences in the distributions of characteristics between BPCI and comparison providers. Third, episodes were constructed for beneficiaries treated by matched comparison providers based on the BPCI program rules. Finally, a sample of episodes was drawn from among those identified in the previous step to match the distribution of BPCI episodes by MS-DRG and date of service. See **Appendix D** for more details regarding each step of the comparison group and episode selection process. **Appendix E** shows the calipers chosen for each propensity score matching model as well as the standardized differences of each covariate included in the matching models between BPCI EIs and matched comparison providers for each strata.

[11] We did not conduct any impact analysis on Model 2 or 3 PGP EIs because the methodology for identifying BPCI PGP episodes was being finalized during the time the analyses were being conducted for this report. The Year 5 BPCI evaluation report will include the impact of BPCI on episodes initiated by M2 PGPs. We did not analyze Model 4 in this report due to small sample sizes; please see the Year 3 BPCI evaluation annual report for the most recent Model 4 results.

[12] Groups were considered meaningful for the analysis if there was enough participation in BPCI, but no formal power calculation was conducted to assess minimum sample size.

2. Measurement Periods

We defined two sets of *measurement periods* for which we calculated the outcomes of interest: the *bundle timeline* and the *patient timeline*. The bundle timeline measurement periods vary by Model and episode length because they are defined relative to the BPCI bundle period (i.e., pre-bundle, post-bundle). In contrast, the patient timeline measurement periods are consistent across Models and episode lengths because they depend on the patient's transition through the episode of care (e.g., post-hospital discharge), so they allow us to compare outcomes across Models and episode lengths.

Every outcome was calculated for one or more defined measurement periods. For example, Model 2 unplanned readmission rates were calculated for two patient timeline measurement periods, within 30 days and within 90 days of hospital discharge. These measurement periods are labeled 30 day post-discharge and 90 day post-discharge. Exhibits 2 and 3 describe the bundle and the patient timeline measurement periods for Models 2 and 3.

Exhibit 2: Definition of Measurement Periods Relative to the Bundle Timeline across Models

Model	Pre-bundle Period	Bundle Start	Bundle End	Post-bundle Period
2	Anchor IP stay admission date minus 30 days	Anchor IP stay admission date	Anchor IP stay discharge date plus bundle length (30, 60, or 90 days)	30 and 90 days after the end of the bundle
3	EI PAC admission date minus 30 days	EI PAC admission date	EI PAC admission date plus bundle length (30, 60, or 90 days)	30 and 90 days after the end of the bundle

Notes: IP = inpatient, EI=episode initiator, PAC=post-acute care

Exhibit 3: Definition of Measurement Periods Relative to the Patient Timeline across Models and Episode Lengths

Model	Inpatient Hospitalization	Post-discharge Period
2	Anchor/qualifying IP stay from IP admission date to IP discharge date	30, 90, and 120 days after anchor/qualifying IP discharge date
3		

3. Outcome Definitions

Exhibit 4 summarizes the key outcome measures by domain. **Appendix F** provides detailed definitions of each outcome measure.

Exhibit 4: Claim-based Measures used to Evaluate the Impact of BPCI, by Domain

Domain	Outcome
Payment[a]	Total Medicare standardized allowed payment for inpatient stay plus 90 and 120 days post discharge
	Total Medicare standardized allowed payment included in the bundle definition
	Total Medicare standardized allowed payment not included in the bundle definition
	Medicare standardized allowed payment, 30 day pre-bundle period
	Medicare standardized allowed payment, 30 and 90 day post-bundle period
	Total Medicare Part A standardized allowed payment (by various settings)
	Total Medicare Part B standardized allowed payment (by various service categories)
Utilization	Acute inpatient length of stay
	Number of days in institutional PAC setting (total and for SNF)
	Number of home health visits
	First PAC setting following inpatient discharge
	Patients discharged to an institution relative to discharged home with home health
	Patients discharged to any PAC
Quality	Unplanned readmission rate
	Emergency department use without hospitalization
	All-cause mortality rate

[a] These amounts combine the Medicare payments with the patient coinsurance and copayment amounts and then adjust for Medicare payment policies to ensure that any differences across time and providers reflect real differences in resource use rather than Medicare payment policies (e.g., teaching payments or differential payment updates).
PAC=post-acute care, SNF=skilled nursing facility

4. Difference-in-differences Methodology

The DiD approach quantifies the impact of BPCI by comparing changes in claim- and assessment-based outcomes for the BPCI episodes with changes in outcomes for the comparison episodes, between the baseline and intervention periods. This approach eliminates biases from time invariant differences between the BPCI and comparison episodes and controls for trends in the BPCI population.[13] The risk adjustment regression model incorporates data from two periods prior to BPCI implementation (baseline and Phase 1) as well as the intervention period. Phase 1 was initiated when BPCI was announced and encompasses the one year period prior to the BPCI intervention period. Because BPCI participants started implementing changes during Phase 1 in preparation for Phase 2, the risk-bearing or intervention phase, the Phase 1 period was excluded from the DiD baseline. Including Phase I in the DiD baseline would likely underestimate the BPCI effect given that participants started to prepare for the intervention during that period. Thus, the DiD compares changes in outcomes from the baseline period to the intervention period.

- The DiD baseline period was from October 2011 through September 2012.
- Phase 1 was from October 2012 through September 2013.

[13] While the DiD model controls for unobserved heterogeneity that is fixed over time, there is no guarantee that this unobserved heterogeneity is, in fact, fixed. It could be the case, for example, that providers with improving outcomes are relatively more likely to sign up for the Model, introducing correlation between BPCI participation and outcomes, which could bias the results.

- The BPCI intervention period was from October 2013 through September 2016.

Consider the following linear model to illustrate the DiD calculation in a regression framework:

$$Y_{i,k,t} = \alpha + \beta_1 BPCI_{i,k} + \beta_2 T_t + \delta BPCI_{i,k} \cdot T_t + \boldsymbol{X}_{i,k,t}'\beta + u_{i,k,t}$$

Where $Y_{i,k,t}$is the outcome of interest for individual *i* with provider *k* in quarter *t*, $BPCI_{i,k}$ is an indicator variable taking the value of 1 if individual *i* was treated by a BPCI provider, T_t indicates the period (i.e., baseline, Phase I, or intervention), and $X_{i,k,t}$ are beneficiary demographics, clinical characteristics observed before hospitalization, and provider characteristics. The vector β is a vector of regression coefficients that captures the impact of risk factors $\boldsymbol{X_{i,k,t}}$ on the outcome of interest. The regression coefficient β_1 captures any inherent, time invariant differences between the control and the treatment groups, β_2 provides an estimate of the potential time trends in the outcome of interest over the period before and after the intervention that is common to both the control and treatment groups, while $u_{i,k,t}$ represents a random error term. In this linear example, the DiD estimate is coefficient δ, which determines the differential in outcome Y experienced by beneficiaries receiving services from BPCI providers during the intervention period relative to beneficiaries receiving services from providers in the comparison group. Thus, BPCI impact estimates, which are the estimated causal effects that are due to BPCI, are referred to as "DiD estimates" throughout the report except otherwise noted.

We used multivariate regression models to control for differences in beneficiary demographics, clinical characteristics, and prior care use before the hospitalization, along with provider characteristics that might be related to the outcome (see Exhibit 5). We used a variety of empirical specifications including ordinary least squares (OLS) and logistic regressions, duration, and two-part models. Models were estimated depending on the type and characteristics of the outcome measure.

Exhibit 5: Predictive Risk Factors Used to Risk Adjust Claims Outcomes

Domain	Variables
Service Mix	• Alternative specifications ▪ Anchor MS-DRG ▪ MS-DRG group: anchor MS-DRG with and without complications grouped together
Patient Demographics & Enrollment	• Age (under 65, 65-79, 80+) • Gender • Medicaid status • Disability status • Alignment to Medicare Shared Savings Program or Pioneer ACO during BPCI episode
Prior health conditions	• Alternative specifications ▪ HCC indicators from qualifying services and diagnoses from claims and data for six months preceding the anchor admission or qualifying stay ▪ HCC indicators aggregated to risk variable groups (RV-HCC) according to NQF measure 1789 (**Appendix G** shows a crosswalk from 2013 HCC indicators to RV-HCC) ▪ HCC index, HCC indicators weighted by their relative weight in the 2013 CMS-HCC model
Utilization measures preceding the start of the anchor stay/ qualifying inpatient stay	• Alternative specifications ▪ Binary indicators for utilization of ED, inpatient, SNF, nursing facility, IRF, HHA services in the **six months** preceding the start of the episode ▪ Number of days of ED, inpatient, SNF, IRF, HHA service use in the **one month** preceding the start of the episode, and ever in a NF/SNF in the six months preceding the start of the episode ▪ Number of days of ED, inpatient, SNF, IRF, HHA service use in the **six months** preceding the start of the episode, and ever in a NF/SNF in the six months preceding the start of the episode
Geography	• Alternative specifications ▪ State indicators ▪ Census region indicators
Provider Characteristics	• Size • Ownership status • For Model 2 episodes, the hospital was in a Comprehensive Care for Joint Replacement Model market • For Model 3 episodes, the hospital of the qualifying hospitalization or the episode initiator (SNF/HHA) was located in a Comprehensive Care for Joint Replacement Model market

MS-DRG=Medicare severity diagnosis related group, ACO=accountable care organization, HCC=hierarchical condition category, NQF=National Quality Forum, ED=emergency department, SNF=skilled nursing facility, IRF=inpatient rehabilitation facility, HHA=home health agency, NF=institutional nursing facility

We attempted to construct a comparison group of providers that closely matched BPCI providers in key characteristics, however, we could not guarantee that BPCI and comparison providers would have parallel trends during the baseline period for every outcome. This is because we could not include every outcome in the matching and some outcomes fluctuate over time. Because it was not feasible to test the null hypothesis that BPCI participants and comparison providers had parallel trends during the baseline for every outcome and every strata, we focused on the claim-based quality outcomes and the total payment for the inpatient stay plus 90 days post-discharge in all strata. If we rejected the null hypothesis that there were parallel trends in the baseline (at the .10 level) and the DiD estimate was statistically significant (positive or negative), we attempted to find an alternative risk adjustment model where we failed to reject the null hypothesis of parallel trends. We also tested the null hypothesis of parallel trends in baseline for any outcomes where there was

visual evidence that the direction of change from baseline to intervention for BPCI differed from the change for the comparison group. In this report, we report all DiD estimates, but we note when we rejected the null hypothesis that there were parallel trends in baseline.

There are some outcomes for which we do not report the DiD estimate because of small sample sizes. We report DiD estimates for each outcome if the sample exceeds 30 BPCI episodes during the intervention period for outcomes evaluated using duration, logistic, and OLS models. In contrast, we used a minimum of 100 BPCI episodes with a positive value of the outcome during the intervention period to report DiD estimates for outcomes using two-part models. Some outcomes, including IRF and LTCH payments during the 90-day post-discharge period and payment outcomes that are stratified by bundle length, suffer from small sample sizes, and consequently, DiD estimates for these outcomes typically were not reported.

C. Beneficiary Survey

1. Study population

This annual report includes the results from the fourth through eighth waves of the beneficiary survey, covering episodes starting between May 2015 and November 2016. In this section, we describe the sampling for these five waves. The sample includes 21 of the 48 Model 2 ACH clinical episodes, which included 91% of all BPCI Model 2 ACH episodes that occurred between the start of the initiative and the fourth quarter of 2016. Results based on this sample approximate differences between patients treated by BPCI ACH EIs and those treated by comparison hospitals.

The survey sample was constructed at the clinical episode level beginning in Wave 4, which was the first wave with sufficient BPCI volume to support clinical episode level sampling. Each survey stratum comprised sampled patients with a unique combination of BPCI Model, type of EI, and clinical episode (e.g., Model 2 – ACH – MJRLE).

We used a stratified random sampling approach to select patients for each stratum. Within each stratum, BPCI and comparison patients were matched within cells. For strata that could achieve enough completed surveys (310) with a single wave of data collection, cells were defined by presence of major complication or comorbidity (MCC), patient age group, provider size, and hospital academic status. For strata that required more than one wave to reach 310 completes, we omitted hospital academic status from the cell matching procedure, due to the smaller number of such cases. Because CMMI's Comprehensive Care for Joint Replacement (CJR) model began April 1, 2016, we excluded MJRLE episodes initiated by CJR hospitals from the Model 2 ACH comparison sample beginning in Wave 7 (May/June 2016).

For all waves, Model 2 strata were constructed using Medicare FFS claims from two "rolling" one month samples, and the beneficiaries selected for the two one-month samples received their surveys one month apart.[14] For example, the first month's sample for Wave 4 used claims from May 2015 that were pulled in early June 2015, and surveys were mailed in the first week of July 2015. The sample for the second month of Wave 4 used claims from June 2015 that were pulled

[14] One month of claims provides insufficient volume to reach the necessary sample size for the clinical episodes used to define the strata.

in early July 2015, and surveys were mailed the first week of August 2015. This rapid sampling process was deliberately used to reduce recall bias. This does, however, limit the sample to patients whose claims were filed within one month of discharge.[15] The time period covered by each survey wave is listed in Exhibit 6.

Exhibit 6: Period Covered by Each Beneficiary Survey Wave

Survey Wave	Months in which sampled episodes were initiated
4	May/June 2015
5	October/November 2015
6	February/March 2016
7	May/June 2016
8	October/November 2016

Across all waves and strata, the overall survey response rate was 46.4%. Response rates for individual strata ranged from 34.6% (urinary tract infection) to 73.4% (major joint replacement of the upper extremity). Response rates for all 21 strata are reported in **Appendix H**.

2. Outcome definitions

Exhibit 7 summarizes the key survey-based measures by domain.

Exhibit 7: Survey-based Measures of the Impact of BPCI

Domain	Outcome Measures
Change in Functional Status	Bathing, dressing, using the toilet, or eating
	Planning regular tasks
	Use of a mobility device
	Walking without rest
	Using stairs
	Physical/emotional problems limit social activities
	Pain limits regular activities
Overall Health	Depression
	Overall physical health
	Overall mental health

[15] Although claims submitted within one month may not represent the entire Medicare population, due to provider delays in submitting claims, this issue should affect BPCI and comparison samples equally and not bias our estimates.

Domain	Outcome Measures
Care Experience	Never received conflicting medical advice
	Services always appropriate for level of care patient needed
	Medical staff always spoke in patient's preferred language
	Discharged at the right time
	Medical staff took patient's preferences into account in deciding what health care services were needed after leaving the hospital
	Patient or caregiver had a good understanding of how to take care of patient before going home
	Medical staff clearly explained how to take medications before going home
	Medical staff clearly explained what follow-up appointments or treatments would be needed before going home
	Patient has been able to manage health needs since returning home
Overall Satisfaction	Overall, patient was "quite a bit" or "extremely" satisfied with recovery since leaving the hospital

3. Survey Analytic Methodology

a. Weighting

We adjusted the survey data with both sampling weights and nonresponse weights.[16] We calculated the sampling weight as the inverse of the selection probability within each sampling cell. We also used nonresponse weights to control for potential bias from differential response rates between BPCI and comparison groups. Nonresponse weights were calculated for all survey respondents who answered at least one question and reflect the inverse of the probability of response among eligible members of the sample (i.e., with deceased respondents removed) within each sampling stratum.

Nonresponse-adjusted weights should improve balance between BPCI and matched comparison samples on the variables used to define the sampling cells. However, differential ineligibility (i.e., rates of deceased) and nonresponse on individual survey items can still create imbalance. To address this, we included the variables used to define cells in the risk-adjustment model (discussed below).

b. Risk-adjustment

We used multivariate regression models to control for differences in patient demographics, clinical characteristics, prior care use before the hospitalization, and provider characteristics that might affect the survey-based outcomes (see Exhibit 8).

Because all survey measures were collapsed into binary variables, all models were estimated using logit regression.

[16] For both the BPCI and comparison groups, the sampling weights sum to the population size of the BPCI group in each wave.

We estimated a fully interacted model that allowed the regression coefficients for each risk-adjustment variable, and the difference between BPCI and comparison respondents, to vary by stratum. This is shown in the equation below:

$$Y_{i,j,k,t} = \delta_k BPCI_{i,j} * S_{i,j,k} + \beta_k X_{i,j} * S_{i,j,k} + DRG_i + T_i + \varepsilon_{i,j,k,t}$$

$Y_{i,j,k,t}$ is the outcome of interest for individual *i*, treated at provider *j*, in stratum *k*, during time *t*. $X_{i,j}$ refers to the risk-adjustment variables listed in Exhibit 8, DRG_i indicates individual indicators for each MS-DRG,[17] T_i reflects wave (time) fixed effects, $BPCI_{i,j}$ is an indicator for patients who were treated by a BPCI participating hospital, and $S_{i,j,k}$ is a set of indicator dummies for each of the K strata. The relationship between Y and X (indicated by β_k) is unique for each stratum (e.g., HCC score may affect changes in functional status differently for patients with sepsis than patients with MJRLE). δ_k indicates the difference between BPCI and comparison respondents in stratum *k*.

The average difference between BPCI and comparison respondents across all strata is equal to the sum of all K values of δ_k, with each δ_k weighted according to the proportion of BPCI episodes in stratum *k* relative to the total number of BPCI episodes across all K strata. That is, each stratum's estimate is weighted according to the size of the stratum relative to the entire underlying Model 2 ACH population represented by the survey.[18] Standard errors are clustered at the hospital level.

[17] Because DRGs are unique to each stratum, these do not need to be interacted with the stratum indicators. Moreover, because they are unique within strata, they function as "stratum fixed effects" without the need for separate, non-interacted strata indicators in the equation.

[18] As a simple example, suppose there were two strata: sepsis and MJRLE. Suppose sepsis had 1000 episodes during survey waves 4-8 and MJRLE had 2000 episodes during the same time. The full BPCI population covered by the survey is then 3000. In such a case, the sepsis weight would be equal to 1000/3000 = 1/3 and the MJRLE weight would be equal to 2000/3000 = 2/3.

Exhibit 8: Predictive Risk Factors Used to Risk Adjust Survey Outcomes

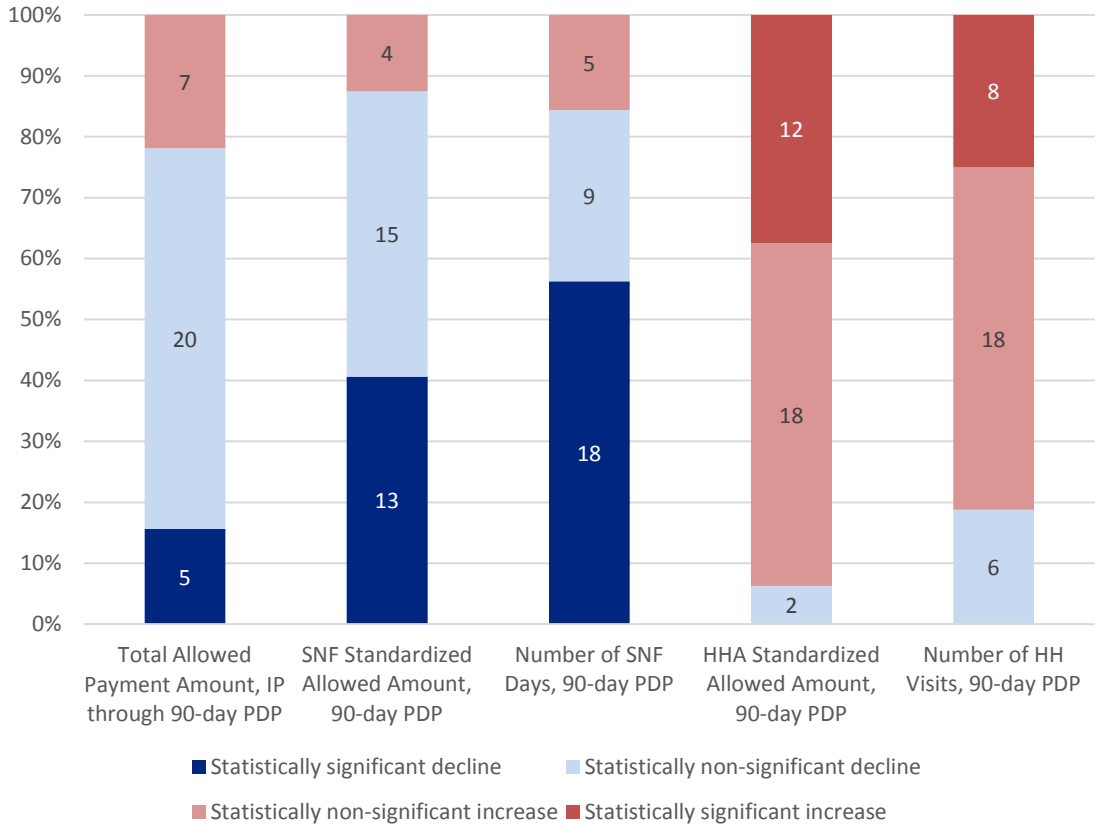
Domain	Variables
Other Risk Factors	• Self-reported baseline functional status • Survey completed by proxy
Service Mix	• Anchor MS-DRG • Hip fracture (MJRLE and hip & femur procedures only) • Ischemic versus intracerebral stroke (stroke only)
Patient Demographics and Enrollment	• Age (under 65, 65-74,75-84, 85+) • Gender • Dual Medicare/Medicaid status
Prior health conditions	• HCC Score
Utilization measures preceding the start of the anchor stay/qualifying inpatient stay	• SNF days in prior 90 days • Inpatient days in prior 90 days • Admitted to hospital from institutional facility
Provider Characteristics	• Size • Ownership status • Academic status

MS-DRG=Medicare severity diagnosis related group, MJRLE=major joint replacement of the lower extremity, HCC=hierarchical condition category, SNF=skilled nursing facility

III. Model 2 Impact of BPCI

Across the 32 Model 2 clinical episodes initiated by acute care hospitals during the first three years of the BPCI initiative (Q4 2013 through Q3 2016) for which we had sufficient sample, five had a statistically significant relative decline in total payments for the inpatient stay plus 90 days post-discharge (Exhibit 9).[19] Another 20 had a decline in total payments that did not reach statistical significance. There was a strong trend toward reduced SNF payments, consistent with fewer SNF users and fewer days for those with a SNF stay. There were higher HHA payments, consistent with more home health users. Reduced institutional PAC use and increased HHA use contributed to lower total episode payments. In general, quality of care was maintained under BPCI. Additionally, there was no clear evidence that participating providers changed their mix of patients under the initiative to reduce episode costs. (See **Appendix I** for more detail on all results.)

Exhibit 9: Relative Change in Payment and Utilization by Clinical Episode, Baseline to Intervention, Model 2 ACH EIs, Q4 2013 - Q3 2016



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark red shading. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark blue shading. Medicare payment outcomes are standardized to remove the

[19] Please note, the term clinical episode refers to the 48 groups of MS-DRGs that participants could choose under BPCI. MJRLE and CABG were also divided for analytic purposes to better represent differences in payments and treatment. Stratified estimates for Model 2 fracture/non-fracture MJRLE episodes and emergent/non-emergent CABG episodes are in **Appendix I**. Stratified estimates for Model 3 SNF fracture/non-fracture MJRLE episodes are in **Appendix K**.

effect of geographic and other adjustments. These payment measures are not conditional upon use of the service, whereas these utilization measures are conditional upon the use of the service.

ACH = acute care hospital, EI = episode initiator, PDP = post-discharge period, IP = inpatient, SNF = skilled nursing facility, HHA = home health agency

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

A. Sample Characteristics

Before discussing the impact of BPCI on payments, utilization, and quality, we present some basic statistics by clinical episode (see Exhibit 10) to better understand the BPCI sample used for the impact analysis. For the clinical episodes for which we had sufficient sample size, the number of matched EIs ranged from 26 to 303. These EIs initiated from 1,089 to 97,922 episodes, depending on clinical episode, during the first three years of the initiative. Because providers were allowed to join BPCI over an extended period, these data represent an average of five quarters of participation. Providers can also stop participating in a clinical episode on a quarterly basis or can stop participating in the initiative altogether at any time. During the first three years, 81 of 422 Model 2 ACH EIs (19%) withdrew entirely from the initiative. Among the 406 matched EIs participating in any of the 32 clinical episodes analyzed in this report, approximately 52% stopped participating in at least one clinical episode during the first three years. For four clinical episodes, hospital EIs that stopped participating in the clinical episode had contributed 50% or more of the episodes during the intervention period. An additional seven clinical episodes had hospital EIs that stopped participating that contributed over 40% of the episodes.

Exhibit 10: Characteristics of the Matched BPCI Providers Included in the DiD Estimates, Model 2, Q4 2013 – Q3 2016

Clinical Episode	Matched EIs (#)	Matched Intervention Period Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (%)	Episodes from EIs that Terminated (%)
Acute myocardial infarction	93	5,337	5	41%	36%
Cardiac arrhythmia	70	6,029	5	51%	45%
Cardiac valve	31	3,957	6	48%	53%
Cellulitis	79	5,474	5	43%	50%
Cervical spinal fusion	34	1,190	5	44%	34%
Congestive heart failure	173	31,858	5	30%	29%
COPD, bronchitis, asthma	133	18,331	6	31%	33%
Coronary artery bypass graft	43	3,242	6	28%	32%
Diabetes	45	1,423	5	38%	24%
Esophagitis, gastroenteritis & other digestive disorders	58	4,104	4	53%	48%
Fractures of the femur and hip or pelvis	47	1,092	5	34%	36%
Gastrointestinal hemorrhage	58	4,386	4	66%	51%
Gastrointestinal obstruction	51	1,735	4	53%	47%
Hip & femur procedures except major joint	101	7,446	5	36%	26%
Lower extremity and humerus procedure except hip, foot, femur	37	1,089	6	38%	28%
Major bowel procedure	46	3,029	5	39%	37%
Major joint replacement of the lower extremity	303	97,922	6	19%	14%
Major joint replacement of the upper extremity	26	1,337	5	31%	21%
Medical non-infectious orthopedic	94	6,588	5	43%	35%
Nutritional and metabolic disorders	57	2,727	4	47%	51%
Other respiratory	62	4,700	5	42%	32%
Other vascular surgery	36	1,590	5	44%	38%
Percutaneous coronary intervention	45	4,745	5	24%	23%
Renal failure	75	7,474	5	43%	45%

Clinical Episode	Matched EIs (#)	Matched Intervention Period Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (%)	Episodes from EIs that Terminated (%)
Revision of the hip or knee	32	1,146	5	50%	37%
Sepsis	119	26,046	5	45%	42%
Simple pneumonia and respiratory infections	132	22,556	6	28%	24%
Spinal fusion (non-cervical)	46	3,417	5	39%	31%
Stroke	77	11,357	5	38%	23%
Syncope & collapse	37	1,364	5	43%	46%
Transient ischemia	30	1,099	6	43%	41%
Urinary tract infection	83	8,010	5	35%	27%

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q3 2016 for BPCI providers. This exhibit is limited to the matched BPCI providers used to calculate the DiD results in the reminder of this section.

B. Key Payment, Utilization, and Quality of Care Outcomes

1. How have the average standardized allowed amounts (Medicare payments and coinsurance/copayments combined) changed under BPCI?

In the three years since implementation of BPCI, the total allowed payment amount declined from the baseline to the intervention period for the majority of BPCI clinical episodes relative to the comparison group. Reduced SNF payments were the major contributor to these declines, which were often accompanied by concurrent increases in HHA payments (Exhibits 11 & 12).

The total allowed payment amount for the inpatient stay plus 90 days post-discharge decreased in 25 of 32 clinical episodes (78%); the decline for five clinical episodes was statistically significant at the 5% level. The decrease in the total allowed payment amount included in the bundle for 90-day episodes, which excludes certain readmissions and Part B services and aligns with the target price definition, was also statistically significant for the same five clinical episodes. The average reduction in Medicare payments across these five clinical episodes was 6.7% of what was expected without BPCI (Exhibit 13). These declines were driven by reductions in SNF payments, which decreased in 28 of 32 clinical episodes (88%); the decline in 13 clinical episodes was statistically significant ($p<0.10$). The HHA standardized allowed amount increased relative to the change in the comparison group in 30 of the 32 clinical episodes (94%); the increase in 12 clinical episodes was statistically significant ($p<0.10$). IRF payments decreased relative to the change in the comparison group for 15 of the 21 (71%) clinical episodes with sufficient sample size to examine; the decline in five clinical episodes was statistically significant ($p<0.10$). These patterns suggest that BPCI resulted in reduced institutional PAC use and increased HHA use, a strategy mentioned by numerous BPCI participants in site visits and interviews.

We identified the 10 clinical episodes with historically the highest proportion of PAC payments to determine if these clinical episodes were more likely to experience greater declines in total payments. Indeed, total allowed payments for the inpatient stay and the 90-day PDP declined relative to the change in the comparison group in all but one of these clinical episodes; the decline in four clinical episodes was statistically significant ($p<0.05$). SNF payments declined in eight of these clinical episodes: the decline was statistically significant for five clinical episodes ($p<0.05$). HHA payments increased in nine of these clinical episodes; the increase was statistically significant for six ($p<0.10$). (See **Appendix J** for more details on the highest PAC payment clinical episodes.)

Exhibit 11: Impact of BPCI on Medicare Allowed Payment Outcomes, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013-Q3 2016

Clinical Episode	Number of Episodes Q4 2013 - Q3 2016	Total Amount Included in Bundle Definition[1]	Total Allowed Payment Amount, IP through 90-day PDP	Readmissions Standardized Allowed Amount, 90-day PDP	SNF Standardized Allowed Amount, 90-day PDP[2]	HHA Standardized Allowed Amount, 90-day PDP[2]	IRF Standardized Allowed Amount, 90-day PDP[2]
Acute myocardial infarction	5,337	-$145	-$281	-$11	-$66	$45	$156
Cardiac arrhythmia	6,029	-$149	-$177	$192	-$319	$48	-$7
Cardiac valve	3,957	-$813	-$1,268	-$438	-$298	$114	-$280
Cellulitis	5,474	-$291	-$478	$19	**-$550**	$68	-$138
Cervical spinal fusion	1,190	$464	$812	-$325	-$203	$121	$546
Congestive heart failure	31,858	-$429	-$400	-$218	**-$302**	$51	-$85
COPD, bronchitis, asthma	18,331	-$386	-$395	$30	-$171	**$105**	-$102
Coronary artery bypass graft	3,242	-$1,092	-$571	$81	-$559	$9	**-$689**
Diabetes	1,423	-$800	-$974	-$174	-$744	$159	
Esophagitis, gastroenteritis & other digestive disorders	4,104	-$705	-$813	-$7	**-$832**	$89	
Fractures of the femur and hip or pelvis[3]	1,092	-$811	-$813	$533	-$1,038	$111	-$410
Gastrointestinal hemorrhage	4,386	-$426	-$259	$335	**-$720**	**$118**	
Gastrointestinal obstruction	1,735	$227	$610*	$352	-$111	$119	
Hip & femur procedures except major joint[3]	7,446	**-$1,828**	**-$1,832**	$65	**-$2,182**	**$188**	$108
Lower extremity and humerus procedure except hip, foot, femur[3]	1,089	-$63	-$309	-$86	$629	-$181	
Major bowel procedure	3,029	$221	-$357	$439	-$128	**$199**	-$12
Major joint replacement of the lower extremity[3]	97,922	**-$1,115**	**-$1,222**	-$59	**-$762**	$61	**-$370**
Major joint replacement of the upper extremity	1,337	-$498	$294	**$516**	-$116	$36	

Clinical Episode	Number of Episodes Q4 2013 - Q3 2016	Total Amount Included in Bundle Definition[1]	Total Allowed Payment Amount, IP through 90-day PDP	Readmissions Standardized Allowed Amount, 90-day PDP	SNF Standardized Allowed Amount, 90-day PDP[2]	HHA Standardized Allowed Amount, 90-day PDP[2]	IRF Standardized Allowed Amount, 90-day PDP[2]
Medical non-infectious orthopedic[3]	6,588	**-$1,641**	**-$1,884**	-$290	**-$1,480**	**$148**	-$280
Nutritional and metabolic disorders[3]	2,727	$854	$1,045	$208	-$252	**$161**	
Other respiratory	4,700	$51	-$446	-$236	-$48	**$121**	**-$251**
Other vascular surgery	1,590	-$136	$726*	$225	$245	$8	$301
Percutaneous coronary intervention	4,745	-$15	$230*	**$697**	$25	$2	
Renal failure	7,474	-$810	-$668	$37	-$551	$47	-$122
Revision of the hip or knee	1,146	-$402	$332	$512	-$451	$105	
Sepsis[3]	26,046	-$382	-$413	$36	**-$487**	**$104**	-$16
Simple pneumonia & respiratory infections	22,556	-$361	-$203	$92	-$212	$50	$18
Spinal fusion (non-cervical)	3,417	-$675	-$1,181	-$174	-$464	-$55	-$369
Stroke[3]	11,357	-$466	-$297	-$25	$99	$55	-$364
Syncope & collapse[3]	1,364	$140	-$96	-$236	-$431	$116	
Transient ischemia	1,099	**-$1,973**	**-$2,541**	-$659	**-$995**	$76	
Urinary tract infection[3]	8,010	**-$908**	**-$860***	$13	**-$917**	**$114**	$9

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented due to insufficient sample size. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. PDP = post-discharge period. IP = inpatient.

[1] The total amount included in the bundle definition is based on only the 90-day episodes.

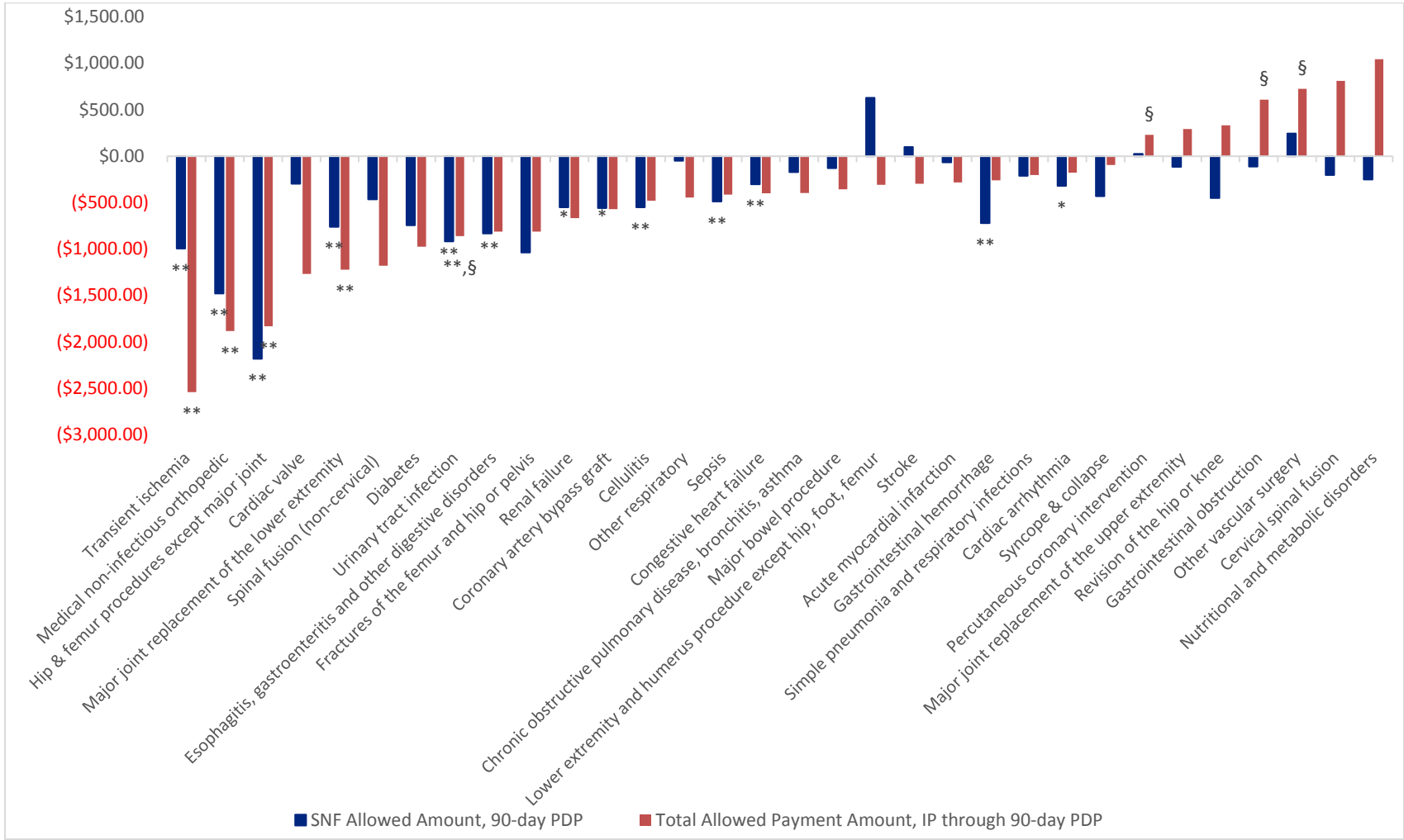
[2] These payment measures are not conditional upon use of the service.

[3] This clinical episode is one of the 10 clinical episodes with the historically highest proportion of PAC payments.

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

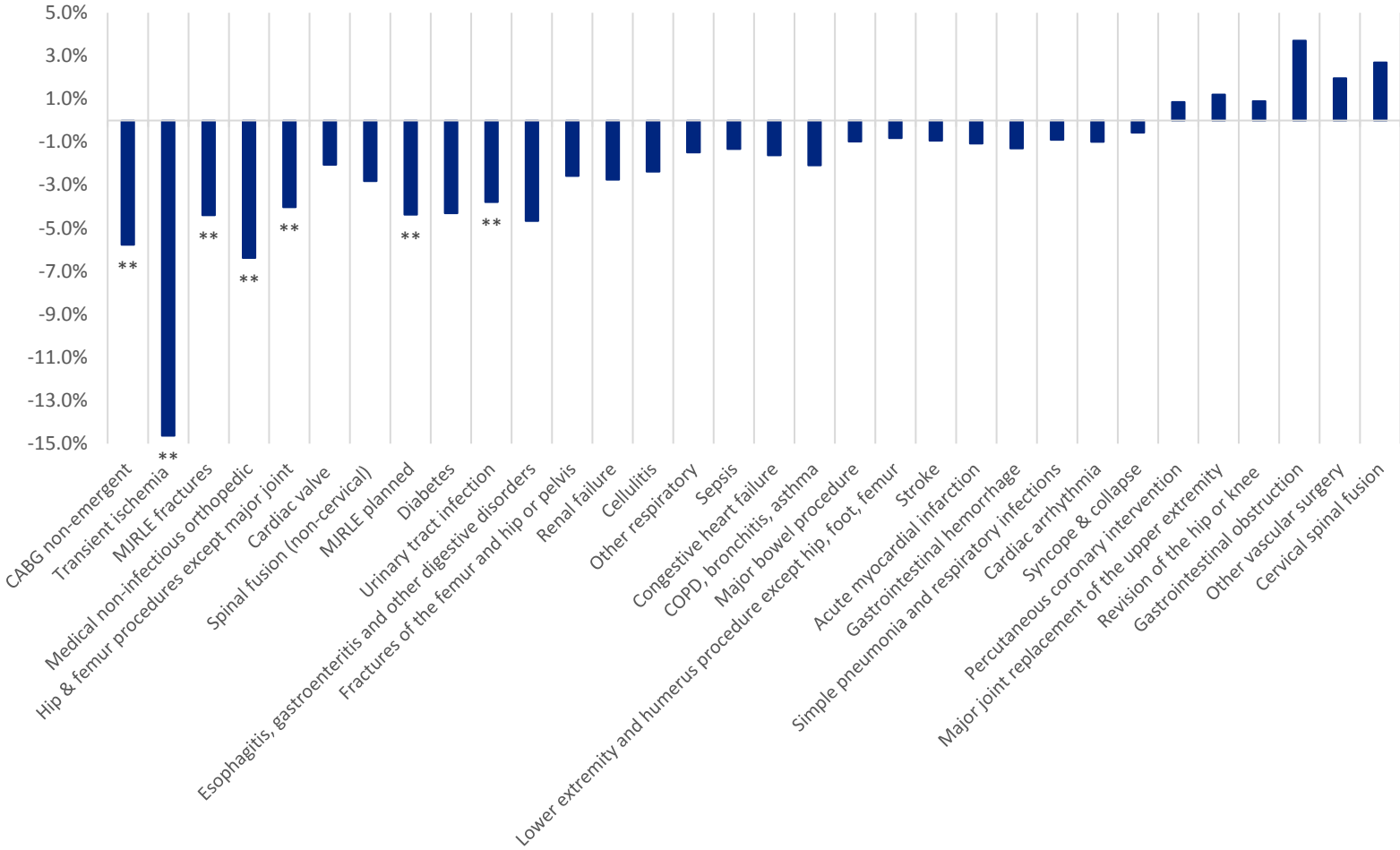
Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Exhibit 12: Impact of BPCI on SNF Payments and Total Allowed Payments, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013 - Q3 2016



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. These payment measures are not conditional upon the use of the service.
**Indicates DiD estimates are statistically significant at the 5% level.
*Indicates DiD estimates are statistically significant at the 10% level.
§ Data from the baseline period shows BPCI and matched comparison providers 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 episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Exhibit 13: Percent Change in BPCI Episode Payments from What Payments Would have Been Absent BPCI, by Clinical Episode, Model 2 ACH, Q4 2013 - Q3 2016



Note: The payments in this exhibit are the risk-adjusted standardized allowed amounts for the inpatient stay plus 90-day PDP. 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). Results are sorted by the allowed payment amount DiD estimate. EI=episode initiator, ACH=acute care hospital.

**Indicates DiD estimates are statistically significant at the 5% level.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

2. How have the services changed under BPCI?

Consistent with the changes in payments, we observed a shift in the use of PAC services. For most clinical episodes there was no statistically significant change in the proportion of patients discharged to PAC, however, among patients who received any PAC, the proportion discharged to institutional PAC declined in 28 of the 32 clinical episodes (Exhibits 14 and 15). The decline was statistically significant ($p<0.10$) in eight clinical episodes. There appeared to be an overall decline in the use of SNF care and an increase in home health across the clinical episodes (Exhibits 14 and 16). The number of SNF days for BPCI patients who used SNF care declined relative to SNF days for the comparison group in 27 clinical episodes; the decline was statistically significant in 18 clinical episodes ($p<0.10$). The number of home health visits among BPCI patients who had at least one visit increased from the baseline to the intervention period relative to the change in the comparison group in 26 of the 32 clinical episodes; this difference was statistically significant in eight clinical episodes ($p<0.10$).

The shift in PAC use is particularly notable in the 10 clinical episodes with the highest proportion of historical PAC payments (Exhibit 14 and **Appendix J)**. In all of these clinical episodes, the number of SNF days for SNF users declined relative to the comparison group: the decline was statistically significant for seven clinical episodes ($p<0.10$). Although an increase in HH visits does not necessarily result in an increase in HHA payments, the number of HH visits for HHA users increased; the increase was statistically significant for three clinical episodes ($p<0.10$).

Across all clinical episodes, there was no systematic change in the inpatient hospital length of stay (LOS). Hospitals have had financial incentives to reduce LOS since the inpatient prospective payment system was implemented in 1983. Furthermore, while reducing hospital LOS may reduce internal hospital costs (and the BPCI initiative allows such savings to be shared with partnering providers), it will have no effect on NPRA. Therefore, it is not surprising to see minimal change in this utilization outcome. Even among the 10 clinical episodes with over 40% of their historical payments attributable to the anchor stay (**Appendix J**), there was no clear pattern for changes in LOS.

Exhibit 14: Impact of BPCI on Utilization Outcomes, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Q4 2013-Q3 2016	Percent Discharged to PAC	Percent Discharged to an Institution out of Those who Received any PAC	Anchor Hospital Stay LOS	Number of HH Visits, 90-day PDP[1]	Number of SNF Days, 90-day PDP[2]	Number of Institutional Days, 90-day PDP[3]
Acute myocardial infarction	5,337	-0.4 pp	-1.4 pp	-0.1	0.1	-0.4	-0.3
Cardiac arrhythmia	6,029	-0.9 pp	0.2 pp	0.0	**1.1**	**-2.9**	**-3.3**
Cardiac valve[5]	3,957	1.4 pp	**-10.5 pp**	0.0	-0.4	1.0	1.7
Cellulitis	5,474	0.1 pp	-1.1 pp	0.0	-0.1	-2.3	-2.5
Cervical spinal fusion[5]	1,190	-0.3 pp	2.2 pp	0.0	0.7	-0.4	-0.6
Congestive heart failure	31,858	0.7 pp	-0.4 pp	0.0	**0.8**	**-1.6**	**-1.4**

Clinical Episode	Number of Episodes Q4 2013-Q3 2016	Percent Discharged to PAC	Percent Discharged to an Institution out of Those who Received any PAC	Anchor Hospital Stay LOS	Number of HH Visits, 90-day PDP[1]	Number of SNF Days, 90-day PDP[2]	Number of Institutional Days, 90-day PDP[3]
COPD, bronchitis, asthma	18,331	**1.6 pp**	-1.9 pp	0.0	**0.7**	-1.5	**-1.6**
Coronary artery bypass graft[5]	3,242	-2.7 pp	**-6.0 pp**	0.0	0.7	-1.1	-0.4
Diabetes	1,423	-0.1 pp	-1.3 pp	0.0	-1.7	**-6.3**	-4.6
Esophagitis, gastroenteritis & other digestive disorders	4,104	-0.2 pp	**-7.5 pp**	-0.1	1.1	**-6.5**	**-6.1**
Fractures of the femur and hip or pelvis[4]	1,092	-0.9 pp	-2.0 pp	0.0	0.9	-3.0	-2.6
Gastrointestinal hemorrhage	4,386	-0.1 pp	-2.3 pp	0.0	-0.2	**-5.7**	**-5.2**
Gastrointestinal obstruction	1,735	0.7 pp	**-10.5 pp**	0.0	1.1	1.0	2.0
Hip & femur procedures except major joint[4]	7,446	-0.1 pp	-0.4 pp	0.0	**1.1**	**-4.2**	**-4.4**
Lower extremity and humerus procedure except hip, foot, femur[4]	1,089	-3.9 pp	**9.7 pp**	-0.1	0.2	-2.4	-1.4
Major bowel procedure[5]	3,029	**5.3 pp**	-1.6 pp	0.3	0.4	-2.0	-0.9
Major joint replacement of the lower extremity[5]	97,922	**-3.0 pp**	**-5.6 pp**	**-0.1**	0.1	**-2.0**	**-1.4**
Major joint replacement of the upper extremity[5]	1,337	1.6 pp	1.5 pp	-0.1	0.8	-1.5	-1.7
Medical non-infectious orthopedic[4]	6,588	-2.2 pp	-2.1 pp	0.0	0.8	**-3.0**	**-2.6**
Nutritional and metabolic disorders[4]	2,727	3.0 pp	**-6.2 pp**	0.0	0.2	-1.8	-1.0
Other respiratory	4,700	1.3 pp	-1.7 pp	0.1	1.0	-1.5	-1.2
Other vascular surgery	1,590	-3.8 pp	-3.7 pp	-0.4	-0.2	1.7	2.1
Percutaneous coronary intervention[5]	4,745	-0.9 pp	-1.3 pp	-0.1	0.4	0.3	0.0
Renal failure	7,474	0.5 pp	-1.6 pp	0.0	0.2	**-2.9**	-1.9

Clinical Episode	Number of Episodes Q4 2013-Q3 2016	Percent Discharged to PAC	Percent Discharged to an Institution out of Those who Received any PAC	Anchor Hospital Stay LOS	Number of HH Visits, 90-day PDP[1]	Number of SNF Days, 90-day PDP[2]	Number of Institutional Days, 90-day PDP[3]
Revision of the hip or knee[5]	1,146	0.9 pp	**-11.2 pp**	0.0	0.4	2.6	2.9
Sepsis[4]	26,046	0.7 pp	**-2.6 pp**	0.2	0.5	**-1.7**	**-1.5**
Simple pneumonia & respiratory infections	22,556	1.2 pp	-0.8 pp	0.0	0.3	**-1.6**	-1.4
Spinal fusion (non-cervical)[5]	3,417	-4.7 pp	-0.1 pp	0.1	-0.6	-2.7	-1.7
Stroke[4]	11,357	0.3 pp	-1.1 pp	-0.1	0.8	-0.8	0.0
Syncope & collapse[4]	1,364	-1.2 pp	-1.1 pp	-0.1	0.7	**-5.8**	-4.3
Transient ischemia	1,099	-0.1 pp	-3.0 pp	0.0	0.6	-5.5	-5.1
Urinary tract infection[4]	8,010	0.7 pp	-2.1 pp	0.0	0.4	**-3.9**	**-3.7**

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. PAC = post-acute care. PDP = post-discharge period. LOS = length of stay. HH = home health. SNF = skilled nursing facility.

[1] Beneficiaries must have spent a minimum of one day in a HH setting during the 90-day PDP.

[2] Beneficiaries must have spent a minimum of one day in a SNF setting during the 90-day PDP.

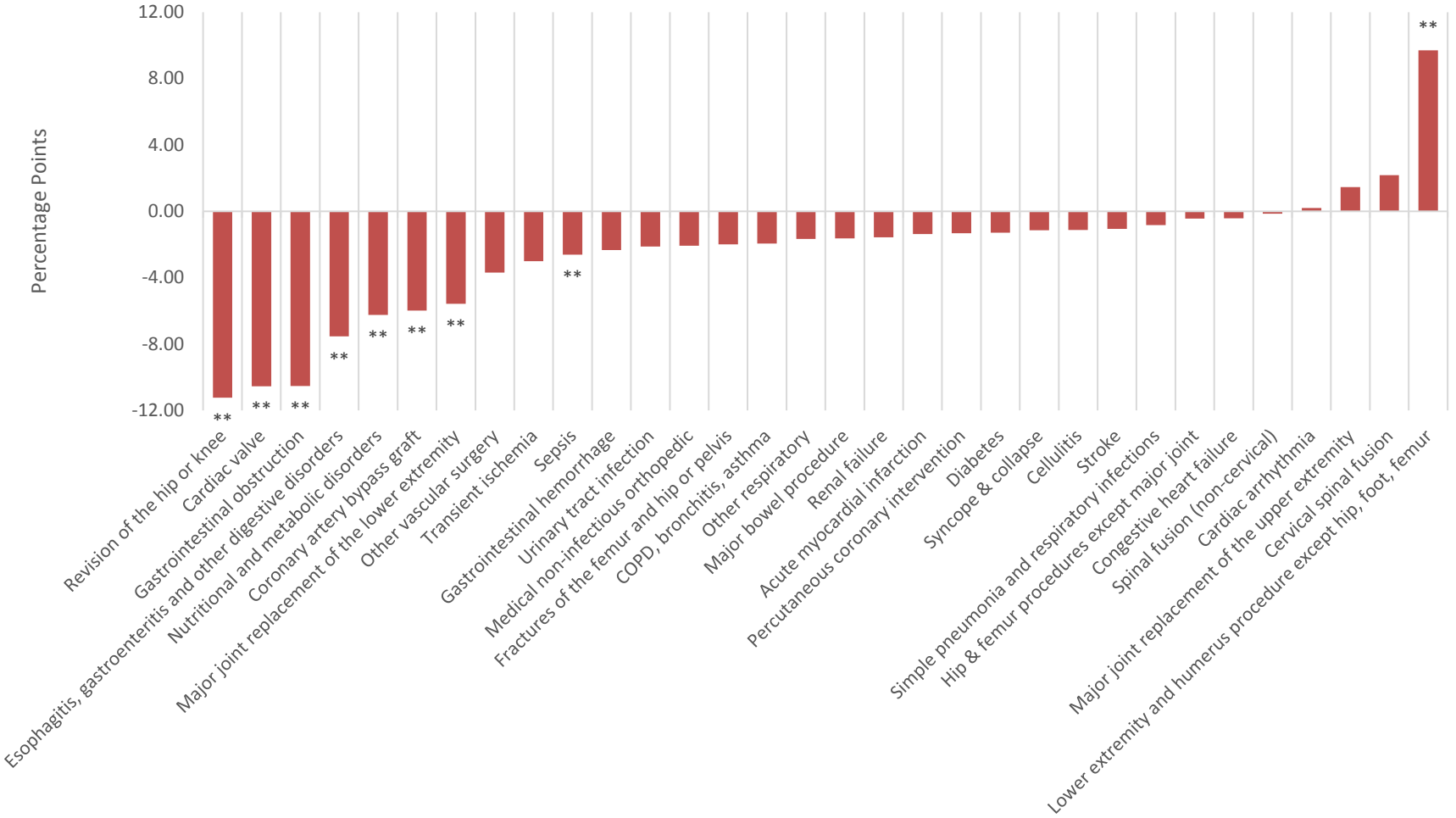
[3] Beneficiaries must have spent a minimum of one day in a SNF, IRF, or LTCH setting during the 90-day PDP.

[4] This clinical episode is one of the 10 with the highest proportion of total baseline episode payments driven by PAC payments.

[5] This clinical episode is one of the 10 with the highest proportion of total baseline episode payments driven by the anchor inpatient stay (>40%).

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Exhibit 15: Impact of BPCI on the Percent of Beneficiaries Discharged to Institutional PAC out of those who Received any PAC, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013 - Q3 2016



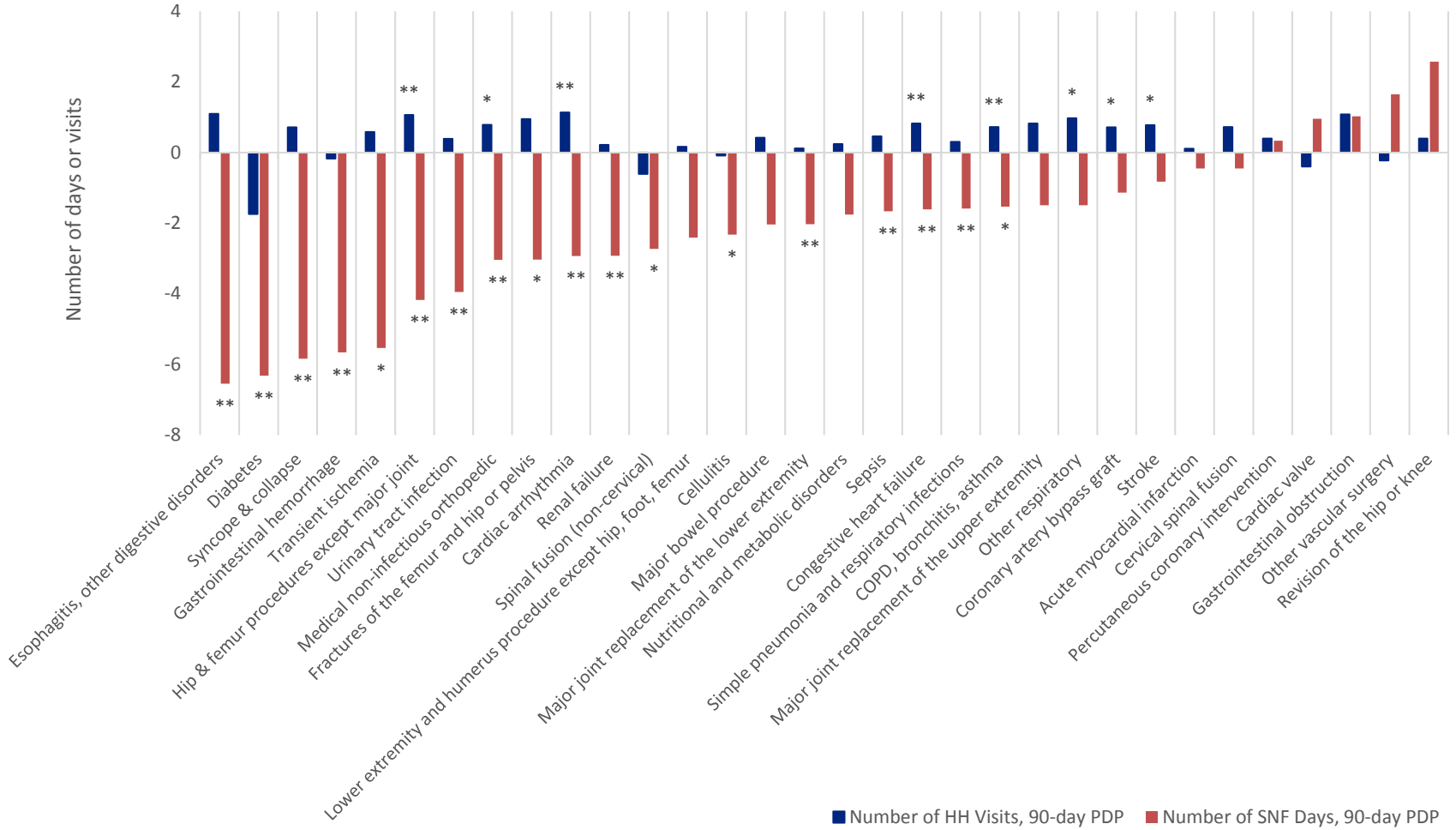
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model.

**Indicates DiD estimates are statistically significant at the 5% level.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Exhibit 16: Impact of BPCI on SNF Days and HH Visits, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013 – Q3 2016



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. These utilization measures are conditional upon the use of the service. Beneficiaries must have spent a minimum of one day in a skilled nursing facility (SNF) or home health (HH) setting during the 90 day post-discharge period (PDP) to be included in the DiD estimate for number of SNF days or HH visits, respectively. ACH= acute care hospital.

**Indicates DiD estimates are statistically significant at the 5% level.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

3. Are participants shifting services outside of the episode period or increasing services not included in the bundle, which may reduce overall savings to Medicare?

BPCI participants may attempt to reduce episode payments by changing the timing of services that would be otherwise included in the bundle so that they are furnished before the anchor hospitalization or after the end of the bundle period. Although these tactics could reduce episode payments, they would not necessarily result in Medicare savings and could even increase total Medicare spending.

Among the nine clinical episodes triggered by hospitalizations that may be planned or elective (see **Appendix J**), the anchor hospitalization could be timed to ensure that outpatient services were provided prior to the BPCI anchor stay, which could reduce within-bundle costs. However, there was no indication that this occurred in planned episodes, based on both the lack of statistically significant results and the lack of any patterns in the changes in statistically non-significant results in the 30-day pre-bundle Part B payment outcome (Exhibit 17).

There was also no pattern of increased post-episode spending across clinical episodes. However, there was a statistically significant decline ($p<0.10$) for three clinical episodes that also had statistically significant declines in the total amount included in the bundle (congestive heart failure, MJRLE, and medical non-infectious orthopedic). This suggests that BPCI participants in these episodes were not only successful in reducing payments during the episode period but during the 30-day post-bundle period as well.

Certain readmissions and services for specific conditions are excluded from the bundle; these exclusions vary by clinical episode.[20] We monitored the change in payments for excluded services because of general concerns that providers that receive bundled payments have no incentives to control services that are not included under the bundle. If the volume of such services went up due to BPCI, this could result in higher total Medicare spending. Payments for these services represent on average only 4% of total payments for the acute stay plus the 90-day post-discharge period and range from 1% (revision of the hip or knee) to 8% (gastrointestinal obstruction) of total payments. Given the small portion of payments it represents, an increase in payments for services excluded from the bundle if any, on net Medicare savings would be minimal. Payments for services not included in the bundle increased for 19 of the 32 clinical episodes. The increases were statistically significant only for fractures of the femur and hip or pelvis, simple pneumonia and respiratory infections, and stroke ($p<0.10$). However, the payments for services not included in the bundle for simple pneumonia and stroke episodes may be biased because the underlying DiD assumption of parallel baseline trends in the treatment and comparison groups were violated. The higher payments for services not included in the bundle for fractures of the femur and hip or pelvis were due to higher payments for excluded readmissions, not a higher probability of having an excluded readmission. This means that for BPCI episodes, excluded readmissions were classified into higher weighted MS-DRGs than for comparison group episodes. It is not clear whether this could be attributed to the BPCI intervention.

[20] Centers for Medicare & Medicaid Services. Bundled Payments for Care Improvement Learning & Resources Area. Retrieved from https://innovation.cms.gov/initiatives/Bundled-Payments/learning-area.html

Exhibit 17: The Impact of BPCI on Allowed Payments Outside of the Bundle, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013-Q3 2016

Clinical Episode	Number of Episodes Q4 2013-Q3 2016	Total Amount not Included in Bundle Definition[1]	Part B, 30-day Pre-bundle Period	Total Part A & B 30-day Post-bundle
Acute myocardial infarction	5,337	-$5	$52	-$75
Cardiac arrhythmia	6,029	-$41	-$55	$7
Cardiac valve	3,957	$58	$127	-$228
Cellulitis	5,474	-$60	**$122**	-$198
Cervical spinal fusion	1,190	$152	-$129	-$268
Congestive heart failure	31,858	$24	$34	-$186
COPD, bronchitis, asthma	18,331	$53	$3	-$16
Coronary artery bypass graft	3,242	$44	$94	$126
Diabetes	1,423	-$264	**-$309**	-$240
Esophagitis, gastroenteritis & other digestive disorders	4,104	-$48	-$8	-$142
Fractures of the femur and hip or pelvis	1,092	**$360**	$45	$356
Gastrointestinal hemorrhage	4,386	$62	-$44	$197
Gastrointestinal obstruction	1,735	$16	-$62	-$227
Hip & femur procedures except major joint	7,446	-$11	-$7	-$166
Lower extremity and humerus procedure except hip, foot, femur	1,089	$69	$137	$66
Major bowel procedure	3,029	-$19	**-$163**	-$13
Major joint replacement of the lower extremity	97,922	-$2	$14	-$60
Major joint replacement of the upper extremity	1,337	$160	-$14	$151
Medical non-infectious orthopedic	6,588	-$52	-$66	**-$431**
Nutritional and metabolic disorders	2,727	$110	-$9	**$800**
Other respiratory	4,700	-$151	-$16	$14
Other vascular surgery	1,590	$74	$13	-$164
Percutaneous coronary intervention	4,745	$143	$7	$28
Renal failure	7,474	$159	-$30	-$176
Revision of the hip or knee	1,146	$35	-$3	$104

Clinical Episode	Number of Episodes Q4 2013-Q3 2016	Total Amount not Included in Bundle Definition[1]	Part B, 30-day Pre-bundle Period	Total Part A & B 30-day Post-bundle
Sepsis	26,046	$40	$5	-$15
Simple pneumonia & respiratory infections	22,556	**$116***	$19	-$46
Spinal fusion (non-cervical)	3,417	-$90	-$24	$66
Stroke	11,357	$96*	-$41	$15
Syncope & collapse	1,364	-$133	$79	$212
Transient ischemia	1,099	-$89	-$45	-$351
Urinary tract infection	8,010	$15	-$67	-$244

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented, either due to insufficient sample size or the type of episodes initiated during the time period.

[1] The total amount not included in bundle definition values include 90-day episodes only.

*This might be a biased estimate because we rejected the null hypothesis that BPCI and matched comparison providers had parallel trends for this outcome (with 90% confidence), which is required for an unbiased estimate.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

4. How has quality of care changed under BPCI?

Claim-based results suggest that the quality of care generally was maintained under BPCI Model 2 (Exhibit 18). BPCI-participating providers had a relative decline in mortality that was statistically significant in three clinical episodes ($p<0.05$). There was a statistically significant relative reduction in emergency department use for congestive heart failure ($p=0.09$). Three clinical episodes exhibited statistically significant relative declines in readmission rates ($p<0.10$). Percutaneous coronary intervention clinical episodes exhibited a statistically significant increase ($p=0.03$), although BPCI and comparison provider readmission rates for percutaneous coronary intervention clinical episodes were not on parallel trends during the baseline period, so this result may be biased.[21]

Exhibit 18: Impact of BPCI on Claim-based Quality Outcomes, by Clinical Episode, Model 2 ACH, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Q4 2013 - Q3 2016	All-cause Mortality Rate, 90-day PDP (pp)	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)
Acute myocardial infarction	5,337	-1.3	-0.7	0.4
Cardiac arrhythmia	6,029	0.6	0.0	1.4
Cardiac valve	3,957	-1.0	-0.6	-2.2*
Cellulitis	5,474	0.7	-1.4	0.4
Cervical spinal fusion	1,190	0.1	-1.1	-1.8
Congestive heart failure	31,858	0.1	-0.9	0.2
COPD, bronchitis, asthma	18,331	0.0	-0.3	0.5
Coronary artery bypass graft	3,242	0.7	1.7	0.0
Diabetes	1,423	0.0	2.3*	-4.2
Esophagitis, gastroenteritis & other digestive disorders	4,104	-0.3	0.0	-0.7
Fractures of the femur and hip or pelvis	1,092	2.1	0.6	4.2
Gastrointestinal hemorrhage	4,386	0.1	-1.0	0.5
Gastrointestinal obstruction	1,735	-3.2	1.4	3.1
Hip & femur procedures except major joint	7,446	-1.0	1.3	-0.3
Lower extremity and humerus procedure except hip, foot, femur	1,089	-1.7	-0.6*	1.2
Major bowel procedure	3,029	-0.2	0.6	2.1
Major joint replacement of the lower extremity	97,922	-0.1	0.0	-0.5
Major joint replacement of the upper extremity	1,337		-2.9	-0.3
Medical non-infectious orthopedic	6,588	-1.4	0.4	-0.3
Nutritional and metabolic disorders	2,727	-1.1	-1.0	-0.1
Other respiratory	4,700	-1.4	-1.1	-0.1
Other vascular surgery	1,590	-0.9	-0.1	1.9

[21] Because of the importance of investigating potential declines in quality of care under BPCI, we conducted further analysis on the 90-day readmission rate for PCI episodes. A summary of the analyses and results are included in **Appendix I**.

Clinical Episode	Number of Episodes Q4 2013 - Q3 2016	All-cause Mortality Rate, 90-day PDP (pp)	ED Use, 90-day PDP (pp)	Unplanned Readmission Rate, 90-day PDP (pp)
Percutaneous coronary intervention	4,745	0.0	0.8	**2.5***
Renal failure	7,474	0.0	-0.6	0.6
Revision of the hip or knee	1,146	0.1	1.8	2.5
Sepsis	26,046	-0.4	-0.3	**-1.5**
Simple pneumonia & respiratory infections	22,556	0.1	0.3	0.0
Spinal fusion (non-cervical)	3,417	-0.4	1.6	-1.5
Stroke	11,357	0.1	-0.1	-0.7
Syncope & collapse	1,364	**-2.3**	-3.1	0.2
Transient ischemia	1,099	-1.0	2.1	-1.7
Urinary tract infection	8,010	-0.3*	0.7	0.4

Note: The estimates in this exhibit are the results of a difference-in-differences model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented due to insufficient sample size. ED=emergency department. PDP = post-discharge period. pp = percentage points.

* Data from the baseline period shows BPCI and matched comparison providers 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 episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Survey results indicate that BPCI respondents reported similar changes in self-reported functional status as comparison respondents (Exhibit 19). This is particularly notable given the reduction in PAC use summarized above in Exhibit 14. Survey respondents with BPCI episodes were 2.3% less likely to report decreased use of a mobility device (p=0.08) and 3.4% more likely to report increased use of a mobility device (p=0.04) relative to the comparison group. However, there were no statistically significant differences in rates of improvement or decline for any of the other six measures of changes in self-reported functional status. Estimated differences in the other outcomes were small and did not follow any consistent pattern.

The proportion of respondents with favorable care experiences was slightly smaller for BPCI survey respondents than comparison respondents. BPCI survey respondents were 1.9% less likely than comparison respondents to report that they never received conflicting medical advice (p=0.04); 2.3% less likely to report that their level of care was always appropriate to their needs (p=0.06); 2.0% less likely to agree that they were discharged at the right time (p<0.01); 1.2% less likely to agree that they had a good understanding of how to take care of themselves before going home (p<0.01); 0.7% less likely to agree that medical staff clearly explained how to take their medications before they went home (p=0.06); and 0.7% less likely to agree that medical staff clearly explained their follow-up appointments or treatments before they went home (p=0.08). BPCI survey respondents were also 2.6% less likely than comparison respondents to report that they were "extremely" or "quite a bit" satisfied with their overall recovery since leaving the hospital (p<0.01).

Although BPCI respondents reported slightly worse care experience and overall satisfaction with recovery than comparison respondents, differences were small and were not accompanied by

worse functional status outcomes. Because survey responses were not collected prior to BPCI, we cannot definitely attribute these differences to BPCI. We match respondents based on provider and episode characteristics, however, we cannot be sure that the differences between BPCI and comparison beneficiaries were not present prior to the BPCI initiative. Additional detail for these results can be found in **Appendix H.**

Exhibit 19: Differences in Survey-based Quality Outcomes between BPCI and Comparison Respondents, Model 2 ACH, Q2 2015 – Q4 2016

Domain	Outcome Measure	Difference (pp)	Percentage Difference Relative to Comparison Group	p-value
Changes in Functional Status	Improvement in bathing, dressing, using toilet, or eating	0.1	0.2	0.84
	Decline in bathing, dressing, using toilet, or eating	-0.55	-3.4	0.24
	Improvement in planning regular tasks	-0.9	-1.4	0.17
	Decline in planning regular tasks	0.6	2.6	0.24
	Improvement in use of a mobility device (less likely to use)	-1.2	-2.3	0.08
	Decline in use of a mobility device (more likely to use)	**1.2**	**3.4**	**0.04**
	Improvement in walking without rest	0.8	1.8	0.26
	Decline in walking without rest	-0.1	-0.4	0.83
	Improvement in using stairs	0.5	1.1	0.53
	Decline in using stairs	0.5	1.6	0.40
	Physical/emotional problems limit social activities less frequently	0.0	0.0	0.99
	Physical/emotional problems limit social activities more frequently	0.1	0.7	0.79
	Pain limits regular activities less frequently	0.7	1.1	0.33
	Pain limits regular activities more frequently	-0.5	-2.7	0.30
Overall Health	Depression	0.1	0.4	0.90
	Overall physical health	0.4	0.7	0.50
	Overall mental health	0.6	0.8	0.21

Domain	Outcome Measure	Difference (pp)	Percentage Difference Relative to Comparison Group	p-value
Care Experience	Never received conflicting medical advice	-1.4	-1.9	0.04
	Services always appropriate for level of care patient needed	-1.5	-2.3	0.06
	Medical staff always spoke in patient's preferred language	-0.3	-0.3	0.55
	Agree that patient was discharged at the right time	-1.8	-2.0	<0.01
	Agree that medical staff took patient's preferences into account in deciding post-discharge health care services	-0.6	-0.6	0.13
	Agree that patient had good understanding of how to take care of self before going home	-1.1	-1.2	<0.01
	Agree that medical staff clearly explained how to take medications before going home	-0.7	-0.7	0.06
	Agree that medical staff clearly explained what follow-up appointments would be needed before patient went home	-0.6	-0.7	0.08
	Agree that patient had been able to manage health needs since returning home	0.1	0.1	0.71
Overall Satisfaction	Extremely or quite a bit satisfied with overall satisfaction with recovery since leaving the hospital	-1.9	-2.6	<0.01

Note: The estimates in this exhibit are the results of a cross sectional model. Positive estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. pp = percentage points.

Source: Lewin analysis of beneficiary survey data for episodes that began in May 2015 through November 2016 for BPCI and comparison hospitals.

C. Change in Patient Mix

Episode payments for BPCI participants could decline relative to the target price if their mix of patients during the intervention period is less resource intensive than their mix of patients during the baseline. Similarly, their episode payments could increase relative to the target price if their patient mix was more resource intensive in the intervention than in the baseline. This is because the target price was not risk adjusted. We examined claim-based patient characteristics that are associated with higher resource use to address the question of whether BPCI participant patient mix changed during the intervention. Exhibit 20 shows estimates of the change between the baseline and intervention period for BPCI patients relative to the change in the comparison group of patients for demographic characteristics, count of HCCs, and the utilization of care in the six months prior to the anchor hospitalization. For each of the measures in Exhibit 20, a negative value indicates a decline in the resource intensity of the BPCI patients during the intervention from the baseline period relative to the comparison group. Similarly, a positive value suggests a relative increase in patient resource intensity. (Please note: the impact analysis on payment, utilization, and quality presented in Section III.B above controlled for changes in these patient characteristics.)

To identify whether there were changes in patient mix across strata, we categorized the 34 Model 2 strata (including MJRLE episodes stratified into fractures and non-fractures and CABG

episodes stratified into emergent and non-emergent), into three broad groups: decline in patient resource intensity, increase in patient resource intensity, and no change.[22] Our categorization was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates (See **Appendix G** for more details.)

Of the 34 Model 2 strata, the majority (29) did not have a consistent pattern of changes across measures. Four strata had indications that BPCI patients were less resource intensive in the intervention period relative to the change for the comparison group. Two of these strata were planned, elective procedures (non-fracture MJRLE and spinal fusion (non-cervical)), and participants may have had the opportunity to choose who they treated. One strata, CABG emergent, had indications that the BPCI-participating hospitals may have had a more resource intensive patient mix in the intervention period relative to the baseline period.

[22] The "no change" category includes strata that do not exhibit a consistent pattern toward a decline or an increase in patient resource intensity. This could be because they have indications of both decreases and increases in patient resource intensity or no indications of changes in either direction.

Exhibit 20: Relative Change in Patient Resource Intensity, by Clinical Episode Strata, Baseline to Intervention, Model 2, Q4 2013 – Q3 2016

Change in Patient Mix	Clinical Episode Strata	Number of episodes, Q4 2013 - Q3 2016	Medicaid Eligibility	Disabled, no ESRD	Age: 80+ years	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
Less resource intensive	Diabetes	1,423	-3.9	-4.0	0.1	0.02	-2.1	**-8.8**	0.7	-0.8
	Major joint replacement of lower extremity – Non-fracture	85,078	-0.3	**-0.9**	0.2	0.00	-0.5	-0.4	0.2	**-0.5**
	Renal failure	7,474	**-3.0**	-1.5	1.3	-0.01	0.6	-1.3	-2.0	1.4
	Spinal fusion (non-cervical)	3,417	**-3.0**	-2.7	0.5	0.02	-0.9	**-3.7**	-0.9	-0.6
More resource intensive	Coronary artery bypass graft – Emergent	1,730	-0.2	-0.4	**6.9**	0.17	0.4	3.0	1.0	**1.8**
No consistent pattern	Acute myocardial infarction	5,337	-1.0	-0.4	1.8	0.02	-1.1	-0.7	-0.7	-0.7
	Cardiac arrhythmia	6,029	0.7	-0.3	0.0	0.08	1.9	1.2	1.9	0.8
	Cardiac valve	3,957	-1.0	-1.3	0.2	-0.08	-2.4	-0.1	**-2.6**	-0.1
	Cellulitis	5,474	1.8	1.7	-0.4	0.03	**3.8**	1.1	**2.9**	1.3
	Cervical spinal fusion	1,190	-4.0	-5.1	1.4	0.08	0.7	-1.7	-1.9	1.5
	Congestive heart failure	31,858	-0.1	-0.1	-0.2	0.02	-0.3	-1.0	1.0	0.1
	COPD, bronchitis, asthma	18,331	0.6	0.7	**-2.0**	0.03	0.8	1.3	**2.0**	0.3
	Coronary artery bypass graft – Non-emergent	1,512	-0.4	-0.9	-1.1	-0.01	-1.8	-2.0	-1.1	-1.0
	Esophagitis, gastroenteritis and other digestive disorders	4,104	0.4	-1.9	0.6	0.05	0.8	0.3	1.7	0.9
	Fractures of the femur and hip or pelvis	1,092	-3.4	0.9	1.6	-0.02	0.8	1.9	0.1	3.2
	Gastrointestinal hemorrhage	4,386	0.8	0.2	-1.6	0.07	2.5	-0.3	1.3	0.8
	Gastrointestinal obstruction	1,735	-0.9	-0.2	-0.5	-0.04	-1.3	-0.9	0.8	-0.9
	Hip & femur except major joint	7,446	1.3	-0.1	-0.6	**0.13**	1.3	-0.8	1.1	0.2

Change in Patient Mix	Clinical Episode Strata	Number of episodes, Q4 2013 - Q3 2016	Medicaid Eligibility	Disabled, no ESRD	Age: 80+ years	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
No consistent pattern (cont'd)	Lower extremity and humerus procedure except hip, foot, femur	1,089	-0.2	-1.3	4.6	0.19	3.0	-2.4	0.2	1.0
	Major bowel procedure	3,029	0.7	-0.7	-1.2	-0.03	0.9	-0.2	0.1	0.2
	Major joint replacement of the upper extremity	1,337	-8.4	-2.6	5.0	0.00	1.3	-2.2	-1.4	-0.7
	Major joint replacement of lower extremity – Fracture	12,844	-0.7	-0.1	-0.3	0.03	-0.2	0.3	-0.3	1.1
	Medical non-infectious orthopedic	6,588	3.4	2.2	-0.3	0.03	1.4	-1.0	-0.3	1.1
	Nutritional and metabolic disorders	2,727	3.0	0.3	-0.3	0.03	1.8	0.0	4.6	2.2
	Other respiratory	4,700	-1.9	1.6	-2.5	0.08	2.2	1.4	0.5	-2.8
	Other vascular surgery	1,590	2.0	-1.8	4.9	-0.20	-2.4	1.2	-0.9	-1.0
	Percutaneous coronary intervention	4,745	1.5	1.3	-0.9	0.01	-1.2	-0.6	-0.5	-0.3
	Revision of hip or knee	1,146	-2.6	-2.7	2.3	0.02	0.6	-2.3	0.5	2.1
	Sepsis	26,046	-1.5	-0.1	0.0	0.08	0.8	0.0	1.1	-0.1
	Simple pneumonia, respiratory infections	22,556	-0.5	0.4	-0.6	0.05	0.3	0.0	0.0	-0.2
	Stroke	11,357	0.1	-1.4	-0.5	0.05	0.8	0.4	0.2	0.1
	Syncope & collapse	1,364	-2.7	-0.9	0.4	-0.04	2.0	-0.5	3.7	-0.4
	Transient ischemia	1,099	-0.7	0.6	-4.7	0.00	-2.7	-1.7	0.1	1.6
	Urinary tract infection	8,010	2.5	0.1	-2.2	0.09	0.5	1.4	1.1	-1.1

Note: Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively.

* These characteristics measure utilization of care in the six months prior to the anchor hospitalization. Count of HCCs is based on the six months prior to the anchor hospitalization.

Source: Lewin analysis of Medicare claims, enrollment, and Minimum Data Set (MDS) data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q3 2016 (intervention period) for BPCI EIs and the matched comparison providers.

IV. Model 3 Impact of BPCI

The impact of BPCI Model 3 was estimated separately for SNF EIs and HHA EIs for the first 12 quarters of the BPCI initiative (Q4 2013 through Q3 2016). There was insufficient sample size to conduct analyses for any clinical episodes for IRF or LTCH EIs.[23] See Section II.B for additional details on the statistical approach. All results for Model 3 clinical episodes are in **Appendix K** and **Appendix L.**

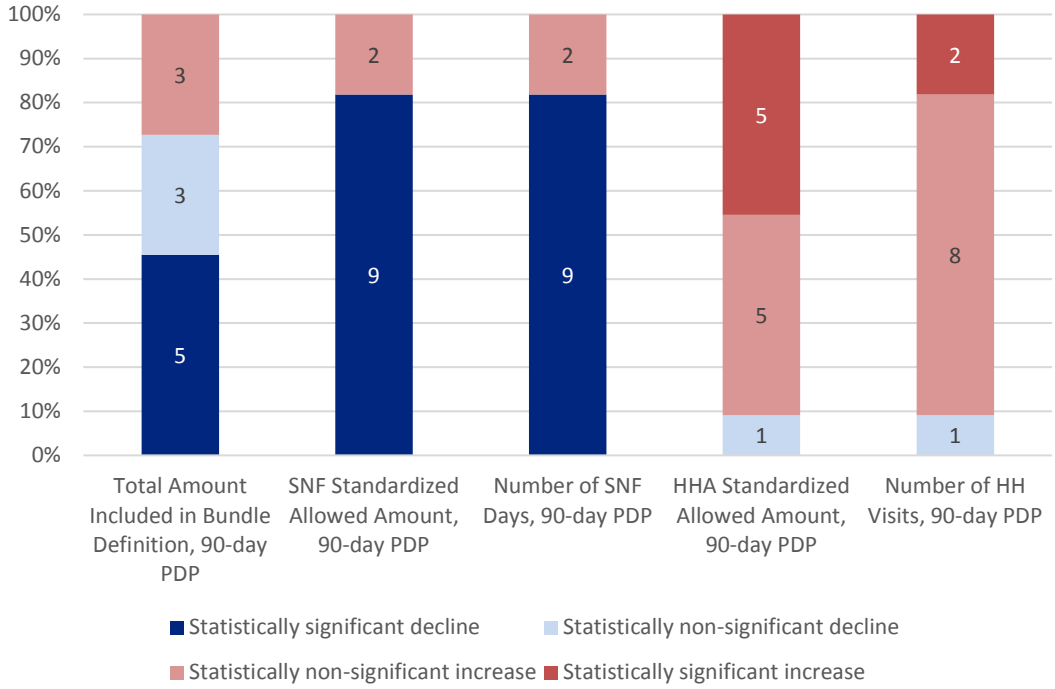
We observed declines in the total amount included in the bundle definition for 90-day episodes in eight of the 11 clinical episodes relative to the change in the comparison group. The reductions in total allowed payments were statistically significant for hip and femur, MJRLE, sepsis, simple pneumonia and respiratory infection, and urinary tract infection clinical episodes. The reduced payments were due to a shorter SNF length of stay; nine of the 11 clinical episodes had statistically significant declines relative to the change in the comparison group. The reduction in SNF use was partially offset by an increase in HHA utilization. HHA payments increased for 10 of the 11 clinical episodes; for five of the clinical episodes, the increase was statistically significant (Exhibit 21).

There were no indications of improved quality of care. There were a few indications of a relative decline in quality of care under Model 3 for SNF-initiated episodes, although further analysis suggests the statistical significance of these estimates are not robust.

For the three HHA episodes with sufficient sample size, there were no statistically significant reductions in total allowed payments for 90-day episodes and few statistically significant relative changes in quality.

[23] We did not analyze PGP EI clinical episodes because of inadequate data for assigning physicians to participating PGPs. PGP results will be included in the next annual report.

Exhibit 21: Clinical Episodes with a Relative Change in Payment and Utilization Outcomes, Baseline to Intervention, Model 3 SNF EIs, Q4 2013- Q3 2016



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark red shading. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark blue shading. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. The total amount included in the bundle definition is based on only the 90-day episodes. These payment measures are not conditional upon use of the service, whereas these utilization measures are conditional upon the use of the service. SNF = skilled nursing facility. EI = episode initiator. PDP = post-discharge period. HHA = home health agency. HH = home health.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

A. BPCI Participating SNFs

1. Sample Characteristics

Before discussing the impact of BPCI on payments, utilization, and quality, we present some basic statistics by clinical episode to better understand the sample included in the SNF analysis (Exhibit 22). Among SNF EIs, the number of matched EIs ranged from 78 to 236 for a given clinical episode; these EIs initiated between 676 and 5,711 episodes over the first twelve quarters of the initiative. Because providers were allowed to join BPCI over an extended period, these data represent an average of five quarters of participation among SNF EIs. Providers are also allowed to stop participation in BPCI for one or more episodes on a quarterly basis or withdraw from the initiative completely at any time. During the first three years (Q4 2013 through Q3 2016), 231 of 873 SNF EIs (26%) withdrew entirely from the initiative. Among the 493 matched EIs participating in any of the 11 clinical episodes analyzed in this report, approximately 21% stopped participating in at least one clinical episode during the first three years. Across the 11 Model 3 SNF clinical episodes, 36% of the intervention episodes were initiated by SNFs that terminated their participation in that clinical episode by the beginning of the fourth quarter of 2016. Across clinical episodes there was a large difference in the percentage of EIs that discontinued their participation

within the initiative, with as few as 6% of EIs in hip and femur procedures except major joint and as many as 40% of EIs participating in COPD.

Exhibit 22: Characteristics of the Matched BPCI Providers Included in the DiD Estimates, Model 3 SNF, Q4 2013 – Q3 2016

Clinical Episodes	Matched EIs (#)	Matched Intervention Period Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (%)	Intervention Period Episodes from EIs that Terminated (%)
Congestive heart failure	181	2,649	6	29%	46%
COPD, bronchitis, asthma	98	826	6	40%	63%
Hip & femur procedures except major joint	119	2,019	5	6%	5%
Major joint replacement of the lower extremity	218	5,711	5	8%	8%
Medical non-infectious orthopedic	125	2,016	5	33%	54%
Other respiratory	78	676	6	38%	64%
Renal failure	97	1,323	6	38%	69%
Sepsis	193	3,863	5	25%	43%
Simple pneumonia and respiratory infections	236	2,569	5	22%	38%
Stroke	98	1,215	5	34%	52%
Urinary tract infection	153	1,813	6	32%	50%

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q3 2016 for BPCI providers. This exhibit is limited to the matched BPCI providers used to calculate the difference-in-differences (DiD) results in the remainder of this section. SNF = skilled nursing facility. EI = episode initiator.

2. Key Payment, Utilization, and Quality Outcomes

a. How have the average standardized allowed amounts (Medicare payments and coinsurance/copayments combined) changed under BPCI?

Across SNF clinical episodes, generally total and SNF allowed amounts declined, and HHA and readmissions payments increased. The total amount included in the bundle definition for 90-day episodes decreased in eight of the 11 episodes relative to the change in the comparison group; the decline was statistically significant for five of the clinical episodes ($p<0.05$) (Exhibit 23). The average reduction in Medicare payments across these five episodes was 7.8% of the payments we would have expected without BPCI (Exhibit 24). Reduced SNF payments contributed to the decline in total payments, which were partially offset by increased HHA payments. The SNF standardized allowed amount went down ($p<0.05$) in all but two clinical episodes relative to the change in the comparison group. The HHA standardized allowed amount increased in 10 of 11 clinical episodes relative to the change in the comparison group; the increase was statistically significant for five clinical episodes ($p<0.10$).

The standardized allowed amount for readmissions increased in eight of 11 episodes; the increase was statistically significant for three clinical episodes ($p<0.10$). The higher readmission payments could be related to shorter SNF LOS. We will continue to monitor this and other quality measures under Model 3.

Exhibit 23: Impact of BPCI on Medicare Allowed Payment Outcomes, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Initiated Q4 2013 – Q3 2016	Total Amount Included in Bundle Definition[1]	SNF Standardized Allowed Amount, 90-day PDP	HHA Standardized Allowed Amount, 90-day PDP[2]	Readmissions Standardized Allowed Amount, 90-day PDP[2]
Congestive heart failure	2,649	-$1,285	**-$1,526**	**$230**	$236
COPD, bronchitis, asthma	826	$1,895	$557	-$23	$918
Hip & femur procedures except major joint	2,019	**-$2,799**	**-$3,753**	**$259**	**$658**
Major joint replacement of the lower extremity	5,711	**-$1,388**	**-$1,867**	**$387**	-$252
Medical non-infectious orthopedic	2,016	-$1,171	**-$1,650**	$65	-$25
Other respiratory	676	$1,304	**-$2,322**	$27	$1,745
Renal failure	1,323	$335	$73	$174	$445
Sepsis	3,863	**-$2,301**	**-$1,892**	**$246**	-$412
Simple pneumonia and respiratory infections	2,569	**-$2,334**	**-$1,895**	$83	$213
Stroke	1,215	-$38	**-$1,810**	$242	**$1,415**
Urinary tract infection	1,813	**-$2,241**	**-$2,208**	$106	$287

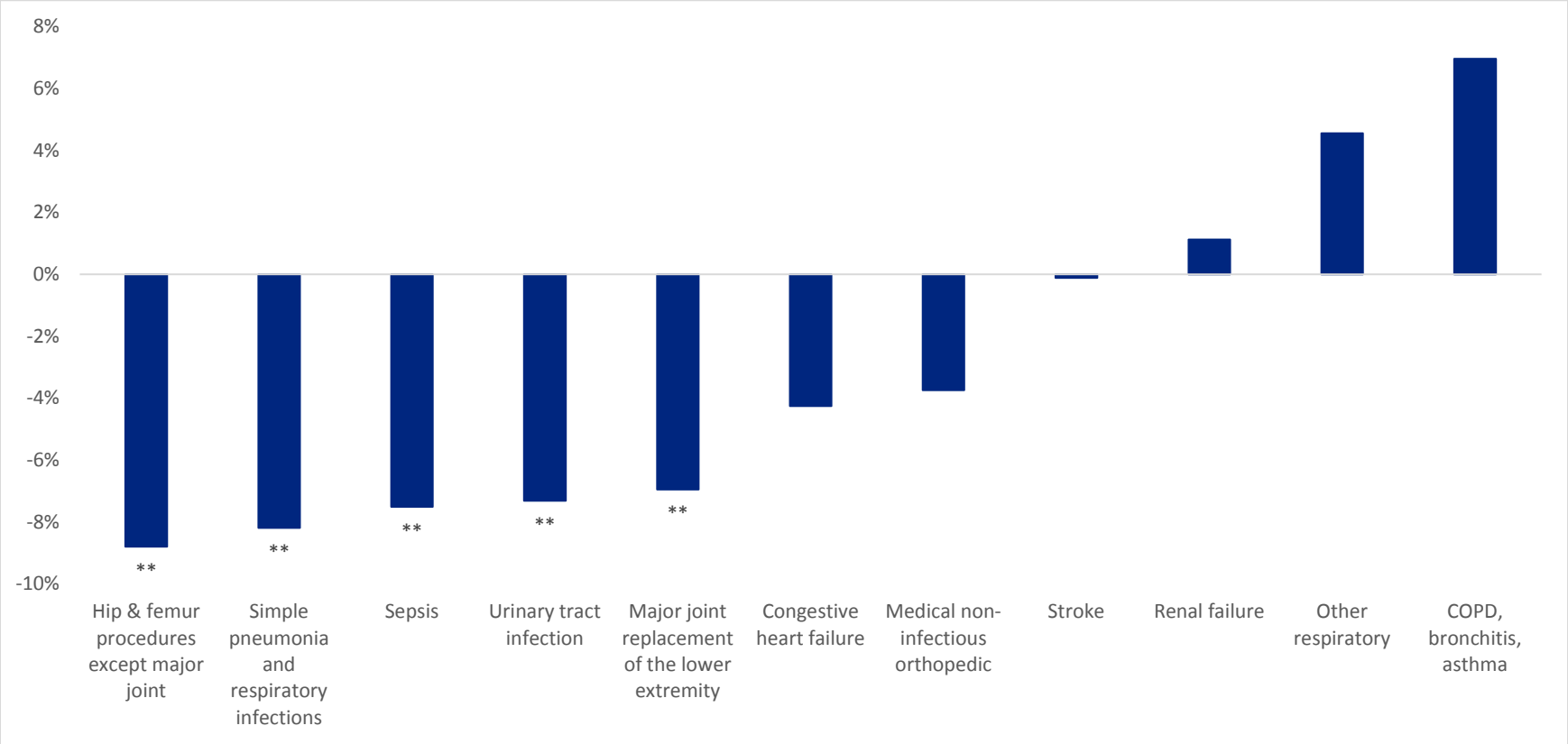
Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. SNF = skilled nursing facility. PDP = post-discharge period. HHA = home health agency.

[1] The total amount included in bundle definition values are based on only the 90-day episodes.

[2] These payments are not conditional upon use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Exhibit 24: Percent Change in BPCI Episode Payments from What Payments Would have Been Absent BPCI, by Clinical Episode, Model 3 SNF, Q4 2013 - Q3 2016



Note: The payments in this Exhibit are the risk-adjusted standardized allowed amounts for services included in the 90 day clinical episodes. 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). Results are sorted by the total amount included in the (90-day) bundle definition DiD estimate. SNF = skilled nursing facility.

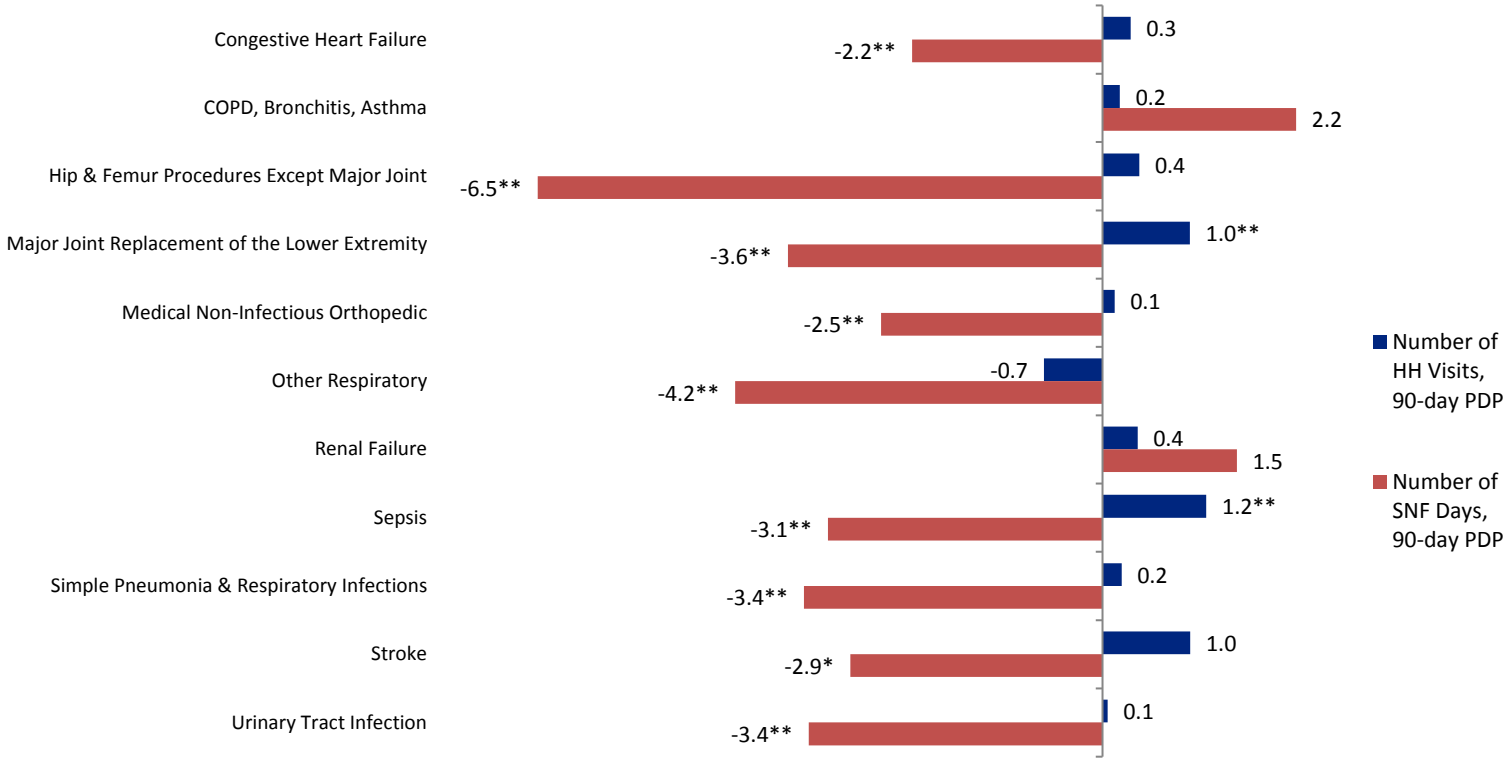
**Indicates DiD estimates are statistically significant at the 5% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

b. How have the services changed under BPCI?

Among the Model 3 SNF clinical episodes we examined, the number of SNF days had a statistically significant decline ($p<0.10$) in nine episodes relative to the change in the comparison group (Exhibit 25). This reduction in SNF days aligns with the reduction in SNF payments. HH visits increased relative to the change for the comparison group in 10 of the 11 clinical episodes. The two clinical episodes that had statistically significant increases in HH visits ($p<0.10$), MJRLE and sepsis, also had a statistically significant decline in the number of SNF days ($p<0.05$), suggesting that HHA visits were substituted for SNF days.

Exhibit 25: Impact of BPCI on Utilization Outcomes, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 – Q3 2016



Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. These utilization measures are conditional upon the use of the service. Beneficiaries must have spent a minimum of one day in a skilled nursing facility (SNF) or home health (HH) agency during the 90 day post-discharge period (PDP) to be included in the DiD estimate for number of SNF days or HH visits, respectively.

**Indicates DiD estimates are statistically significant at the 5% level.

*Indicates DiD estimates are statistically significant at the 10% level.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

c. *Are participants shifting services outside of the episode period or increasing services not included in the bundle, which may reduce overall savings to Medicare?*

BPCI participants may attempt to reduce episode payments by changing the timing of services that would be otherwise included in the bundle so that they are furnished before the initiating SNF stay or after the end of the bundle period. Although these tactics could reduce episode payments, they would not necessarily achieve Medicare savings. If this were happening, payments prior to the episode or after the episode would increase. However, for Model 3 SNF episodes, payments in the 30-day pre-bundle period decreased in seven of 11 clinical episodes; the decline was statistically significant for MJRLE ($p=0.07$). Post-bundle payments decreased in six clinical episodes; again, MJRLE was the only clinical episode with a statistically significant decrease ($p=0.10$) (Exhibit 26). One clinical episode, urinary tract infection, had a statistically significant increase of $581 in post-bundle payments for BPCI relative to the comparison group from baseline to intervention ($p=0.09$).

We monitored the change in payments for excluded services because of general concerns that providers that receive bundled payments have no incentives to control services that are not included under the bundle, which could result in higher total Medicare spending. Payments for these services on average are only 3% of total payments in the bundle period and range from 1% (hip and femur procedures except major joint) to 8% (renal failure), so an increase in payments for excluded services, if any, would be small. The standardized allowed amount for services not included in the bundle for 90-day episodes increased in seven of 11 clinical episodes relative to the change in the comparison group (Exhibit 26). For two of these clinical episodes, COPD and stroke, the increase was statistically significant ($p<0.10$) and the higher payments were due to excluded readmissions. For COPD, there was a higher probability of an excluded readmission for BPCI episodes relative to the comparison group (3.0 percentage points, $p=0.05$). For stroke episodes, the increase in standardized allowed payment for excluded services was likely due to excluded readmissions in higher weighted MS-DRGs for BPCI episodes relative to comparison episodes. It is not clear whether this could be attributed to the BPCI intervention.

Exhibit 26: The Impact of BPCI on Allowed Payment Outcomes Outside of the Bundle, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Initiated Q4 2013 – Q3 2016	Standardized Allowed Amount not Included in Bundle Definition[1]	Standardized Allowed Amount Part A & B, Days 1-30 Pre-bundle Period	Standardized Allowed Amount Part A & B, Days 1-30 Post-bundle Period
Congestive heart failure	2,649	$205	-$779	-$50
COPD, bronchitis, asthma	826	**$671**	-$552	$120
Hip & femur procedures except major joint	2,019	$173	$537	$181
Major joint replacement of the lower extremity	5,711	-$28	-$334	-$246
Medical non-infectious orthopedic	2,016	$166	-$199	-$161
Other respiratory	676	-$149	$901	-$786
Renal failure	1,323	-$338	-$542	$404
Sepsis	3,863	-$183	$416	-$347
Simple pneumonia and respiratory infections	2,569	$179	-$558	-$479
Stroke	1,215	$335	$876	$288
Urinary tract infection	1,813	$157	-$495	$581

Note: The estimates in this exhibit are the result of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark light orange shaded cells, respectively. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments. A blank cell indicates that the outcome cannot be presented, either due to insufficient sample size or the type of episodes initiated during the time period. SNF = skilled nursing facility.

[1] The standardized allowed amount not included in bundle definition is based on only 90-day episodes.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

d. How has quality of care changed under BPCI?

Across the claim-based outcomes, there was some evidence that quality may have declined due to BPCI (Exhibit 27). There were statistically significant increases in the unplanned readmission rate relative to the change in the comparison group in hip and femur procedures except major joint (3.6 percentage points, p=0.06) and stroke (6.9 percentage points, p=0.01) clinical episodes. Mortality increased in the COPD (5.7 percentage points, p=0.03) clinical episode. Relative increases in emergency department use were statistically significant in MJRLE (2.0 percentage points, p=0.06) and stroke (5.3 percentage points, p=0.03) clinical episodes. These results could signal inadequate care during the episode. Emergency department use for MJRLE episodes rose at the same time that Medicare payments and SNF use declined. That unplanned readmissions and emergency department use were both statistically significantly higher for BPCI stroke episodes may be of particular concern, but mortality declined (although this change was not statistically significant).

Exhibit 27: Impact of BPCI on Claim-based Quality Outcomes, by Clinical Episode, Model 3 SNF, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Initiated Q4 2013 – Q3 2016	Unplanned Readmission Rate, 90 Days from Episode Start Date (pp)	Emergency Department Use, 90 Days from Episode Start Date (pp)	All-cause Mortality Rate, 90 Days from Episode Start Date (pp)
Congestive heart failure	2,649	0.8	0.8	-0.2
COPD, bronchitis, asthma	826	4.7	4.2	**5.7**
Hip & femur procedures except major joint	2,019	3.6	-1.4*	2.0
Major joint replacement of the lower extremity	5,711	-0.6	2.0	-0.4
Medical non-infectious orthopedic	2,016	-2.0	-1.3*	-2.1*
Other respiratory	676	3.5	4.1	-1.5*
Renal failure	1,323	2.9	1.5	-2.7
Sepsis	3,863	-0.5	1.8	0.5*
Simple pneumonia and respiratory infections	2,569	0.1	-0.4	0.3
Stroke	1,215	**6.9**	**5.3**	-2.6
Urinary tract infection	1,813	1.5	0.1	0.9*

Note: The estimates in this exhibit are the result of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. SNF = skilled nursing facility. pp = percentage points.

* Data from the baseline period shows BPCI and matched comparison providers were not on parallel trends for this outcome, which is required for an unbiased estimate. Equal trends test was conducted for ED, readmission, and mortality outcomes.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Because of the importance of quality of care, we conducted further analyses of the outcomes that suggest that quality may have declined for Model 3 SNF clinical episodes. The comparison episodes for SNFs were not as close a match as we would like, even after multiple attempts to improve the match. To test the impact of this on our results, we conducted sensitivity analyses with 1,000 random samples of episodes from comparison providers and compared the DiD results across the random samples. We found that 97% of the samples yielded results in the same direction as presented in Exhibit 27. However, the results were statistically significant in less than 50% of the random samples. None of the samples produced statistically significant results in the opposite direction. Although the statistical significance of these estimates are in question, we conclude that these providers may be unable to reduce costs and consistently maintain quality as measured with these three outcomes. The overall patterns, therefore, suggest that these episodes should be subject to continued scrutiny.

3. Changes in Patient Mix

Episode payments for BPCI participants could decline relative to their target price if their mix of patients during the intervention period is less resource intensive than their mix of patients during the baseline. This is because the target price was not risk adjusted. Relative to Model 2, under Model 3, there may be more opportunities for participants to change their mix of patients in response to BPCI because PAC providers choose who they will treat. In this section, we examine

changes in patient characteristics that are associated with higher resource use across Model 3 SNF strata (including MJRLE stratified into fractures and non-fractures) to assess if the resource intensity of the average patient changed under the initiative. We compared the change between the baseline and intervention period for BPCI patients relative to the change in the matched comparison group of patients in basic demographic characteristics, count of HCCs, utilization of care in the six months prior to the qualifying hospitalization, and diagnostic and functional information from the initial patient assessment conducted at their episode-initiating SNF. For each of the measures in Exhibits 28 and 29, a negative value indicates a decline in the resource intensity of the BPCI patients during the intervention from the baseline period relative to the comparison group. Similarly, a positive value suggests a relative increase in patient resource intensity. (Please note, the impact analysis on payment, utilization, and quality presented above controlled for changes in the claim-based patient characteristics.)

To identify whether there were changes in patient mix across strata, we categorized the 12 Model 3 strata into three broad groups: decline in patient resource intensity, increase in patient resource intensity, and no change.[24] Our categorization was based on statistically significant changes in patient characteristics associated with higher resource use as well as the direction and average magnitude of the estimates (See **Appendices G and M** for more details).

Four of the 12 Model 3 strata had indications that BPCI patients were less resource intensive in the intervention period relative to the comparison group. COPD appeared to have a shift towards a more resource intensive patient mix and seven strata had no change in resource intensity.

Among episodes triggered by a SNF stay, comparisons between the patient mix measures based on claims and assessment data demonstrate the differences in information available. For example, the claims data indicated no changes toward a less intensive patient mix, but the addition of assessment measures suggests a less intensive patient mix for four strata. This finding indicates particular challenges with risk-adjusting Model 3 target prices using only claim-based measures.

The four assessment-based functional status measures (moving in bed, transferring, walking in room, and toileting) indicate that BPCI-participating SNFs may have been treating patients who required less assistance after joining BPCI relative to the change for the comparison group. Six of the 12 strata had a decline in the proportion of patients who needed assistance in all functional status outcomes relative to the comparison group. These declines were statistically significant for three or more measures in four of the six strata ($p<0.10$). These functional measures, however, may be more subjective than measures indicating the presence of comorbidities, so these results should be interpreted with caution. Additional information on which characteristics had a statistically significant change can be found in Exhibits 28 and 29. See **Appendix M** for definitions of the MDS characteristics.

[24] The "no change" category includes strata that do not exhibit a consistent pattern toward a decline or an increase in patient resource intensity. This could be because they have indications of both decreases and increases in patient resource intensity or no indications of changes in either direction.

Exhibit 28: Relative Change in Patient Resource Intensity from Claim Measures, by Clinical Episode Strata, Baseline to Intervention, Model 3 SNF, Q4 2013 – Q3 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes, Q4 2013- Q3 2016	Medicaid Eligibility	Disabled, no ESRD	Age: 80+ years	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
Less resource intensive	Congestive heart failure	2,649	-1.8	1.8	0.2	-0.05	0.4	-1.4	4.2	-2.6
	Major joint replacement of lower extremity – Fracture	1,500	0.8	1.2	-3.0	0.02	-0.1	-0.4	1.2	-2.8
	Medical non-infectious orthopedic	2,016	2.1	0.9	-6.3	0.08	2.4	-0.4	-3.5	-1.8
	Stroke	1,215	5.7	2.4	-2.2	-0.21	-4.6	1.4	-5.0	-5.0
More resource intensive	COPD, bronchitis, asthma	826	-0.9	3.5	-2.6	0.41	4.3	4.8	-1.1	7.7
No consistent pattern	Hip & femur procedures except major joint	2,019	3.6	-0.4	-1.4	-0.01	0.8	-0.4	0.1	-0.1
	Major joint replacement of lower extremity – Non-fracture	4,211	-2.4	-1.0	-4.1	0.03	-1.1	-1.2	2.9	-1.9
	Other respiratory	676	0.0	2.1	-5.1	0.18	1.6	0.5	5.4	3.2
	Renal failure	1,323	-2.1	0.4	1.8	-0.19	2.0	-4.5	-2.1	-1.8
	Sepsis	3,863	1.2	-0.7	1.4	0.01	3.1	-0.1	2.4	-0.1
	Simple pneumonia and respiratory infections	2,569	-1.5	0.0	-0.3	0.07	1.7	-7.0	-0.9	-1.0
	Urinary tract infection	1,813	3.8	1.0	-2.7	-0.04	-0.6	0.2	-1.1	1.1

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. SNF = skilled nursing facility. ESRD = end stage renal disease. HCC= Hierarchical Condition Category. PAC = post-acute care.

* These characteristics measure utilization of care in the six months prior to the qualifying hospitalization. Count of HCCs is based on the six months prior to the qualifying hospitalization.

Source: Lewin analysis of Medicare claims, enrollment, and Minimum Data Set (MDS) data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q3 2016 (intervention period) for BPCI EIs and the matched comparison providers.

Exhibit 29: Relative Change in Patient Resource Intensity from Assessment Measures, by Clinical Episode Strata, Baseline to Intervention, Model 3 SNF, Q4 2013 – Q3 2016

Change in Patient Mix	Clinical Episode Strata	Number of Episodes, Q4 2013- Q3 2016*	Need Extensive Assistance or are Totally Dependent Moving in Bed	Need Extensive Assistance or are Totally Dependent Transferring, e.g., Between Bed and Wheelchair	Need Extensive Assistance or are Totally Dependent Walking in Room	Need Extensive Assistance or are Totally Dependent Using the Toilet	Not Currently Married	Moderate to Severe Cognitive Impairment	Moderate to Severe Depression	Rejected Necessary Evaluation or Care at Least Once	Unhealed Pressure Ulcer	Incontinence	Active Diagnosis of Alzheimer's	Active Diagnosis of Dementia	Shortness of Breath	Require Special Treatment [1]
Less resource intensive	Congestive heart failure	2,594	-3.8	-1.8	-4.5	-3.0	1.2	2.6	-0.9	-0.5	-1.6	-1.1	-0.2	2.4	-5.0	2.4
	Major joint replacement of the lower extremity – Fracture	1,445	-3.4	-4.2	-8.0	-2.9	0.6	-5.3	-2.0	-1.0	-1.3	-9.1	-3.4	-7.9	-3.2	0.1
	Medical non-infectious orthopedic	1,965	-4.6	-6.4	-5.0	-6.0	1.0	-3.5	-0.5	-0.4	-1.7	-5.9	-1.2	3.4	-6.0	-4.5
	Stroke	1,193	-2.9	-3.4	-3.9	-3.3	-2.3	-3.9	0.8	0.8	-1.5	-4.9	-0.1	-6.2	-2.0	-5.5
More resource intensive	COPD, bronchitis, asthma	812	-6.6	-6.3	-4.7	-6.0	6.5	-1.2	-3.5	0.0	-1.8	1.7	2.3	3.3	1.3	1.3
No consistent pattern	Hip & femur procedures except major joint	1,978	-0.5	0.9	-0.8	-1.5	0.9	-6.6	-1.6	-1.1	-1.6	-6.3	-2.6	-2.6	-2.0	-0.1
	Major joint replacement of the lower extremity - Non-Fracture	4,088	0.9	-1.2	-0.3	-1.7	-0.7	-0.5	-0.7	-0.6	-0.3	0.1	-0.1	-0.6	-0.8	-2.1
	Other respiratory	666	3.6	2.4	-5.3	4.0	-0.8	-2.9	-3.1	0.1	0.3	-3.2	0.5	-0.1	11.3	9.4
	Renal failure	1,297	-4.2	-2.7	0.5	-2.3	1.5	3.2	1.9	0.4	-1.1	0.6	0.0	-3.8	-1.6	-5.5
	Sepsis	3,764	-0.7	-1.2	-2.4	-0.3	3.4	0.1	0.0	1.4	-1.6	-0.7	0.9	3.3	-1.0	-0.5
	Simple pneumonia and respiratory infections	2,512	-2.3	0.3	-0.4	-0.6	-3.0	2.7	-1.4	0.0	-0.5	-1.2	0.0	3.4	-0.7	0.8
	Urinary tract infection	1,760	1.9	0.9	-1.8	0.7	2.3	5.5	1.8	-0.2	0.6	0.3	0.3	2.4	-1.7	1.2

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively). MDS = Minimum Data Set. ADL=Activities of Daily Living.

The MDS assessment is administered within 5 days (plus 3 days grace) of the start of care date. The measurement period for outcomes listed above is the seven-day "look-back" period preceding the assessment (the first week of the SNF stay).

*Assessment data was not available for all episodes. This table is limited to the episodes where we had the initial patient assessment data.

[1] Examples include chemotherapy, oxygen therapy, and transfusions.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

B. BPCI Participating HHAs

1. Sample Characteristics

The number of matched HHA EIs ranged from 37 to 46 across the three clinical episodes for which there was sufficient sample (Exhibit 30). Because providers were allowed to join BPCI over an extended period, these data represent an average of five quarters of participation. Providers are also allowed to withdraw from the initiative completely at any time. During the first three years (Q4 2013 through Q3 2016), 35 of 116 HHA EIs (30%) withdrew entirely from the initiative. Providers are also allowed to stop participation in BPCI for one or all episodes at any time, and the percent of EIs that exited varies across the three HHA clinical episodes. Among the 71 matched EIs participating in any of the three clinical episodes analyzed in this report, approximately 46% stopped participating in at least one clinical episode during the first three years. Over half of the HHA EIs stopped participating in the CHF clinical episode, and these EIs accounted for 21% of intervention episodes. For simple pneumonia and respiratory infection, terminated EIs (57%) accounted for 58% of the episodes initiated during the intervention period. Retention was higher in the MJRLE clinical episode, where only five (14%) EIs stopped participating in MJRLE through the third year of BPCI.

Exhibit 30: Characteristics of the Matched BPCI Providers included in the DiD Estimates, Model 3 HHA, Q4 2013 – Q3 2016

Clinical Episodes	Matched EIs (#)	Matched Intervention Period Episodes (#)	Average Length of Participation (Quarters)	EIs that Terminated Participation in the Clinical Episode (%)	Intervention Period Episodes from EIs that Terminated (%)
Congestive heart failure	46	4,119	5	57%	21%
Major joint replacement of lower extremity	37	2,931	5	14%	5%
Simple pneumonia and respiratory infections	37	1,208	5	57%	58%

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2013 through Q3 2016 for BPCI providers. This exhibit is limited to the matched BPCI providers used to calculate the difference-in-differences (DiD) results in the remainder of this section.

2. Key Payment, Utilization, and Quality Outcomes

a. How have the average standardized allowed amounts (Medicare payments and coinsurance/copayments combined) changed under BPCI?

Total allowed amounts in the bundle declined in two of the three clinical episodes under BPCI, but there were no statistically significant changes relative to the comparison group. The only statistically significant change in payments was a relative decline in the SNF standardized allowed amount for MJRLE episodes (p=0.02) (Exhibit 31).

Exhibit 31: Impact of BPCI on Allowed Payment Outcomes, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Initiated Q4 2013 – Q3 2016	Total Amount Included in Bundle Definition[1]	HHA Standardized Allowed Amount, 90-day PDP[2]	SNF Standardized Allowed Amount, within the bundle[1,2]	Readmissions Standardized Allowed Amount, 90-day PDP[2]
Congestive heart failure	4,119	-$656	-$147	-$71	-$306
Major joint replacement of lower extremity	2,931	-$396	-$25	-$327	-$57
Simple pneumonia and respiratory infections	1,208	$262	$196	-$290	$119

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented, either due to insufficient sample size or the type of episodes initiated during the time period. Medicare payment outcomes are standardized to remove the effect of geographic and other adjustments and are trended to 2015. PDP = post-discharge period.

[1] The total amount included in bundle definition values are based on only the 90-day episodes.

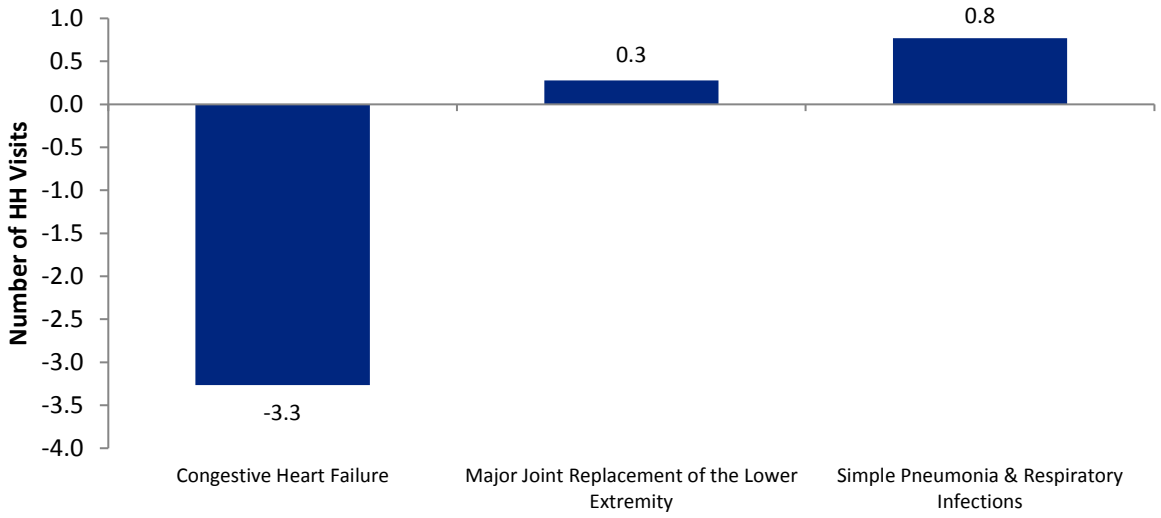
[2] These payment measures are not conditional upon use of the service.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

b. How have the services changed under BPCI?

There were no statistically significant changes in the number of HH visits relative to the change in the comparison group across the three clinical episodes (Exhibit 32). Changes in the number of HH visits does not necessarily translate into changes in HHA payments because Medicare typically makes a single payment for 60 days of HHA care.

Exhibit 32: Impact of BPCI on Home Health Utilization, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013 - Q3 2016



Note: The estimates in this table are the results of a difference-in-differences (DiD) model. HHA = home health agency.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

c. *Are participants shifting services outside of the episode period or increasing services not included in the bundle, which may reduce overall savings to Medicare?*

Across the HHA clinical episodes we examined, payments in the 30-day pre-bundle period increased in all three clinical episodes relative to the change in the comparison group, though none of the differences were statistically significant (Exhibit 33). The standardized allowed amount for services not included in the bundle for 90-day episodes decreased in all three clinical episodes. For simple pneumonia and respiratory infection episodes, the difference was statistically significant (p=0.07) (Exhibit 33).

Exhibit 33: Impact of BPCI on Allowed Payment Outcomes Outside of the Bundle, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Initiated Q4 2013 – Q3 2016	Total Amount not Included in Bundle Definition[1]	Standardized Allowed Amount Part A & B, Days 1-30 Pre-bundle Period	Standardized Allowed Amount Part A & B, Days 1-30 PBP
Congestive heart failure	4,119	-$82	$307	-$452
Major joint replacement of lower extremity	2,931	-$26	$486	$105
Simple pneumonia and respiratory infections	1,208	-$409	$788	-$113

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented, either due to insufficient sample size or the type of episodes initiated during the time period. HHA = home health agency. PBP = post-bundle period.

[1] The standardized allowed amount not included in bundle definition values include 90-day episodes only.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

d. *How has quality of care changed under BPCI?*

Across clinical episodes, there were no patterns in the change in readmissions rates, emergency department use, or mortality rates (Exhibit 34). For CHF episodes, emergency department use increased relative to the change in the comparison group (2.2 percentage points, p=0.08) and readmissions and mortality declined, although these changes were not statistically significant. Emergency department use declined for the simple pneumonia and respiratory infection clinical episodes (-4.5 percentage points, p=0.10).

Exhibit 34: Impact of BPCI on Claim-based Quality Outcomes, by Clinical Episode, Model 3 HHA, Baseline to Intervention, Q4 2013 - Q3 2016

Clinical Episode	Number of Episodes Initiated Q4 2013 – Q3 2016	Unplanned Readmission Rate, 90 Days from Episode Start Date	Emergency Department Use, 90 Days from Episode Start Date	All-cause Mortality Rate, 90 Days from Episode Start Date
Congestive heart failure	4,119	-1.6 pp	2.2 pp	-1.3* pp
Major joint replacement of lower extremity	2,931	-1.5* pp	-0.8 pp	0.1 pp
Simple pneumonia and respiratory infections	1,208	0.9* pp	-4.5 pp	-1.9 pp

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. A blank cell indicates that the outcome cannot be presented, either due to insufficient sample size. HHA = home health agency.

* Data from the baseline period shows BPCI and matched comparison providers 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 episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

3. Changes in Patient Mix

Episode payments for BPCI participants could decline relative to their target price if their mix of patients during the intervention period is less resource intensive than their mix of patients during the baseline. This is because the target price was not risk adjusted. Based on claim- and assessment-based characteristics that are associated with higher resource use, there is evidence that the resource intensity of CHF episodes declined for BPCI-participating HHAs from the baseline to the intervention period relative to the comparison group. Five of the OASIS measures indicated a less resource intensive patient mix and were statistically significant (see Exhibit 36). There was no consistent pattern among MJRLE and SPRI patients. See **Appendices G and M** for additional information.

Exhibit 35: Relative Change in Patient Resource Intensity from Claim Measures, by Clinical Episode Strata, Baseline to Intervention, Model 3 HHA, Q4 2013 – Q3 2016

Change in Patient Mix	Clinical Episode	Number of Episodes, Q4 2013- Q4 2016	Medicaid Eligibility	Disabled, no ESRD	Age: 80+ years	Count of HCC Indicators *	Inpatient Acute Care Hospital *	Emergency Room without Admission *	Home Health *	Institutional Nursing Facility *
Less resource intensive	Congestive heart failure	4,119	4.0	1.1	-2.3	-0.01	0.3	-0.5	-1.7	-0.1
No consistent pattern	Major joint replacement of lower extremity	2,931	-1.7	-1.2	2.6	0.06	2.4	-0.2	1.3	-1.3
	Simple pneumonia and respiratory infections	1,208	-4.4	-1.8	8.2	-0.16	-3.1	1.1	-0.5	-5.6

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively.

HHA = home health agency. ESRD = end stage renal disease. HCC = hierarchical condition categories. ED = emergency department. PAC = post-acute care.

* These characteristics measure utilization of care in the six months prior to the qualifying hospitalization. Count of HCCs is based on the six months prior to the qualifying hospitalization.

Source: Lewin analysis of Medicare claims and enrollment data for episodes that began Q4 2011 through Q3 2016 for BPCI and comparison providers.

Exhibit 36: Relative Change in Patient Resource Intensity from Assessment Measures, by Clinical Episode Strata, Baseline to Intervention, Model 3 HHA, Q4 2013 – Q3 2016

Change in Patient Mix	Clinical Episode	Number of Episodes, Q4 2013- Q3 2016*	Poor Overall Status	Require Use of Bedside Commode or are Totally Dependent in Toileting	Require Assistance Transferring or are Unable to Transfer (e.g. from bed to wheelchair)	Require Walker or More Assistance Ambulating	Dependent in Maintaining Self-care[4]	Dependent in Ambulating	Dependent in Transferring	Impaired Vision or Hearing	Impaired Cognition	Unhealed Pressure Ulcer	Short of Breath from Moderate to No Exertion	Not Likely to Receive Assistance in ADL	Caregiver Needs Training to Provide Supervision and Safety, is Unlikely to Provide Help, or is not Present	Incontinence	Depressive Symptoms
Less resource intensive	Congestive heart failure	4,119	-12.4	-7.7	-29.7	-4.3	-1.4	0.1	-0.8	0.0	-0.5	-0.7	-3.4	1.5	-0.6	0.3	0.5
No consistent pattern	Major joint replacement of the lower extremity	2,931	-0.9	-8.9	-2.9	-0.3	-0.7	-0.2	-0.4	0.7	1.8	-0.3	3.9	1.1	1.1	-2.3	0.5
	Simple pneumonia and respiratory infections	1,208	-2.6	5.0	0.7	-0.4	3.4	2.6	2.5	0.3	-1.9	-1.5	1.9	1.1	-0.9	0.6	-2.8

Note: The estimates in this exhibit are the results of a difference-in-differences (DiD) model. Positive DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light green shaded cells, respectively. Negative DiD estimates that are significant at the 5% or 10% significance level are indicated by dark and light orange shaded cells, respectively. HHA = home health agency. ADL = Activities of Daily Living.
The Outcome and Assessment Information Set (OASIS) assessment is administered within five days of the start of care date. The measurement period for outcomes listed above is upon assessment, except for depressive symptoms, for which the measurement period covers the 14 days prior to the assessment date.

*Assessment data was not available for all episodes. This table is limited to the episodes where we had the initial patient assessment data.

Source: Lewin analysis of Medicare claims, enrollment, and (OASIS data for episodes that began Q4 2011 through Q3 2012 (baseline) and Q4 2013 through Q3 2016 (intervention period) for BPCI EIs and the matched comparison providers.

V. Net Savings to Medicare

As demonstrated in this evaluation, BPCI participants have successfully reduced Medicare allowed payments across a variety of clinical episodes. Under Model 2, 25 of the 32 ACH clinical episodes we examined had declines in Medicare allowed payments among BPCI participants relative to the comparison group. For five of these clinical episodes, the relative declines in payments were statistically significant. Under Model 3, there were relative declines in allowed payments for eight of the 11 SNF-initiated clinical episodes, six of which were statistically significant, and two of the three HHA-initiated clinical episodes. To determine if these largely encouraging outcomes translate to Medicare program savings, the estimated reductions in payments need to be adjusted by the net payment reconciliation amounts (NPRA) that Medicare paid or recovered from participants.

The NPRA is the difference between the episode target amount and the participant's Medicare episode payments. Participants with episode payments below the target receive the NPRA, while those with payments above the target have to repay the NPRA to Medicare. These NPRA payments have not been accounted for in the payment estimates presented thus far.

In the preliminary analysis presented below, we adjusted the estimated change in episode allowed payments from the DiD methodology by the average NPRA by clinical episode. Based on this analysis, even though Medicare allowed payments declined for BPCI participants relative to the comparison group for most clinical episodes, the BPCI initiative did not always result in net savings to Medicare. For many Model 2 and Model 3 clinical episodes Medicare NPRA payments to BPCI participants were larger than the reductions in allowed Medicare payments due to BPCI. As a result, Medicare did not achieve expected savings from the discount applied to historical episode payments.[25] In fact, Medicare likely incurred additional costs under Model 3. This finding suggests that the Medicare program will not benefit to the full extent of the discount unless there are adjustments to the BPCI target price and reconciliation methodologies.

A. Methods

We compared the DiD estimate of the relative change in Medicare allowed payments with the average NPRA for all Model 2 ACH and Model 3 SNF and HHA clinical episodes that had sufficient volume,[26] pooling across all episode lengths, for BPCI episodes initiated between Q4 2013 and Q3 2016.

[25] BPCI Medicare discount is 2% for 90-day episodes in Model 2 and 3% for Model 2 episodes of other lengths as well as for all Model 3 episodes.

[26] Each unique combination of Model-EI type-clinical episode was considered to have a sufficient sample size for meaningful analysis if there were at least 20 EIs with a minimum of 1,000 episodes. Though these participation criteria were applied, no formal power calculation was conducted to assess minimum sample size.

The DiD estimate is an estimate of the change in Medicare standardized allowed amounts for the inpatient stay plus 90 days post-discharge for the BPCI episodes relative to the change for comparison episodes.[27] The DiD approach provides the best estimate of BPCI causal effect, that is, the change in Medicare allowed episode payments due to BPCI. More precisely, the DiD approach measures the change in risk-adjusted Medicare allowed payments for BPCI episodes relative to the change in risk-adjusted episode payments for a matched comparison group of episode as the counterfactual. It accounts for self-selection bias due to voluntary participation in BPCI, Medicare payment updates, changes in patient mix, changes in medical practice, and other secular changes in health care delivery that would affect episode payments. We then multiply the DiD by -1 so that a positive value indicates savings; this measure is referred to as the DiD savings estimate.

We obtained the average NPRA per clinical episode from CMS. We assumed that participants repaid any negative NPRA through the entire intervention period, even though downside risk was waived from Q4 2013 through Q4 2014.[28]

The net savings to Medicare is estimated as the difference between the DiD savings estimate and average NPRA for each clinical episode. Exhibits 37 and 38 present six potential scenarios of the net Medicare savings estimate based on various relationships between the DiD savings estimate and average NPRA. A positive DiD savings estimate indicates an estimated decrease in Medicare allowed payments per clinical episode. A negative DiD savings estimate indicates an estimated increase in Medicare allowed payments per clinical episode. A positive average NPRA is the amount per episode paid by Medicare to participants. A negative average NPRA is the amount per episode that participants repaid to Medicare. A positive value for the net Medicare savings outcome indicates aggregate savings to the program; a negative value indicates that the Medicare program paid out more than it would have absent BPCI.

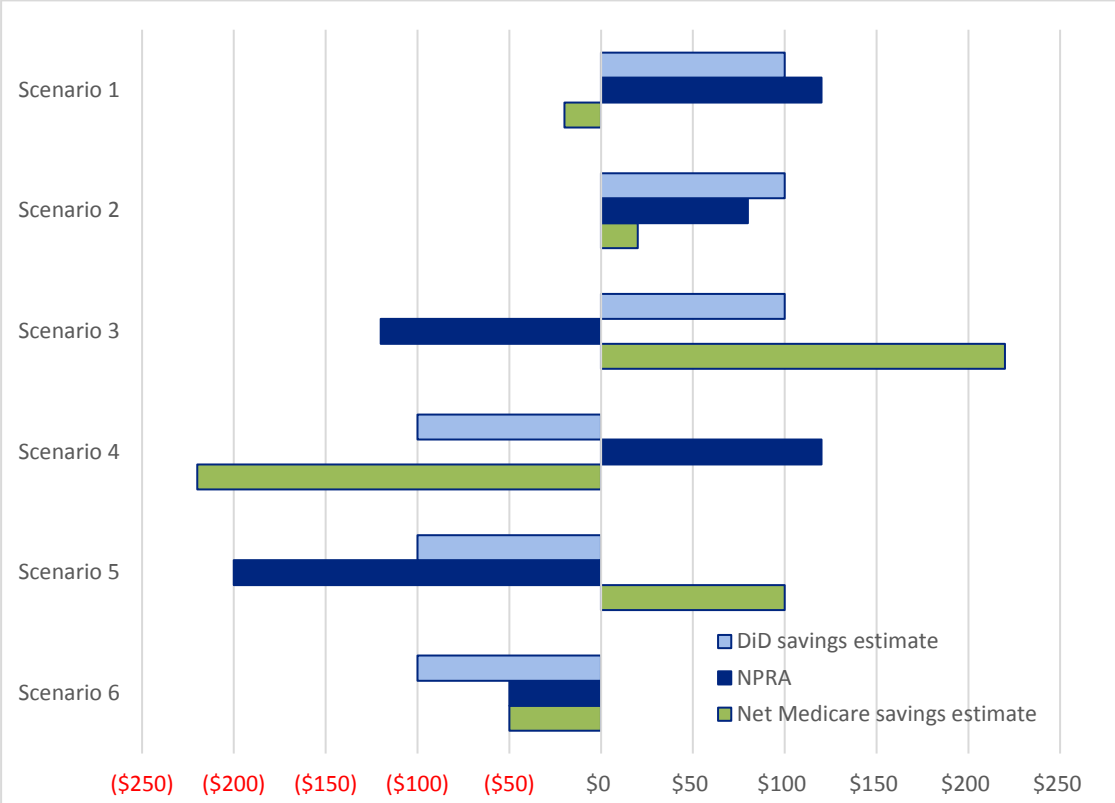
Exhibit 37: Potential Scenarios of Net Medicare Savings Estimates

Outcome	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
DiD savings estimate	$100	$100	$100	-$100	-$100	-$100
NPRA	$120	$80	-$120	$120	-$200	-$50
Net Medicare savings estimate	-$20	$20	$220	-$220	$100	-$50

[27] The DiD savings estimate used in this analysis was defined the same across Models 2 and 3. This outcome differs from the Model 3 payment outcome presented in Sections IV.A and IV.B that only included the Medicare allowed payment for services provided within the bundle for 90 day episodes. The most notable difference between these two outcomes is that the qualifying inpatient stay for Model 3 episodes is included in the Medicare Savings analysis.

[28] CMS also eliminated negative NPRA for any episode of care that was initiated as a result of the episode attribution issues caused by the incorrect PGP Reassignment Lists for the period of January 1, 2015 through September 30, 2016.

Exhibit 38: Graphical Depiction of Potential Scenarios of Net Medicare Savings Estimates



B. Results

1. Model 2 ACH

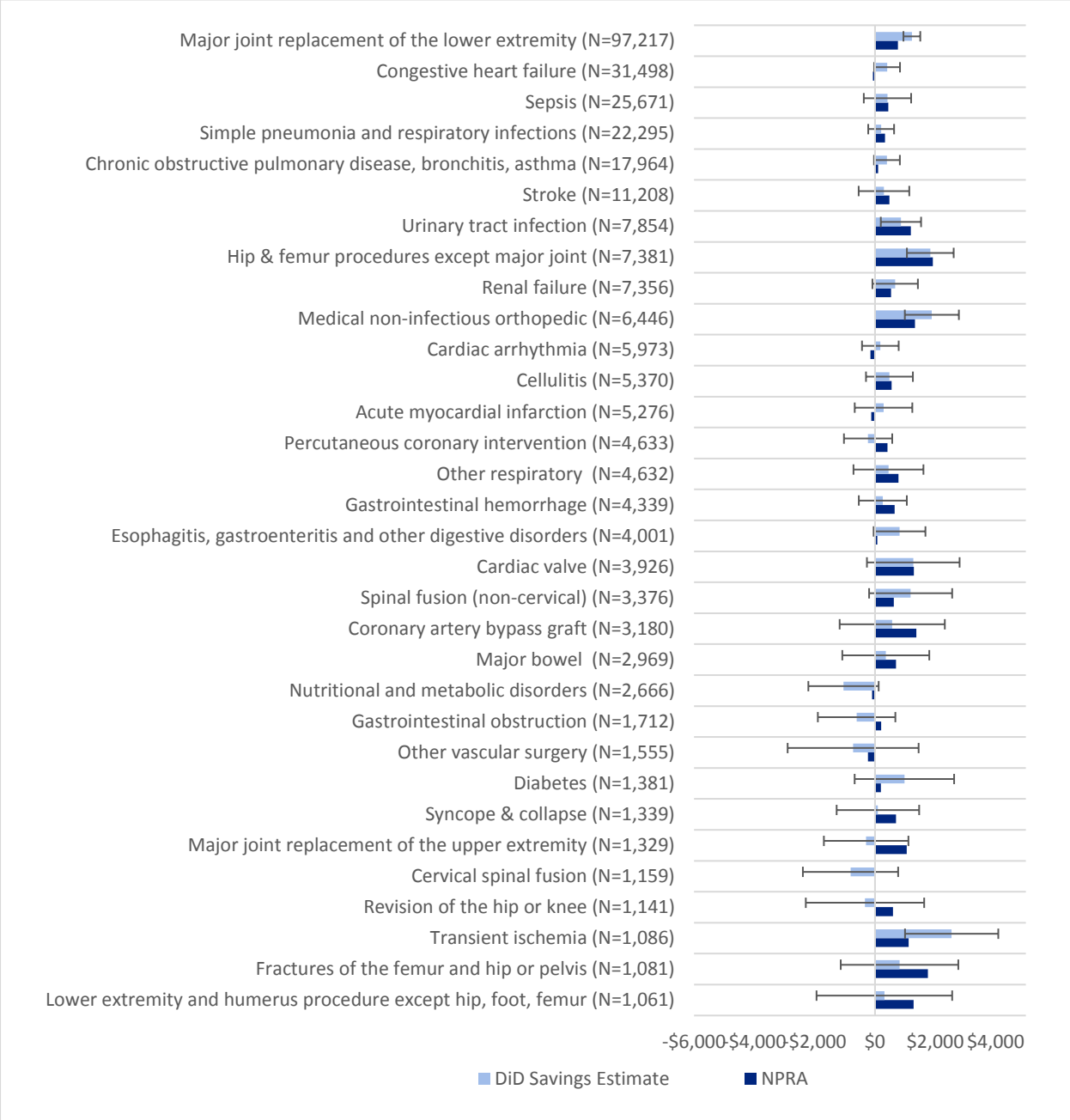
Across the 32 Model 2 ACH clinical episodes examined, the average NPRA ranged from -$230 for other vascular surgery to $1,919 for hip and femur procedures except major joint clinical episodes. For 26 of these clinical episodes, the average NPRA was positive.

The DiD savings estimate was less than the average NPRA for 21 of the 32 clinical episodes, which means that the average NPRA paid to participants was more than the estimated change in payments due to BPCI (Exhibit 39 and **Appendix N**). The difference between the DiD savings estimate and the average NPRA across these 21 clinical episodes ranged from -$23 for cardiac valve to -$1,348 for major joint replacement of the upper extremity clinical episodes, but none of the differences between the DiD savings estimate and the average NPRA were statistically significant. Although the DiD savings estimate indicated a relative reduction in Medicare allowed payments for 14 of these 21 clinical episodes, the average NPRA paid to participants was higher than the decline in payments. For the other seven clinical episodes, the DiD savings estimate indicated a relative increase in allowed payments. For three of these clinical episodes the NPRA was negative, indicating that NPRA was owed to Medicare, however, the repayments were not sufficient to offset the estimated increase in Medicare allowed payments.

The DiD savings estimate was greater than the average NPRA for 11 clinical episodes that contributed over 60% of the episode volume included in this analysis. The difference between the

DiD savings estimate and NPRA, which is an estimate of net savings to Medicare, ranged from $141 for renal failure to $1,423 for transient ischemia clinical episodes. The net Medicare savings estimate was positive and statistically significant for MJRLE and CHF clinical episodes ($p<0.10$).

Exhibit 39. DiD Savings Estimate and Average NPRA by Clinical Episode, Model 2 ACH Episodes, Q4 2013 - Q3 2016



Note: The DiD savings estimates reflect total Medicare allowed Part A and B payments for the qualifying inpatient stay plus 90 day post discharge period. The line represents the 90th percent confidence interval for the DiD savings estimate. A positive DiD savings estimate indicates an estimated decrease in Medicare allowed payments per clinical episode. A negative DiD savings estimate indicates an estimated increase in Medicare allowed payments per clinical episode. The average NPRA per episode is calculated as the target price minus the actual Medicare episode payments divided by the total number of episodes. A positive average NPRA is the amount per episode paid by Medicare to participants. A negative average NPRA is the amount per episode that participants repaid to Medicare. Results are sorted in order of episode volume.

Source: The DiD savings estimates are based on Lewin analysis of Medicare claims and enrollment data for episodes that began in Q4 2011 through Q3 2016 for BPCI and comparison providers. The average NPRA per episode is based on Lewin analysis of reconciliation data from the BPCI program for BPCI episodes that began in Q4 2013 through Q3 2016.

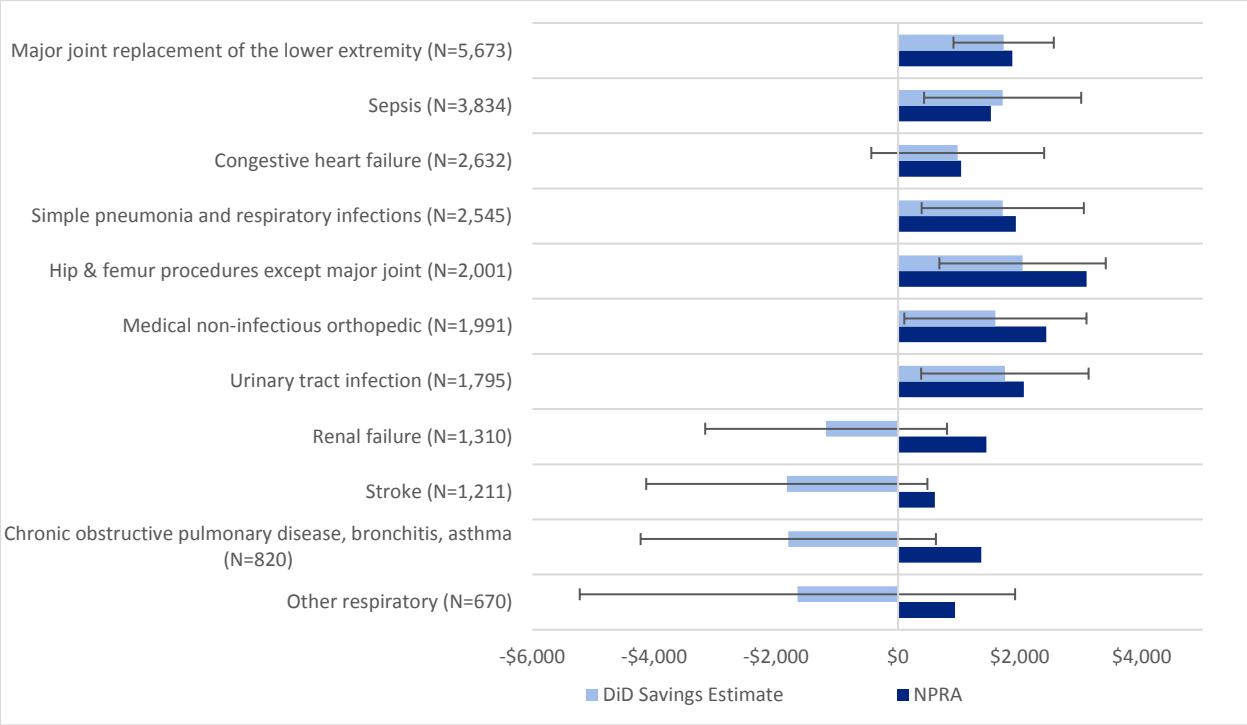
2. Model 3 SNFs

The average NPRA was positive for all 11 Model 3 SNF clinical episodes we examined, ranging from $603 for stroke to $3,094 for hip & femur procedures except major joint (Exhibit 40).

The DiD savings estimate was less than the NPRA for 10 of 11 clinical episodes (see Exhibit 40 and **Appendix N**). This analysis implies that the Medicare program likely paid out more for these 10 clinical episodes under BPCI than it would have otherwise, although the probability is greatest for COPD, renal failure, and stroke clinical episodes because the differences between the DiD savings estimate and NPRA were statistically significant ($p<0.10$). For six of these clinical episodes, we estimated a relative decline in Medicare allowed payments, but the average NPRA payments were higher. There were four clinical episodes where the DiD savings estimate was negative, meaning there was an estimated increase in Medicare allowed payments. In addition, Medicare paid NPRA to participants for these four clinical episodes, increasing Medicare spending above the DiD savings estimate.

Sepsis was the only clinical episode where the DiD savings estimate was higher than the average NPRA. This suggests that the Medicare program may have saved on this clinical episode, although the difference between the DID savings estimate and the NPRA was not statistically significant.

Exhibit 40. DiD Savings Estimate and Average NPRA by Clinical Episode, Model 3 SNFs, Q4 2013 - Q3 2016



Note: The DiD savings estimates reflect total Medicare allowed Part A and B payments for the qualifying inpatient stay plus 90 day post discharge period. The line represents the 90th percent confidence interval for the DiD savings estimate. A positive DiD savings estimate indicates an estimated decrease in Medicare allowed payments per clinical episode. A negative DiD savings estimate indicates an estimated increase in Medicare allowed payments per clinical episode. The average NPRA per episode is calculated as the target price minus the actual Medicare episode payments divided by the total number of episodes. A positive average NPRA is the amount per episode paid by Medicare to participants. A negative average NPRA is the amount per episode that participants repaid to Medicare. Results are sorted in order of episode volume.

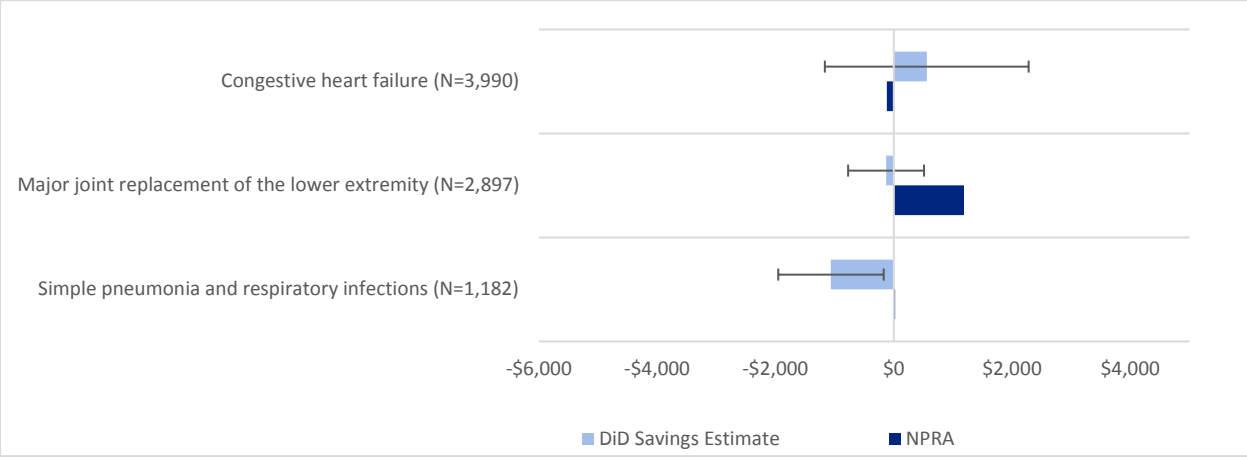
Source: The DiD savings estimates are based on Lewin analysis of Medicare claims and enrollment data for episodes that began in Q4 2011 through Q3 2016 for BPCI and comparison providers. The average NPRA per episode is based on Lewin analysis of reconciliation data from the BPCI program for BPCI episodes that began in Q4 2013 through Q3 2016.

3. Model 3 HHAs

The average NPRA was positive for two of the three Model 3 HHA clinical episodes we examined. NPRA Medicare paid to the participants varied from $23 for simple pneumonia and respiratory infections to $1,188 for MJRLE clinical episodes (Exhibit 41 and **Appendix N**). For these two clinical episodes, the DiD savings estimates were also negative, further increasing the additional costs to Medicare (by $1,086 per simple pneumonia and respiratory infections episodes and by $1,319 per MJRLE episodes). This analysis indicates that the Medicare program likely paid out more for these two clinical episodes under BPCI than it would have otherwise. This is especially likely for MJRLE because the difference between the DiD savings estimate and NPRA was statistically significant ($p<0.10$).

The DiD savings estimate was positive for CHF clinical episodes. Furthermore, the average NPRA was negative (i.e., participants repaid Medicare) adding to the DiD savings estimate. The estimated net savings to Medicare for Model 3 HHA CHF is $675 per episode although the difference between the NPRA and DiD savings estimate was not statistically significant.

Exhibit 41. DiD Savings Estimate and Average NPRA by Clinical Episode, Model 3 HHAs, Q4 2013 - Q3 2016



Note: The DiD savings estimates reflect total Medicare allowed Part A and B payments for the qualifying inpatient stay plus 90 day post discharge period. The line represents the 90th percent confidence interval for the DiD savings estimate. A positive DiD savings estimate indicates an estimated decrease in Medicare allowed payments per clinical episode. A negative DiD savings estimate indicates an estimated increase in Medicare allowed payments per clinical episode. The average NPRA per episode is calculated as the target price minus the actual Medicare episode payments divided by the total number of episodes. A positive average NPRA is the amount per episode paid by Medicare to participants. A negative average NPRA is the amount per episode that participants repaid to Medicare. Results are sorted in order of episode volume.

Source: The DiD savings estimates are based on Lewin analysis of Medicare claims and enrollment data for episodes that began in Q4 2011 through Q3 2016 for BPCI and comparison providers. The average NPRA per episode is based on Lewin analysis of reconciliation data from the BPCI program for BPCI episodes that began in Q4 2013 through Q3 2016.

C. Discussion

Although the DiD savings estimates presented in this report indicate that Medicare allowed payments declined for 25 Model 2 ACH clinical episodes and 10 Model 3 SNF and HHA clinical episodes due to BPCI, the Medicare program probably did not achieve savings for 22 of the 35 clinical episodes. This is because the DiD savings estimates did not account for the NPRA paid to BPCI participants when their episode payments were below their target amount.

The BPCI initiative was designed to reward participants through NPRA payments for redesigning care, improving coordination across providers involved in an episode, and implementing other changes that would lower episode costs while maintaining or improving quality. Participants did indeed respond to these incentives and received positive NPRA for many clinical episodes; yet, the NPRA payments to participants were often higher than the estimated decline in Medicare allowed payments, even though the NPRA calculations incorporated a discount that was intended to translate into automatic savings to the Medicare program.[29]

Medicare did not achieve savings equal to the discount amount under BPCI because of the determination of the benchmark amounts, discounted target prices and the reconciliation process. The benchmark amount, based largely on historical payments, was intended to represent what fee-for-service payments would have been absent the BPCI initiative. There are several reasons

[29] NPRA was calculated as the difference between a target price and actual payments. The target price for each clinical episode was a discounted benchmark that was based on trended, participant-specific, historical episode payments. The discount was intended to ensure that the Medicare program achieved savings.

why the benchmark amount may not have been an accurate representation of this counterfactual. The national growth rate used to update historical payments to the performance period may have overestimated Medicare fee-for-service spending for the providers that chose to participate in BPCI. The participants' mix of patients may have changed, such that historical payments no longer reflected the mix of services used by their patients under BPCI. Other changes in the health care market could affect episode service mix differentially for BPCI participants, such as ACOs or other payer initiatives. Thus, to ensure savings, Medicare may need to make adjustments to the methodology used to calculate the benchmark amount in its episode payment models and the discount it applies to calculate the target price.

The BPCI benchmark methodology could have resulted in episode payment amounts that were higher or lower than what they would have been absent the BPCI initiative. It is less likely that the Medicare program would have been advantaged by benchmarks that were too low and more likely that Medicare would have been disadvantaged by benchmarks that were too high. This is because BPCI participants had multiple options under the initiative, including the ability to drop clinical episodes on a quarterly basis, and they were unlikely to stay with a configuration of options under BPCI[30] that was not beneficial to them. As a result, any errors with the benchmark methodology disproportionately resulted in higher Medicare payments, even after factoring in the discount to the benchmark to calculate the target price.

There are limitations with this analysis. First, the DiD savings estimate is based on standardized payments, which underestimates the savings. Preliminary analyses for the BPCI participants found that the actual Medicare payments are approximately 5% higher than the standardized payments. This may be due to higher representation of large, urban providers among BPCI participants. On the other hand, because the DiD savings estimate is based on Medicare allowed amounts, they include what the Medicare program paid plus beneficiary out-of-pocket expenses. To the extent the BPCI initiative changed the beneficiary out-of-pocket expenses, the DiD savings estimate based on allowed payments may overstate the change to Medicare payments associated with BPCI. In addition, the net savings to Medicare assumes there is no change in episode volume. If the volume of episodes increased due to BPCI, any savings to Medicare would decrease. Finally, the NPRA used in this analysis assumes that BPCI participants repaid negative NPRA even though downside risk was waived from Q4 2013 through Q4 2014.[31] In the fifth annual report, we will refine the analysis to address these limitations as appropriate.

[30] BPCI participants could choose to participate in any of the 48 clinical episodes, one of three options for bundle length and risk track. Risk track refers to the level of winsorization, that is, the outliers that are excluded from the reconciliation payment calculation.

[31] CMS also eliminated negative NPRA for any episode of care that was initiated as a result of the episode attribution issues caused by the incorrect PGP Reassignment Lists for the period of January 1, 2015 through September 30, 2016.

VI. Discussion and Conclusion

A. Discussion

This fourth annual report focuses on the impact of BPCI on payment, utilization, and quality for 32 clinical episodes under Model 2 and 14 clinical episodes under Model 3 during the first three years of the initiative, from Q4 2013 through Q3 2016.[32] In the prior annual report, we began to see the impact of changes providers were making in response to the BPCI incentives. With the inclusion of another year of experience under the initiative, we now see further evidence of participants responding to the incentives as hypothesized.

Under Model 2, 422 BPCI-participating hospitals initiated over 353,000 episodes during the first 12 quarters of the initiative. Our study sample included 406 hospitals with an average of five quarters of experience in the initiative, and approximately 300,000 episodes of care. The average length of participation was less than the full three years of the initiative because providers could begin the risk-bearing phase of BPCI over a nine quarter period and because they could terminate their participation at any time. By October 2016, 213 of the 406 hospitals stopped participating in at least one of the clinical episodes included in our study sample, though 334 remained in the BPCI initiative by participating in other clinical episodes.

We observed a statistically significant decline in Medicare allowed payments for five ACH-initiated Model 2 clinical episodes: transient ischemia, MJRLE, medical non-infectious orthopedic, hip and femur procedures except major joint, and urinary tract infection. In the Year 3 annual report, based on the first two years of the initiative, only MJRLE had a statistically significant decline in Medicare allowed payments. The average reduction in Medicare payments across these five clinical episodes was 6.7% greater than what we would have expected without BPCI. The statistically significant declines in total payments for the inpatient stay and 90 day post-discharge period were driven by a reduction in PAC utilization, particularly IRF and SNF use. Even beyond the five clinical episodes with statistically significant declines in total payments, we observed a strong trend across most clinical episodes in reducing SNF and IRF utilization and increasing HH utilization, which corresponded to reductions in SNF and IRF payments and increases in HHA payments. The changes in PAC use and reductions in total payments do not appear to have a negative impact on the average quality of care received by patients in BPCI-participating hospitals. In general, BPCI did not appear to have a systematic effect, either positive or negative, on the quality of care delivered. One possible exception is for self-reported care experience. Survey results reveal that respondents treated at BPCI hospitals reported slightly worse care experiences and lower overall satisfaction with recovery than did comparison survey respondents treated at non-BPCI hospitals. However, these differences were small and were not accompanied by worse changes in self-reported functional status among survey respondents.

Under Model 3, 873 BPCI-participating SNFs and 116 BPCI-participating HHAs initiated nearly 50,000 and 15,000 episodes, respectively, during the first three years of the initiative. Our study sample included 493 SNFs and 71 HHAs that initiated approximately 24,680 and 8,258 episodes of care. The ability of participants to begin the risk-bearing phase of BPCI over a period of nine

[32] We stratified MJRLE into fractures/non-fracture episodes and CABG into emergent/non-emergent, so our analysis was based on 34 Model 2 clinical episode strata and 15 Model 3 clinical episode strata. The patient survey included respondents from 21 Model 2 clinical episodes.

quarters and terminate their participation at any time, results in an average length of participation in BPCI of five quarters for SNFs and HHAs.

After three years of BPCI, we found statistically significant reductions in Medicare payments for the following Model 3 SNF clinical episodes: hip and femur procedures except major joint, MJRLE, sepsis, simple pneumonia and respiratory infections, and urinary tract infections. After two years of BPCI, only MJRLE showed a statistically significant decline in the total allowed amount included in the bundle definition among 90-day episodes. Given that 36% of SNF episodes included in the impact estimates were associated with SNFs that stopped participating in that clinical episode, it is unclear whether the estimated impact of BPCI is due to care redesign, additional experience with the initiative, or by changes in the composition of the SNFs that remained in the initiative.

After three years of BPCI, there continued to be no statistically significant change in payments for the three largest HHA-initiated clinical episodes or in the use of services.

The statistically significant declines in Medicare allowed payments do not directly translate into Medicare program savings. This is because the DiD savings estimates do not account for the NPRA paid to BPCI participants. After making this adjustment, we determined that Medicare achieved program savings only for MJRLE and CHF clinical episodes under Model 2 ACH. However, it should be noted that our estimate of savings may be too high because the NPRA used in this analysis does not account for the fact that participants were not required to repay NPRA to Medicare for a portion of the initiative.

There were a few indications in the claim-based measures that BPCI had a negative effect on quality of care under Model 3. Most changes in quality measures were not statistically significant, and sensitivity analyses also indicated that the statistical significance of some results may have been due to the chance selection of particular comparison episodes. These outcomes will be subject to additional analyses in the next annual report.

B. Limitations

The primary analytic approach for this evaluation relies on the differential change in claim-based measures between the BPCI participants and a comparison group to infer the impact of BPCI. The strength of these results is therefore dependent on how well the comparison group represents what would have happened absent the BPCI initiative. We have matched providers and episodes on key factors expected to affect provider responses to BPCI that are available in Medicare administrative data in other sources. For Model 3 SNF and HHA EIs, however, the comparison episodes were not as close a match as we would like, particularly for quality outcomes, even after multiple attempts to improve the match. Additional sensitivity analyses with repeated samples of comparison episodes on a subset of Model 2 and 3 quality outcomes indicate that the statistical significance of some results may be due to the chance selection of particular comparison episodes.

Because the DiD estimate attributes differences in trends between BPCI and the comparison group during the intervention period to BPCI, it is essential that the comparison and BPCI providers have parallel trends for a given outcome during the baseline period. With this in mind, we matched BPCI participants with other providers on the unadjusted levels in 2011 and changes from 2011 to 2012 in emergency department visits, readmissions, mortality, and total Medicare

payments for the inpatient stay plus 90 days post discharge. Despite this approach, we rejected the null hypothesis that there were parallel trends for 55 of 450 (12%) DiD estimates we tested.[33] Thus, for these estimates, the underlying assumptions of the DiD method were violated, which may bias our results. Furthermore, the baseline levels differed for many of the Model 3 SNF quality outcomes in particular.

With respect to the survey results, because we do not have survey data predating the BPCI initiative, we cannot be certain whether differences in responses between patients treated in BPCI and comparison hospitals were caused by BPCI or were just the continuation of a trend that existed prior to BPCI.

The estimates of the BPCI impact on payment, utilization, and quality of care account for differences in provider and market characteristics, as well as patient mix that is measurable with claims data. As with all risk adjustment models, however, we cannot be sure that we controlled for all characteristics that may affect these outcomes.

The Medicare program savings analysis is based on standardized allowed payments, not actual Medicare program outlays. The net effect of using standardized payments that include beneficiary out-of-pocket amounts on the savings estimates is not known. The NPRA we used in this analysis does not account for the fact that participants were not required to repay NPRA for a portion of this initiative, which results in an overestimate of Medicare program savings.

The BPCI initiative tests a wide range of configurations, including the two bundled payment Models presented in this report and multiple options for providers and other organizations to participate in up to 48 clinical episodes. Because we are measuring multiple outcomes across the range of Model, participant, and clinical episode combinations, by chance alone some results will appear statistically significant, although in reality these are not true effects. However, this is an issue for isolated results and much less so for patterns of results. We do not derive conclusions based on isolated results.

C. Future Analyses

The next summative report for the BPCI evaluation will incorporate results and conclusions based on a multitude of analyses conducted over the five year contract. One of the most important new analyses will be of the impact of BPCI on PGP-initiated episodes.[34] As of July 2016, PGPs accounted for approximately 40% of Model 2 EIs and 13% of Model 3 EIs. We will include the experience of PGPs and the impact of BPCI on Model 2 PGP-initiated episode payments and quality based on claims and beneficiary survey results. This will help complete the picture of the impact of BPCI on multiple outcomes across all types of EIs.

We will strengthen and expand several analyses in the next annual report with the larger volume of episodes due to more BPCI experience. A larger sample will allow us to examine the factors

[33] We could not conduct the test for an additional 14 DiD estimates due to small sample size. Because we tested the null hypothesis that there were parallel trends at the 10% significance level, this proportion is slightly above the 10% that would be observed by chance alone.

[34] The lists of BPCI-participating physicians by PGP from Q1 2016 onward were corrected in Q1 2017. The evaluation team implemented and tested the revised methodology in July 2017.

that contribute to whether a participant achieves success (as measured by NPRA) under BPCI for a broader set of clinical episodes. To date, we have estimated the impact of BPCI on the average beneficiary. The larger sample will allow us to estimate the impact of BPCI on quality of care and beneficiary satisfaction among vulnerable beneficiaries. In addition, new waves of the beneficiary survey will allow us to assess the impact on beneficiary satisfaction and quality of care for more clinical episodes and EI types. Analyses in the next report will also help in understanding which providers end their participation in a clinical episode. We will include DiD estimates of the total payment outcome based on intent-to-treat methods, which will incorporate all episodes from BPCI participants, even after they stopped participating in the clinical episode. For the first time, we will also assess whether BPCI caused an increase in episode volume, which could offset any savings to the Medicare program. We will focus this analysis on estimating the impact of BPCI on the number of MJRLE elective surgeries, because it is the highest volume episode in the initiative and an elective procedure, which may be the most susceptible to volume changes due to BPCI. Finally, we will refine our methodology, addressing the limitations described above, to estimate the impact of BPCI on Medicare program spending.

D. Conclusion

BPCI participants have responded to BPCI incentives as anticipated. Participants reduced Medicare payments relative to the comparison group, primarily by reducing SNF payments. A substantial number of participants stopped participating in a given clinical episode, although they remained in BPCI for other clinical episodes. The analysis indicates that in only limited circumstances does BPCI generate savings to the Medicare program. Key design features of BPCI, such as the ability to withdraw at any time and the calculation of NPRA, appear to have reduced savings to the Medicare program due to BPCI.

CMMI officials reported that the rapid cycle evaluation results from the BPCI evaluation informed several of the design components of CMS' new episode payment model, BPCI Advanced, which will be implemented under the authority of CMMI. BPCI Advanced, which qualifies as an Advanced Alternative Payment Model, will start at the conclusion of BPCI. It is based on BPCI Model 2 because CMS achieved the most favorable evaluation results under Model 2. Even though payments for some clinical episodes declined under Model 3, changes in patient mix for some of the clinical episodes raised the possibility that the reductions in payment may have been at least partly due to changes in patient mix that may not be adequately accounted for by the claims-based risk adjustment used in BPCI Advanced.[35] Other key differences between BPCI and BPCI Advanced include modified target prices in BPCI Advanced that incorporate risk adjustment for patient mix and reflect peer performance and a higher discount. Changes to the target prices are intended to encourage both high and low cost providers to participate, which would lessen the self-selection we have seen in BPCI. Some clinical episodes were not included in BPCI Advanced due to high clinical heterogeneity or small volume. In addition, the participant entry and exit opportunities are scaled back under BPCI Advanced. Like all models tested by CMS, there will be a formal, independent evaluation to assess the impact of BPCI Advanced, including changes in quality of care and Medicare savings as well as any unintended consequences.

[35] For the full details of BPCI Advanced model design, please visit https://innovation.cms.gov/initiatives/bpci-advanced.