LEWINGROUP®

CMS Bundled Payments for Care Improvement Advanced Model: Fourth Evaluation Report

Final

Prepared for:

CMS

Submitted by:

The Lewin Group, Inc. with our partners Abt Associates, GDIT, and Telligen

June 2023

CMS Bundled Payments for Care Improvement Advanced Model: Fourth Evaluation Report

The Lewin Group, Inc.

With our partners Abt Associates, GDIT, and Telligen

Authors:

Julie Somers, Aylin Bradley, Kelsey Bacon, Shannon Bicott, Elizabeth Chung, Timothy Collins, Brady Durst, Katrina Drager, Kassandra Fate, Olivia Hayden, Raina Kelly, Dean Khazanov, Grace Kim, Ian Morrall, Conor Nistler, Lorissa Pagán, Hiral Patel, Patrick Roehm, Yujing Shen, Jingwei Sun, Isabella Swanson, Wafa Tarazi, Megan Tebbenhoff, Matt Trombley, Hannah Ableman, Abbi Barry, Louisa Buatti, Andrew Evans, Andrea Hassol, Catherine Hersey, Caroline Logan, Kelly Necastro, Sharmini Radakrishnan, Brittany Tuwamo, Katharine Witgert, Christine LaRocca, Becky Blystone, Maria Joseph-King, Nyaradzo Longinaker, Lauren Birmingham, Colleen Kummet, Chris Alleman

Lewin's address:

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

Federal Project Officer:

Daver Kahvecioglu
Division of Data, Research, and Analytic Methods (DRAM)
Research and Rapid Cycle Evaluation Group (RREG),
Center for Medicare & Medicaid Innovation (CMMI),
Centers for Medicare & Medicaid Services (CMS)

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2014-000331 Task Order 75FCMC18F0089.

Table of Contents

List of Appendices

Executive Summary

A. Introduction

The Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) launched the Bundled Payments for Care Improvement Advanced (BPCI Advanced) Model on October 1, 2018. BPCI Advanced is an Advanced Alternative Payment Model (Advanced APM), under the Quality Payment Program (QPP), and tests whether linking payments for an episode of care will incentivize health care providers to invest in innovation and care redesign to improve care coordination and reduce expenditures, while maintaining or improving the quality of care for Medicare fee-for-service (FFS) beneficiaries. BPCI Advanced is the next iteration of the Bundled Payments for Care Improvement (BPCI) Initiative Model 2 which ran from October 2013 to September 2018. The BPCI Advanced Model was initially set to continue through December 2023. CMS recently announced that the model will be extended for two years and will now continue through December 2025.

BPCI Advanced is a voluntary episode payment model in which a participant enters into an agreement with CMS to be financially responsible for the cost and quality of health care services during an episode of care. There are two categories of participants: convener participants and non-convener participants. A convener participant (convener) has at least one downstream episode initiator. Downstream episode initiators are either acute care hospitals or physician group practices (PGP). A convener holds financial risk on behalf of its downstream episode initiators and commonly provides guidance or services to help its episode initiators succeed in the model. A non-convener participant is a hospital or PGP that bears the financial risk only for itself. Participants could join the model in Model Year 1 (beginning October 2018), when they chose among 32 clinical episodes, or Model Year 3 (beginning January 2020), when they chose among 34 clinical episodes, and there will be a new opportunity for participants to join in Model Year 7 (beginning January 2024). Beginning in Model Year 4, CMS grouped clinical episodes into eight clinical episode service line groups (CESLGs). Participants were required to choose among the eight CESLGs instead of individual clinical episodes and are accountable for each clinical episode within the CESLGs they selected unless they did not meet the minimum volume threshold for a clinical episode during the baseline period.

BPCI Advanced includes both inpatient and outpatient episodes. An inpatient episode begins with a hospitalization in which the discharge is categorized in a Medicare Severity-Diagnosis Related Group (MS-DRG) that is included in one of the participant's selected clinical episodes, and the episode extends for 90 days post discharge. An outpatient episode begins with a hospital outpatient procedure that is identified by a Healthcare Common Procedure Coding System (HCPCS) code that is included in one of the participant's selected clinical episodes, and the episode extends for 90 days after the procedure. Each episode is attributed to an episode initiator, which may be the hospital where the discharge or procedure occurred or the PGP for the attending or operating clinician.

BPCI Advanced uses a bundled payment approach, wherein CMS may make a payment to participants or participants may owe a payment to CMS after CMS reconciles all FFS expenditures for a clinical episode against a target price. Specifically, at the end of each performance period, episode payments for every episode initiator and their selected clinical episodes are reconciled

against a target price. If the episode initiator's episode payments are above the target price, then the participant may owe CMS a reconciliation amount. Conversely, if an episode initiator's episode payments are below the target price, the participant may receive a reconciliation amount from CMS. Reconciliation amounts are also adjusted by the episode initiator's performance on quality measures. Target prices are calculated for each combination of episode initiator, clinical episode, and hospital where the episode was initiated. Target prices are based on historical payments for the hospital where the episode was initiated, and they are updated to the current performance period based on prospective trends that factor in risk-adjusted spending of similar hospitals. Beginning in Model Year 4, target prices also have a retrospective trends adjustment to account for unanticipated deviations from projected trends. Target prices are also adjusted for differences in patient mix during the performance period relative to the baseline period. For PGP episode initiators, target prices incorporate adjustments for differences between PGP and hospital historical payments, but beginning in Model Year 4, target prices no longer incorporate those adjustments. Target prices are discounted by 3% which is intended to be Medicare savings under the model.

In this evaluation report, we provide estimates of the impact of the model on total payments, utilization, and quality of care (readmission and mortality rates) as measured in Medicare claims data, as well as estimates of Medicare program savings in Model Year 3 (Exhibit ES.1). The report also gives an early look at Model Year 4 by providing estimates of the differences in patient-reported functional status, care experiences, and satisfaction with overall care between BPCI Advanced and comparison respondents. Finally, in this report, we provide results of exploratory analyses of outcomes under BPCI Advanced for beneficiaries from populations that have been historically underserved.[1] Populations analyzed in this report include Black or African American beneficiaries, Hispanic beneficiaries, dual-eligible beneficiaries, beneficiaries living in rural areas, and beneficiaries living in areas with a high Area Deprivation Index (ADI) (see **Appendix C** for details about these definitions).

The analyses presented in this report cover Model Years 3 and 4, 2020 and 2021, respectively, two years that were impacted by the COVID-19 public health emergency (PHE). We include controls for COVID-19 in the regressions that estimate the results presented in this report, but the COVID-19 PHE had widespread effects on the entire health care system that may not be adequately captured. There was an overall decrease in patient volume which coincided with CMS recommendations that hospitals delay elective surgeries and non-essential medical and surgical procedures during the COVID-19 PHE. During site visits and interviews conducted by the evaluation team, participants reported that patients may have delayed care or wanted to avoid engaging with post-acute care (PAC) services, and they noted difficulties with staffing, burnout, and labor shortages. Because these factors would likely make it more difficult for participants to engage with the BPCI Advanced Model, CMS allowed participants to select participation agreement amendments that removed episodes with a COVID-19 diagnosis from reconciliation for Model Year 3 or removed all episodes from reconciliation. For Model Years 4 and 5, CMS continued the policy to exclude COVID-19 episodes from reconciliation. CMS later offered a voluntary bilateral amendment that would allow participants to be held accountable for COVID-19

[1] CMS' strategy introduced in 2021 seeks to promote equitable outcomes through high-quality, affordable, person-centered care, with a special focus on underserved communities. See https://innovation.cms.gov/strategic-direction for more information.

episodes in Model Year 5 (2022), but beginning in Model Year 6 (2023), all participants will be held accountable for all episodes with a COVID-19 diagnosis during the episode.

We estimate the impact of the BPCI Advanced Model by pooling episodes into different groups. We estimate impacts for all episodes pooled across the clinical episodes evaluated and by clinical episode type (medical or surgical) and episode initiator type (hospital or PGP), as well as by clinical episode. We selected these groupings because strategies for care redesign may vary by these categories resulting in different impacts of the model.

Exhibit ES.1: Analyses and Model Years Reflected in This Report

Component	Model Years 1 and 2 (2018-2019)	Model Year 3 (2020)	Model Year 4 (2021)
Impact of the Model on Payments, Utilization, and Quality		•	
Changes in Patient-reported Functional Status, Care Experiences, and Satisfaction			•
Medicare Program Savings	•	•	
Analyses of Beneficiaries From Populations That Have Been Historically Underserved	•	•	•

CMS made changes to the BPCI Advanced Model in Model Year 3, including but not limited to changing and adding clinical episodes. Model Year 3 was initially the last opportunity for organizations to enroll in the BPCI Advanced Model. In October 2022, CMS announced a two-year extension of the model and a new opportunity for participants to enroll for Model Year 7. In Model Year 4, CMS made significant changes to the BPCI Advanced Model pricing methodology, including adjusting the calculation of target prices to add a retrospective trend adjustment and adding new risk-adjustment factors for the *major joint replacement of the lower extremity* clinical episode. As mentioned above, CMS also required participants to select CESLGs instead of individual clinical episodes. There were other changes in Model Year 4 including changing the episode overlap methodology, removing the PGP offset, and adding new alternative quality measures. Beginning in Model Year 6, CMS implemented additional changes to improve the pricing methodology and keep providers and suppliers engaged in value-based care including a reduction in the CMS discount from 3% to 2% for medical episodes, a reduction in the peer group trend factor adjustment cap from 10% to 5% for all episodes, and converting the *major joint replacement of the upper extremity* clinical episode to a multi-setting clinical episode by allowing outpatient total shoulder arthroplasty procedures in the model.

B. Summary of Results

1. What was the impact of BPCI Advanced on episode payments, utilization, and quality of care during Model Year 3 (January 1, 2020 through December 31, 2020)?

Consistent with prior evaluation reports, for Model Year 3, the BPCI Advanced Model reduced total episode payments, PAC payments, discharges to institutional PAC settings, and skilled nursing facility (SNF) days relative to the comparison group. Across the clinical episodes

evaluated in Model Year 3, the BPCI Advanced Model reduced average standardized episode payments by $1,028 per episode, or 3.8% of the baseline mean, relative to the comparison group. Surgical clinical episodes had per-episode reductions that were more than twice as large as medical clinical episodes. The BPCI Advanced Model reduced average standardized payments by $796 per episode (or 3.1% of the baseline mean) for medical clinical episodes and by $1,800 per episode (or 5.8% of the baseline mean) for surgical clinical episodes.

Hospital and PGP episode initiators reduced per-episode payments for medical clinical episodes by roughly similar amounts (-$756 or -2.9% of the baseline mean for hospital episode initiators compared to -$667 or -2.7% of the baseline mean for PGP episode initiators). However, for surgical clinical episodes, the reduction in per-episode payments for PGP episode initiators was more than double the reduction of hospital episode initiators (-$2,147, or -6.9% of the baseline mean, for PGP episode initiators compared to -$933, or -3.0% of the baseline mean, for hospital episode initiators).

As in prior evaluation reports, the reductions in payments were primarily driven by reductions in SNF and inpatient rehabilitation facility (IRF) payments. SNF was a larger driver of payment reductions for medical clinical episodes (-$452 or -8.6% of the baseline mean for SNF payments and -$109 or -10.8% of the baseline mean for IRF payments) while IRF was a larger driver of payment reductions for surgical clinical episodes (-$533 or -9.0% of the baseline mean for SNF payments and -$754 or -44.1% of the baseline mean for IRF payments).

For medical clinical episodes, hospital and PGP episode initiators reduced the share of episodes first discharged to an institutional PAC setting (including SNF, IRF, and long-term care hospital) by similar amounts. However, for surgical clinical episodes, the reduction was larger for PGPs than for hospitals (the estimate for hospitals also was not statistically significant). Among episodes with at least one day of SNF use, hospital and PGP episode initiators reduced SNF days for both medical and surgical clinical episodes. For medical clinical episodes, the reduction in SNF days was larger for hospitals than for PGPs (the estimate for PGPs also was not statistically significant), while for surgical clinical episodes, the reduction was similar for hospitals and PGPs.

In Model Year 3, there was some evidence of a small improvement in quality. There was a reduction in the unplanned readmission rate (-0.83 percentage points (pp) or -2.8% of the baseline mean) and in the mortality rate (-0.64 pp or -3.4% of the baseline mean) for PGP medical clinical episodes. There was also a reduction in the readmission rate for all clinical episodes (-0.25 pp or -0.9% of the baseline mean). The reduction in the readmission rate for all clinical episodes was mostly driven by PGP medical clinical episodes.

2. Were there differences in patient-reported functional status, care experiences, and overall satisfaction with care between BPCI Advanced and comparison group respondents under BPCI Advanced Model Year 4 (January 1 through December 31, 2021)?

In Model Year 4, there was some evidence of unfavorable quality results from the beneficiary survey. BPCI Advanced respondents were slightly more likely to report unfavorable changes in functional status than comparison respondents for both hospital- and PGP-initiated episodes (Exhibit ES.2). However, differences between BPCI Advanced and comparison respondents were small, approximately 1.5 pp on average. For hospital-initiated episodes, BPCI Advanced

respondents were less likely to report favorable care experiences and the highest levels of satisfaction with care than comparison respondents (differences were roughly 1 pp). For PGP-initiated episodes, there was a mix of favorable and unfavorable results for care experiences and satisfaction with care across CESLGs.

Exhibit ES.2: Differences in Patient-Reported Change in Functional Status, Care Experiences, and Satisfaction with Care Between BPCI Advanced and Comparison Respondents, July – August 2021

Clinical Episode Type	Functional Status	Care Experiences & Satisfaction
Hospital Clinical Episodes	■	■
Medical	■	■
Surgical	○	○
PGP Clinical Episodes	■	○
Medical	■	▲
Surgical	○	■

While BPCI Advanced respondents were slightly less likely to report improvement in functional status than comparison respondents and slightly less likely to report favorable care experiences and satisfaction with care for some episode groupings, the differences were within 1 to 2 pp, on average.

***Note*:** This exhibit summarizes differences between BPCI Advanced and comparison respondents in patient-reported measures from the beneficiary survey. The functional status column provides a summary across 7 measures, and the care experiences and satisfaction column provide a summary across 10 measures. A solid red square represents generally unfavorable results across the measures, a solid green triangle represents generally favorable results across the measures, and the grey outlined circles represent mixed results or generally no differences between BPCI Advanced and comparison respondents. PGP = physician group practice; pp = percentage point(s).

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Small differences in patient-reported outcomes reflect meaningfully large differences that affect a small proportion of individuals because the survey measures were constructed from questions with distinct response categories (such as, agree versus disagree, some help needed with task versus complete help needed with task). Average differences in functional status and care experiences roughly translate to one or two additional respondents out of 100 reporting an unfavorable outcome relative to comparison respondents.

3. Did BPCI Advanced result in savings to Medicare during Model Year 3 (January 1 through December 31, 2020)?

In Model Year 3, BPCI Advanced had a net loss of $113.7 million, or 0.8% of Medicare payments under the counterfactual (that is, what Medicare payments would have been if the BPCI Advanced Model had not occurred), ranging from a loss of $20.1 million to $207.3 million based on a 90% confidence interval. Medical clinical episodes resulted in an estimated net loss to Medicare (-$200.5 million, or -1.9% of payments under the counterfactual), while surgical clinical episodes resulted in net savings to Medicare ($71.3 million, or 2.3% of payments under the counterfactual). The evidence suggests that target prices may have been too high for medical clinical episodes but more accurate for surgical clinical episodes in Model Year 3. The findings for Model Year 3 are similar to those for Model Years 1 and 2 as might be anticipated since the target price methodology remained unchanged in Model Years 1 through 3. To improve the model's ability to achieve

savings to Medicare, CMS made significant changes to the target price methodology for Model Year 4 and onward. Therefore, estimates of savings may differ for Model Year 4 and will be reported in the Fifth Evaluation Report.

4. Were there differences in outcomes under the BPCI Advanced Model for beneficiaries from populations that have been historically underserved?

In 2021, CMS launched a new strategy with a focus on advancing health equity and announced that all new models will be designed to reduce inequities in health care outcomes. We have begun to assess health equity under the BPCI Advanced Model for the first time in this evaluation report. Because the model was not designed to address health equity, these analyses assess potential unintended consequences of the model. We provide results from exploratory analyses of beneficiaries from populations that have been historically underserved, for whom we had sufficient data to analyze. Populations analyzed in this report include Black or African American beneficiaries, Hispanic beneficiaries, dual-eligible beneficiaries, beneficiaries living in rural areas, and beneficiaries living in areas with a high Area Deprivation Index (ADI) (Exhibit ES.3).

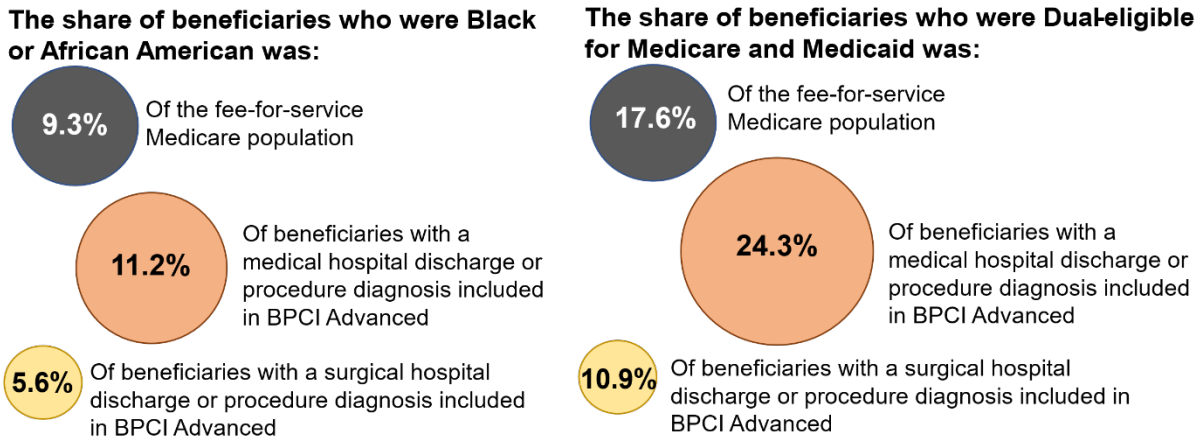
Exhibit ES.3: Analyses of Beneficiaries From Populations That Have Been Historically Underserved

Population Analyzed	Representation Under BPCI Advanced (MY 1-3)	Changes in Payments, Readmission Rates, and Mortality Rates (MY3)	Changes in Patient-reported Functional Status, Care Experiences, and Satisfaction (MY4)
Black or African American Beneficiaries	•	•	•
Hispanic Beneficiaries			•
Dual-Eligible Beneficiaries	•	•	•
Beneficiaries Living in Rural Areas			•
Beneficiaries Living in Areas with a High ADI			•

***Note*:** ADI = Area Deprivation Index; MY = Model Year.

Compared to the share of all FFS beneficiaries that were Black or African American, the share of FFS beneficiaries with medical stays that were Black or African American was higher (11.2% compared to 9.3%), while the share of FFS beneficiaries with surgical stays or procedures that were Black or African American was lower (5.6% compared to 9.3%) between October 2018 and December 2020 (Model Years 1 through 3) (Exhibit ES.4). These large representational differences between medical and surgical clinical episodes also existed for dual-eligible beneficiaries. Compared to the share of all FFS beneficiaries that were dual eligible, the share of FFS beneficiaries with medical stays or procedures that were dual eligible was higher (24.3% compared to 17.6%), while the share of FFS beneficiaries with surgical stays or procedures that were dual eligible was lower (10.9% compared to 17.6%). The differences in representation between medical and surgical clinical episodes for Black or African American and dual-eligible beneficiaries are features of the underlying FFS healthcare system and are caused by neither patient selection nor features of the BPCI Advanced Model (such as requiring beneficiary Part A and B coverage or excluding Maryland providers).

Exhibit ES.4: Representation of Black or African American Beneficiaries and Dual-Eligible Beneficiaries in the Fee-for-service Medicare Population and Among Beneficiaries With a Hospital Discharge or Procedure Diagnosis Included in BPCI Advanced, Model Years 1-3, October 1, 2018 – December 31, 2020



In the analyses of both Model Years 1 and 2 and Model Year 3, we found that all evaluated populations experienced declines in average episode payments relative to their comparison groups (Black or African American, Non-Hispanic White, dual-eligible, and nondual-eligible beneficiaries). The declines were larger for historically underserved populations (Black or African American and dual-eligible beneficiaries) than for the reference populations (Non-Hispanic White and nondual-eligible beneficiaries).

In Model Year 3, there were no changes in readmission rates for Black or African American beneficiaries or for dual-eligible beneficiaries relative to their comparison groups. In Model Years 1 and 2, the readmission rate declined for dual-eligible beneficiaries with surgical clinical episodes relative to their comparison group. While there were relative declines in readmission rates for some episode groupings of the reference populations—Non-Hispanic White beneficiaries with surgical clinical episodes in Model Years 1 and 2 and with medical clinical episodes in Model Year 3 and nondual-eligible beneficiaries with medical clinical episodes in Model Year 3—changes in readmission rates for the underserved populations (relative to their comparison groups) were not statistically different from that of their reference populations. This finding indicates that underserved populations were not any more or less likely than their reference populations to experience reductions in readmission rates under the model during Model Years 1 and 2 or Model Year 3.

There were no changes in mortality rates for Black or African American beneficiaries or for dual-eligible beneficiaries relative to their comparison groups in Model Years 1 and 2 or Model Year 3. However, in Model Year 3, there were differential increases in the mortality rate for both Black or African American beneficiaries and for dual-eligible beneficiaries with medical clinical episodes compared to their reference populations. These results are due to small, non-statistically significant increases in mortality rates for the underserved populations with medical clinical episodes relative to their comparison groups and statistically significant decreases for their reference populations with medical clinical episodes relative to their comparison groups. These findings suggest that Black or African American beneficiaries and dual-eligible beneficiaries in

medical clinical episodes were less likely than their reference populations to benefit from the BPCI Advanced Model.

Our beneficiary survey analysis includes respondents with episodes during Model Year 4. For patient-reported functional status, findings were neutral for all historically underserved populations analyzed except one. Dual-eligible BPCI Advanced respondents to the survey with hospital-initiated episodes were less likely to report favorable changes in functional status relative to dual-eligible comparison respondents with hospital-initiated episodes (Exhibit ES.5). The average difference was approximately 4 pp, indicating that roughly four additional dual-eligible BPCI Advanced respondents out of 100 reported unfavorable outcomes, relative to the comparison group. For nondual-eligible respondents with hospital-initiated episodes, there was no pattern of changes in functional status relative to nondual-eligible comparison respondents. Therefore, BPCI Advanced had an unfavorable differential impact on patient-reported functional status for dual-eligible respondents compared to the impact on nondual-eligible respondents, an average differential of roughly 3 pp. For two historically underserved populations, respondents living in high-ADI ZIP codes and respondents living in rural ZIP codes (with either hospital- or PGP-initiated episodes), BPCI Advanced had favorable differential impacts on patient-reported functional status compared to the impacts on their reference groups. These favorable differential impacts were driven by unfavorable changes in functional status for the reference populations and were roughly 2 to 3 pp in magnitude, on average.

Exhibit ES.5: Differences in Patient-reported Functional Status Between BPCI Advanced and Comparison Group Respondents, Model Year 4, July – August 2021

Episode-Initiator Type	Population	Difference (BPCI Advanced – Comparison)	Differential Impact (Underserved Difference – Reference Difference)
Hospital-Initiated Episodes	Black or African American	○	○
	Hispanic	○	○
	Dual Eligible	■	■
	High-ADI ZIP Code	○	▲
	Rural ZIP Code	○	▲
PGP-Initiated Episodes	Rural ZIP Code	○	▲

BPCI Advanced dual-eligible respondents with hospital episodes were 4 pp less likely to report favorable changes in functional status relative to comparison duals, and the impact on duals was more unfavorable than the impact on nonduals by about 3 pp. The model had favorable differential impacts on rural respondents and those in high-ADI ZIP codes of 2-3 pp due to unfavorable changes for reference populations.

***Note*:** This exhibit provides a summary of findings across the 7 patient-reported measures of functional status from the beneficiary survey. The underlying results are based on a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. A solid red square represents generally unfavorable results across the measures, a solid green triangle represents generally favorable results across the measures, and a grey outline circle represents mixed results or generally no differences between BPCI Advanced and comparison respondents. ADI = Area Deprivation Index; PGP = physician group practice; pp = percentage point(s).

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

For measures of care experiences and satisfaction in Model Year 4, findings were generally less favorable or mixed. BPCI Advanced respondents with hospital-initiated episodes who were (1) Hispanic, (2) lived in high-ADI ZIP codes, or (3) lived in rural ZIP codes were more likely to

report unfavorable care experiences than their counterparts in the comparison group, by roughly 3 to 4 pp (Exhibit ES.6). For example, Black or African American BPCI Advanced respondents and Hispanic BPCI Advanced respondents were both 7 to 8 pp less likely than comparison respondents to agree that medical staff took their preferences into account in deciding what services they should receive after leaving the hospital. The reference populations for those three underserved populations (Non-Hispanic White respondents, respondents that lived in non-high ADI ZIP codes, and respondents that lived in non-rural ZIP codes, respectively) reported mixed care experiences relative to their counterparts in the comparison group, resulting in larger unfavorable impacts on care experiences for the three historically underserved populations compared to their reference populations, averaging 2 to 3 pp in magnitude. For example, BPCI Advanced Hispanic respondents (relative to respondents in their comparison group) were less likely to agree that they were able to manage their health needs since returning home compared to Non-Hispanic White BPCI Advanced respondents (relative to respondents in their comparison group).

Exhibit ES.6: Differences in Patient-reported Care Experiences and Satisfaction with Care Between BPCI Advanced and Comparison Group Respondents, Model Year 4, July – August 2021

Episode-Initiator Type	Population	Difference (BPCI Advanced – Comparison)	Differential Impact (Underserved Difference – Reference Difference)
Hospital-Initiated Episodes	Black or African American	○	○
	Hispanic	■	■
	Dual Eligible	○	○
	High-ADI ZIP Code	■	■
	Rural ZIP Code	■	■
PGP-Initiated Episodes	Rural ZIP Code	▲	○

Favorable and unfavorable results for care experiences and satisfaction with care reflect relative differences and differential impacts in the range of approximately 2 to 4 pp on average, across measures.

***Note*:** This exhibit provides a summary of findings across the 10 patient-reported measures of care experiences and satisfaction with care from the beneficiary survey. The underlying results are based on a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. A solid red square represents generally unfavorable results across the measures, a solid green triangle represents generally favorable results across the measures, and the grey outlined circles represent mixed results or generally no differences between BPCI Advanced and comparison respondents. ADI = Area Deprivation Index; PGP = physician group practice; pp = percentage point(s).

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

C. Discussion and Conclusion

The BPCI Advanced Model tests whether linking Medicare FFS payments for an episode of care can reduce Medicare expenditures while improving quality of care. In Model Year 3, the model continued to result in statistically significant reductions in average episode payments, which were again mostly driven by reductions in institutionalized PAC payments. As in Model Years 1 and 2, the reduction in average episode payments was over twice as large for surgical clinical episodes as it was for medical clinical episodes. There was evidence of an improvement in quality of care in

Model Year 3. There was a small reduction in the unplanned readmission rate for all clinical episodes in Model Year 3, which was mostly driven by the decline for PGP medical clinical episodes. PGP medical clinical episodes also had a small reduction in the mortality rate.

In Model Year 4, however, performance on quality of care from the beneficiary survey was mixed. BPCI Advanced survey respondents were slightly less likely to report improvement in functional status than comparison respondents and slightly less likely to report favorable care experiences and the highest levels of satisfaction with care. These are the first unfavorable patient-reported functional status results for the model. Results from Model Year 2 (2019) were neutral, and multiple years of survey results from the evaluation of the BPCI Initiative Model 2 showed no evidence of unfavorable functional status results among BPCI respondents relative to comparison respondents. There were changes made to the BPCI Advanced Model in Model Year 4, including the requirement to participate in broad CESLGs rather than individual clinical episodes, which could have contributed to the unfavorable results. We will field additional surveys in Model Year 5 (2022). The results will provide additional evidence on the impact of the model on functional status, care experiences, and satisfaction with care and will be reported in the Fifth Evaluation Report.

As in Model Years 1 and 2, BPCI Advanced resulted in savings to Medicare for surgical clinical episodes ($71.3 million, 2.3%), which were offset by losses from medical clinical episodes (-$200.5 million, -1.9%). Target prices may have been too high for medical clinical episodes but were generally more accurate for surgical clinical episodes in Model Years 1 and 2 and Model Year 3. The target price methodology remained unchanged in Model Years 1 through 3 but changed in Model Year 4. Therefore, estimates of savings may differ for Model Year 4 and will be reported in the Fifth Evaluation Report.

In 2021, CMMI launched a strategy refresh which seeks to promote health equity, and it announced that all new models will be designed to reduce inequities in health care outcomes. We have begun to assess health equity for the first time in this evaluation report. Because the model was not designed to address health equity, we assess health equity as a potential unintended consequence of the model. We found that during Model Years 1 through 3, the representation of Black or African American beneficiaries and of dual-eligible beneficiaries was higher in medical clinical episodes and lower in surgical clinical episodes than their representations in the FFS beneficiary population. The differences in representation are features of the underlying FFS healthcare system and are not caused by features of the BPCI Advanced Model.

In Model Years 1 and 2 and Model Year 3, there were larger relative declines in episode payments for the two underserved populations evaluated, Black or African American beneficiaries and dual-eligible beneficiaries, in both medical and surgical clinical episodes. Overall, the impact on quality as measured by readmission and mortality rates was small. In Model Years 1 and 2, the readmission rate declined for dual-eligible beneficiaries with surgical clinical episodes relative to their comparison group. In Model Year 3, there were differential increases in mortality rates for Black or African American beneficiaries and for dual-eligible beneficiaries with medical clinical episodes compared to their reference populations, suggesting that Black or African American beneficiaries and dual-eligible beneficiaries were less likely than their reference populations to benefit from the BPCI Advanced Model during Model Year 3.

We also analyzed responses to the Model Year 4 beneficiary survey for historically underserved populations. We found evidence of unfavorable impacts on functional status for dual-eligible BPCI Advanced respondents relative to their comparison group and no pattern for nondual-eligible BPCI Advanced respondents, resulting in unfavorable impacts on dual-eligible respondents compared to nondual-eligible respondents. For the other historically underserved populations analyzed, there was generally no pattern of favorable or unfavorable changes in functional status, though there was some evidence of impacts on care experiences. BPCI Advanced respondents with hospital-initiated episodes in some historically underserved populations were more likely to report unfavorable care experiences relative to their counterparts in the comparison group in Model Year 4 (Hispanic beneficiaries, those who lived in rural ZIP codes, and those who lived in high-ADI ZIP codes), and BPCI Advanced had unfavorable differential impacts for these respondents compared to their reference groups. Black or African American BPCI Advanced respondents and Hispanic BPCI Advanced respondents with hospital-initiated episodes were 7 to 8 pp less likely than comparison respondents to agree that medical staff took their preferences into account in deciding what services they should receive after leaving the hospital. In contrast, rural beneficiaries with BPCI Advanced PGP-initiated episodes were more likely to report favorable care experiences and the highest levels of satisfaction with post-discharge care in Model Year 4 relative to the comparison group. This is the first time we analyzed survey outcomes under BPCI Advanced for historically underserved populations. Analyses of survey responses from Model Year 5 will be presented in the next evaluation report and will provide additional evidence on the impact of the model on historically underserved populations.

The analyses in this report cover Model Years 3 and 4, 2020 and 2021, two years that were deeply impacted by the COVID-19 PHE. The PHE had widespread effects on the entire health system that may have made it more difficult for participants to focus on care redesign efforts, and that may not be adequately captured in our analyses.

Prior to the PHE, CMS implemented multiple changes to the BPCI Advanced Model for Model Year 4 including changes to the target pricing methodology and requiring participants to select CESLGs instead of individual clinical episodes. These changes are intended to result in savings for the Medicare program. In Model Year 6, CMS implemented additional changes intended to further improve the pricing methodology and keep providers engaged in value-based care, such as a reduction in the CMS discount from 3% to 2% for medical episodes and a reduction in the peer group trend factor adjustment cap from 10% to 5% for all episodes. Future evaluation reports will assess how the changes for Model Year 4 impacted participation in the model and how changes to the target price methodology in Model Year 4 and Model Year 6 impact savings for the Medicare program. In addition, we will continue to evaluate the impact of the model on outcomes (including payments, utilization, and quality of care), health equity, and participant and beneficiary perspectives.

I. Introduction

The Bundled Payments for Care Improvement Advanced (BPCI Advanced) Model is designed to test whether linking Medicare payments for an episode of care can reduce Medicare expenditures while improving or maintaining quality of care. The Center for Medicare & Medicaid Innovation (CMMI) in the Centers for Medicare & Medicaid Services (CMS) launched BPCI Advanced, an Advanced Alternative Payment Model (Advanced APM), in October 2018 and the model will continue through December 2025 including the two-year extension that CMS announced on October 13, 2022.[2]

The Lewin Group, with our partners Abt Associates, Inc., GDIT, and Telligen, is under contract with CMS to conduct an independent evaluation of the impact of the BPCI Advanced Model. This fourth annual report evaluates the impact of the model in Model Year 3 (2020), with additional analyses of Model Years 1 and 2 (2018 and 2019) as well as Model Year 4 (2021). The report presents the impact of the model on payments, utilization, and quality of care for Medicare fee-for-service (FFS) beneficiaries in Model Year 3; differences in patient-reported functional status and care experiences for Medicare FFS beneficiaries in Model Year 4; Medicare program savings in Model Year 3; and findings from analyses of beneficiaries from populations that have been historically underserved in Model Years 1 and 2, Model Year 3, and Model Year 4.

A. The BPCI Advanced Model

BPCI Advanced is a voluntary model in which participants enter into agreements with CMS to be held accountable for total Medicare FFS payments and quality of health care services for a beneficiary during an episode of care included in the list of BPCI Advanced clinical episodes.[3] If total payments for a participant's chosen clinical episode are below its target price, the participant may receive reconciliation payments from CMS. Conversely, if total payments are above its target price, the participant may owe reconciliation payments to CMS.[4] Reconciliation payments are also adjusted by the episode initiator's performance on quality measures. Thus, participants have financial incentives to ensure efficient, coordinated care delivery throughout the entire episode, which begins with a triggering hospitalization or outpatient procedure and ends 90 days after discharge or completion of the procedure. BPCI Advanced is based on the Bundled Payments for Care Improvement (BPCI) Initiative, one of CMMI's previous bundled payment approaches, which was comprised of four models.[5] BPCI Advanced is similar to BPCI Model 2 and incorporates lessons learned.[6] Exhibit 1 highlights key components of the model.

[2] See **Appendix A** for a glossary of terms and abbreviations used in this report.

[3] In Model Years 1 through 3 (2018 through 2020), participants chose to participate in one or more clinical episodes. Beginning in Model Year 4 (2021), CMS required participants to choose one or more of eight clinical episode service line groups (CESLGs), groupings of clinical episodes.

[4] See the CMS BPCI Advanced website for additional information on the reconciliation specifications: https://innovation.cms.gov/media/document/bpciadvanced-my1-2-reconcilation-specs and https://innovation.cms.gov/media/document/bpcia-my3-reconcilation-specs-rev-nov2021.

[5] BPCI Model 1 began in April 2013, and the final Awardee concluded its participation on December 31, 2016. BPCI Models 2, 3, and 4 began in October 2013 and the initiative ended on September 30, 2018.

[6] See the CMS BPCI website for additional information on the initiative and evaluation reports: https://innovation.cms.gov/innovation-models/bundled-payments.

Exhibit 1: Key Components of BPCI Advanced

Defining Characteristics of the Model

- Voluntary, Advanced Alternative Payment Model (APM)
- CMS calculates payment reconciliation by comparing the aggregate Medicare fee-for-service allowed amounts for episodes attributed to a participant against the target price for those episodes, which determines whether the participant receives a payment from CMS or makes a repayment to CMS
- Hospitals and physician group practices (PGPs) can initiate episodes as episode initiators
- Includes 30 inpatient, 3 outpatient, and 1 multi-setting clinical episodes as of Model Year 3
- Participants are required to participate in clinical episode service line groups rather than individual clinical episodes beginning with Model Year 4

Target Prices

- Preliminary target prices were made available to applicants before they made participation decisions
- Hospital target prices are based on hospital historical payments, patient case mix, peer group historical payments, and a prospective peer group trend factor, and are discounted by 3%
- PGP target prices are hospital target prices adjusted for the PGP's specific patient case mix and differences between PGP and hospital historical payments
- Beginning in Model Year 4, final target prices reflect a realized peer group trend (capped at 10%), and PGP target prices are no longer adjusted for differences between PGP and hospital historical payments
- Beginning in Model Year 6, target prices for medical episodes will be discounted by 2% and the peer group trend factor adjustment will be capped at 5% for all episodes

Entry and Withdrawal Rules

- Participants could join the model at the start of Model Year 1 (October 1, 2018) or Model Year 3 (January 1, 2020)
- Model Year 3 was initially the last enrollment opportunity, but CMS announced a two-year extension to the BPCI Advanced Model, with an opportunity for new participants to join in Model Year 7 (January 1, 2024)
- Participants can terminate participation in the model with 90-days advance written notice
- CMS may terminate participants that do not meet the requirements of the participation agreement
- In response to the COVID-19 public health emergency, participants could retrospectively opt out of reconciliation or exclude episodes with a COVID-19 diagnosis from reconciliation in Model Year 3
- In Model Year 4 (2021), CMS excluded all episodes with a COVID-19 diagnosis from reconciliation
- In Model Year 5 (2022), participants could retrospectively choose whether to exclude episodes with a COVID-19 diagnosis from reconciliation
- Beginning in Model Year 6 (2023), all episodes with a COVID-19 diagnosis will be included in reconciliation for all participants

Source: Centers for Medicare & Medicaid Services (2020, May 5). BPCI Advanced. Retrieved from https://innovation.cms.gov/initiatives/bpci-advanced; Centers for Medicare & Medicaid Services (2021, June). Model Overview Fact Sheet – Model Year 3 (MY3). Retrieved from https://innovation.cms.gov/files/fact-sheet/bpciadvanced-my3-modeloverviewfs.pdf; Centers for Medicare & Medicaid Services (2019, September 14). Bundled Payments for Care Improvement Advanced Amended and Restated Participation Agreement. Retrieved from the BPCI Advanced Participant Portal; Centers for Medicare & Medicaid Services (2022, October 13). Model Extension and Changes for Model Year 6 (2023) Fact Sheet. Retrieved from https://innovation.cms.gov/media/document/bpcia-model-ext-and-changes-fs-my6.

1. Participants and Episode Initiators

Each BPCI Advanced participant, which may be a hospital, PGP, or other eligible entity, enters into an agreement with CMS to be held accountable for performance on quality measures and episode payments relative to their target prices. Participants are expected to coordinate care across the providers involved in an episode to reduce payments and improve the quality of patient care.

Participants are either a convener participant (convener) or a non-convener participant. A convener has at least one downstream episode initiator, which is a hospital or a PGP. A convener bears financial risk on behalf of its downstream episode initiators and often provides services (e.g., data analysis, guidance on clinical episode selection, or case management services) intended to help episode initiators succeed in the model. A non-convener participant is a hospital or PGP episode initiator that bears financial risk only for itself. A convener may have multiple participation agreements with CMS for different downstream episode initiators, but as a single episode initiator, non-convener participants can only have one agreement with CMS.

In Model Year 3 (2020), there were 694 unique participants in BPCI Advanced, up from 334 in Model Years 1 and 2 (2018 and 2019).[7] There were 1,010 hospital episode initiators and 1,031 PGP episode initiators, with 70% of episode initiators participating as downstream episode initiators under a convener. By the end of Model Year 3, 256 participants were no longer participating in the model. This resulted in 438 unique participants in BPCI Advanced in Model Year 4 (2021), including 682 hospital episode initiators and 523 PGP episode initiators.

2. BPCI Advanced Episodes

A BPCI Advanced episode begins with a hospitalization or procedure at a participating hospital episode initiator or when the attending or operating clinician for the hospitalization or procedure is a member of a participating PGP episode initiator. Inpatient episodes start when a Medicare beneficiary is admitted to a hospital (anchor stay), and the resulting Medicare Severity-Diagnosis Related Group (MS-DRG) is in one of the participating episode initiator's selected clinical episodes or CESLGs. Outpatient episodes begin when a beneficiary has an outpatient procedure (anchor procedure) in a hospital outpatient setting that is identified by a Healthcare Common Procedure Coding System (HCPCS) code in the participating episode initiator's selected clinical episodes or CESLGs. All FFS Medicare-covered items and professional services, with certain exclusions, furnished during the anchor stay (or the anchor procedure) from the day prior to admission through 89 days after discharge are included in the episode.[8]

Approximately three quarters (74%) of Model Year 3 BPCI Advanced episodes were in a medical clinical episode (Exhibit 2). Most BPCI Advanced medical episodes were initiated by hospital episode initiators (71%), while PGPs initiated a larger share of surgical episodes (62%). See **Appendix B** for a list of clinical episodes by type.

[7] "Unique participants" refers to unique entities which entered into participation agreements with CMS. For example, in Model Year 3, there were 1,707 participants (participation agreements) from 694 unique participants (unique entities).

[8] The lists of exclusions are available for download from the participant resources page on the BPCI Advanced website at https://innovation.cms.gov/innovation-models/bpci-advanced/participant-resources.

Exhibit 2: BPCI Advanced Volume by Clinical Episode Type, Model Year 3 (2020)

Clinical Episode Type	Number of Episode Initiators	Number of Episodes	Percent of All Clinical Episodes
All Clinical Episodes	1,502	353,609	100%
Medical Clinical Episodes	1,155	262,828	74%
Hospitals	910	187,094	53%
PGPs	245	75,734	21%
Surgical Clinical Episodes	870	90,781	26%
Hospitals	473	34,386	10%
PGPs	397	56,395	16%

Note: The number of episodes presented for "All Clinical Episodes" includes all 34 BPCI Advanced clinical episodes. These numbers may not align with numbers presented in other tables and exhibits in the report due to different data sources. See **Appendix C** for additional information. PGP = physician group practice.

Source: CMS reconciliation data for BPCI Advanced hospitals and PGPs from Model Year 3. Second True-Up for Performance Period 3, 4, and 5.

In Model Year 3 (2020), there were changes to the clinical episodes that were included in the model. Inpatient clinical episodes for *bariatric surgery*, *inflammatory bowel disease*, and *seizures* were added. The three individual inpatient clinical episodes for spinal fusion—*cervical spinal fusion*, *combined anterior posterior spinal fusion*, and *spinal fusion (non-cervical)*—were combined into a single *spinal fusion* clinical episode. An inpatient clinical episode for *transcatheter aortic valve replacement* was carved out of the original inpatient *cardiac valve* clinical episode. Further, the inpatient *major joint replacement of the lower extremity* (*MJRLE*) clinical episode was expanded to be a multi-setting clinical episode and includes total knee arthroplasty (TKA) procedures performed in the hospital outpatient department in addition to inpatient procedures.

In Model Year 4, the previous 34 clinical episodes were grouped into eight CESLGs—*cardiac care*, *cardiac procedures*, *gastrointestinal surgery*, *gastrointestinal care*, *neurological care*, *medical and critical care*, *spinal procedures*, and *orthopedics*. CMS requires that participants select CESLGs and participate in all clinical episodes within them. However, participants are not eligible to participate in clinical episodes within selected CESLGs that do not meet the minimum volume threshold during the baseline period. Shifting to CESLGs is intended to encourage participants to broaden care redesign efforts to a wider range of conditions and limit their ability to participate only in clinical episodes that are most financially advantageous to them. See **Appendix B** for a list of the BPCI Advanced clinical episodes, CESLGs, and associated MS-DRGs and HCPCS codes.

3. Target Prices and Reconciliation

CMS calculates a BPCI Advanced target price for each episode initiator and clinical episode combination. A hospital episode initiator's target price reflects its historical Medicare FFS episode payments during the baseline period, adjusted for its patient mix and its payments relative to national historical payments, which are updated based on the spending trends of its hospital peers. A PGP episode initiator's target price is based on the target price of the hospital where the hospitalization or procedure occurred, adjusted for the PGP's patient case mix and

historical spending patterns. Because a PGP may initiate episodes in different hospitals, it may have different target prices for the same clinical episodes, depending on where the episode was initiated. Target prices incorporate a 3% discount, which is intended to be Medicare savings under the model.[9]

The target price calculation method was designed to support participation from a broad range of providers by accounting for variation in episode payments and factors that contribute to payment differences that are beyond providers' control. The use of hospital-specific historical payments, adjusted for peer group levels, peer group trends, and patient mix, is to encourage participation from a variety of providers, including those with historically high and those with historically low episode payments. The peer adjustments recognize that underlying costs and episode spending trends differ across types of hospitals in different circumstances.[10] The patient case-mix adjustment accounts for variations in payments due to differences in patient needs.

The BPCI Advanced Model is an Advanced APM, in part because participant performance on quality measures is factored into the determination of reconciliation payments. BPCI Advanced incorporates seven claims-based quality measures to calculate each episode initiator's Composite Quality Score (CQS) (Exhibit 3).[11,12]

Exhibit 3: BPCI Advanced Quality Measures for Model Years 1 through 3 (2018-2020)

Measure	Applicable Clinical Episodes
All-Cause Hospital Readmission Measure	All clinical episodes
Advance Care Plan	All clinical episodes
CMS Patient Safety Indicators (CMS PSI 90)	All inpatient clinical episodes, as well as outpatient procedures for major joint replacement of the lower extremity
Hospital-Level Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA)	Double joint replacement of the lower extremity; major joint replacement of the lower extremity (inpatient and outpatient)
Hospital 30-Day, All-Cause, Risk-Standardized Mortality Rate Following Coronary Artery Bypass Graft Surgery	Coronary artery bypass graft
Excess Days in Acute Care after Hospitalization for Acute Myocardial Infarction	Acute myocardial infarction

[9] Beginning in Model Year 6, target prices for medical episodes will incorporate at 2% discount.
[10] Centers for Medicare & Medicaid Services (June 2018). Pricing Methodology for Clinicians and Administrators. Retrieved from https://innovation.cms.gov/Files/slides/bpciadvanced-wc-pricingmethodology-clinadmin.pdf.
[11] An additional set of 23 alternate quality measures, including claims-based and registry-based measures, were available for participants to select for each clinical episode within a clinical episode service line group in Model Year 4.
[12] More information about BPCI Advanced quality measures is available at https://innovation.cms.gov/innovation-models/bpci-advanced/quality-measures-fact-sheets.

Measure	Applicable Clinical Episodes
Perioperative Care: Selection of Prophylactic Antibiotic: First- or Second-Generation Cephalosporin	Back and neck except spinal fusion (inpatient and outpatient); bariatric surgery; coronary artery bypass graft; cardiac valve; double joint replacement of the lower extremity; hip and femur procedures except major joint; lower extremity and humerus procedures except hip, foot, femur; major bowel procedure; major joint replacement of the lower extremity (inpatient and outpatient); major joint replacement of the upper extremity; spinal fusion

Source: Centers for Medicare & Medicaid Services (n.d.). Quality Measures Correlation to Clinical Episodes Model Years 1 through 3. Retrieved from https://innovation.cms.gov/files/x/bpci-advanced-qualmsrcorrclinepi-modelyrs1-3.pdf.

Under the model, providers and suppliers continue to receive Medicare FFS payments for providing Medicare-covered items and services. At the end of each six-month performance period, CMS compares Medicare payments during the episode with the target price for each episode initiator for each of its clinical episodes. When the episode payments for a participant, aggregated across all of its episode initiators and clinical episodes, are below its target amount, the participant will receive a Net Payment Reconciliation Amount (NPRA). When the aggregated episode payments are above the target amount, the participant will owe a repayment amount to CMS. The NPRA or repayment amount includes adjustments for the episode initiator's CQS and for the stop-loss or stop-gain limits of the BPCI Advanced Model.[13,14] Throughout the report, we refer to the NPRA or repayment amounts collectively as "reconciliation payments."

4. Participation and Clinical Episode Selection

BPCI Advanced participants self-selected to participate in the model. To help inform their decision to participate, prospective participants received historical data and preliminary target prices to review and assess their potential success in the model within specific clinical episodes. Prior to the start of Model Years 1 (2018) and 3 (2020), episode initiators could select one or more clinical episodes if the minimum episode thresholds were met. Prior to the start of Model Year 4 (2021), episode initiators could select one or more CESLGs and were not permitted to change their selections in Model Year 5 (2022) or Model Year 6 (2023).

In response to the COVID-19 public health emergency (PHE), CMS allowed BPCI Advanced participants to alter their financial risk for episodes beginning on or after January 1, 2020 and ending by December 31, 2020. Any participant that had not withdrawn from the model before June 28, 2020 could select one of two BPCI Advanced participation agreement amendments. Amendment 1 allowed the participant to forgo reconciliation for all episodes in Model Year 3 (2020), such that the participant would not earn or owe reconciliation payments, but they could continue participation. Amendment 2 allowed the participant to exclude episodes with a COVID-19 diagnosis from reconciliation in Model Year 3. Episodes from participants that did not elect either amendment were subject to the usual reconciliation process of the BPCI Advanced Model. As reported in the Third Evaluation Report, about a quarter (26.1%) of participants selected Amendment 1, to opt out of reconciliation entirely.[15] More than half of participants (57.4%) chose

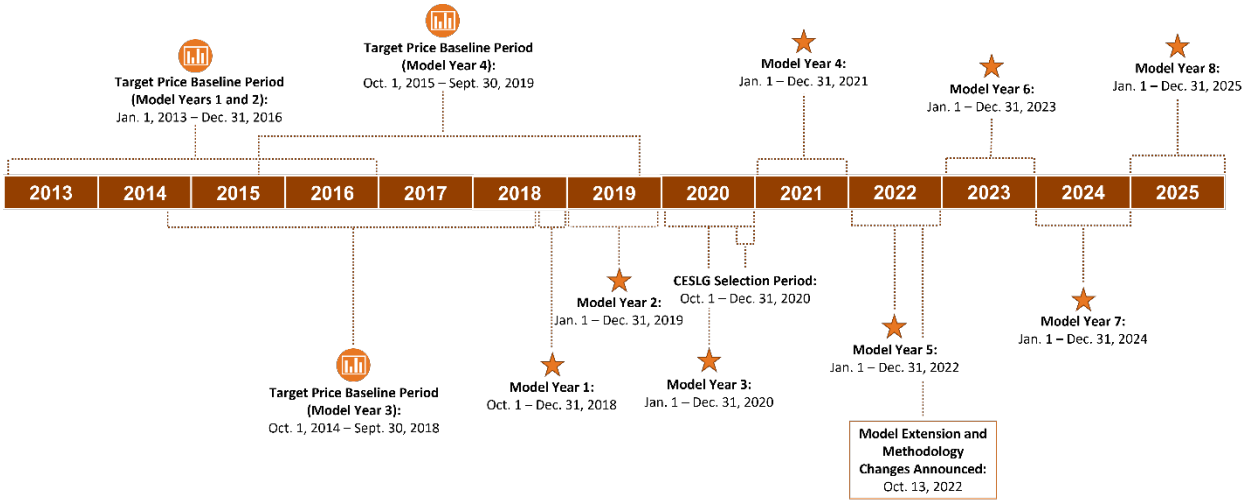
[13] The reconciliation amount has a 20% stop loss/gain applied at the episode initiator level.
[14] The CQS adjustment amount cannot change the NPRA or repayment amount by more than 10%.
[15] The BPCI Advanced Third Evaluation Report is available for download at https://innovation.cms.gov/innovation-models/bpci-advanced.

Amendment 2, to exclude episodes in 2020 with a COVID-19 diagnosis from reconciliation. The remainder (16.5%) did not select either amendment. CMS later issued a unilateral amendment that all episodes with a COVID-19 diagnosis that were initiated in 2020 and reconciled in Performance Period 5 (ending June 30, 2021) would be excluded from reconciliation for those that chose Amendment 1 and those that did not elect either amendment.[16] CMS allowed this policy to continue through the end of 2022, by incorporating it into the Model Year 4 and Model Year 5 BPCI Advanced participation agreement.[17]

5. Model Timeline

The BPCI Advanced Model extends for more than seven calendar years: Model Year 1 began October 1, 2018, and Model Year 8 ends December 31, 2025 (Exhibit 4). Participants and episode initiators had two opportunities to join the model, with a third opportunity in the future. The first cohort of participants began at the start of Model Year 1 (October 2018). The second cohort began at the start of Model Year 3 (January 2020). The third cohort will begin at the start of Model Year 7 (January 2024). The target prices for Model Years 1 and 2 (2018 and 2019) were based on historical payments from January 2013 through December 2016 (target price baseline period). In Model Year 3 (2020), the target price baseline period was October 1, 2014 through September 30, 2018. In Model Year 4 (2021), the target price baseline period was October 1, 2015 through September 30, 2019, which means target prices incorporated payments for episodes that occurred after the model began. The baseline period will continue to shift forward for future model years. In Model Year 4 (2021), participants were required to participate in CESLGs, though no new participants nor episode initiators were allowed to join the model. These CESLG selections are binding until the end of Model Year 6.

Exhibit 4: BPCI Advanced Timeline Through Model Year 8



Note: CESLG = clinical episode service line group.

[16] The unilateral amendment was not applicable to participants that signed Amendment 2 since Amendment 2 already excluded COVID-19 clinical episodes from reconciliation until June 30, 2021.

[17] In August 2022, during Model Year 5, CMS offered a voluntary bilateral amendment that would allow participants to be held accountable for COVID-19 episodes in Model Year 5. Beginning in Model Year 6, all participants will be held accountable for all episodes, regardless of COVID-19 diagnosis.

Source: Centers for Medicare & Medicaid Services (2018, April). BPCI Advanced Model Timeline. Retrieved from https://innovation.cms.gov/Files/x/bpci-advanced-timeline.pdf and Centers for Medicare & Medicaid Services. Pricing Methodology: Frequently Asked Questions (FAQ). Retrieved from https://innovation.cms.gov/Files/x/bpciadvanced-my3-pm-faqs.pdf.

B. Research Questions

As in prior evaluation reports, this report presents the impact of the model on payments, utilization, and quality of care for Medicare FFS beneficiaries (Model Year 3, 2020); differences in patient-reported functional status and care experiences for Medicare FFS beneficiaries (Model Year 4, 2021); and Medicare program savings (Model Year 3). Finally, for the first time in this report, we provide results of exploratory analyses of outcomes under BPCI Advanced for beneficiaries from populations that have been historically underserved. Populations analyzed in this report include Black or African American beneficiaries, Hispanic beneficiaries, dual-eligible beneficiaries, beneficiaries living in rural areas, and beneficiaries living in areas with a high Area Deprivation Index (ADI) (see **Appendix C** for details about these definitions).

Four major research questions provided the framework for our analytic approach for this report.

Research Questions

- What is the impact of BPCI Advanced on episode payments, utilization, and quality of care in Model Year 3?
- Were there differences in patient-reported functional status, care experiences, and overall satisfaction with care between BPCI Advanced and comparison group respondents under BPCI Advanced in Model Year 4?
- Did BPCI Advanced result in savings to Medicare during Model Year 3?
- Were there differences in outcomes under the BPCI Advanced Model for beneficiaries from populations that have been historically underserved?

1. What was the impact of BPCI Advanced on episode payments, utilization, and quality during Model Year 3 (January 1, 2020 through December 31, 2020)?

We estimated the impact of BPCI Advanced on episode payments, utilization of services, and claims-based quality of care, measured by readmission and mortality rates, for Medicare FFS beneficiaries. Medicare claims and enrollment data were used to construct episodes for beneficiaries attributed to BPCI Advanced participating episode initiators and to matched comparison providers.

2. Were there differences in patient-reported functional status, care experiences, and overall satisfaction with care under BPCI Advanced in Model Year 4 (January 1, 2021 through December 31, 2021)?

We conducted a beneficiary survey to explore differences in patient care experiences and functional outcomes between Medicare FFS beneficiaries cared for by BPCI Advanced providers and similar beneficiaries whose providers did not participate in BPCI Advanced.

3. Did BPCI Advanced result in savings to Medicare during Model Year 3 (January 1, 2020 through December 31, 2020)?

Net Medicare savings is defined as the change in non-standardized payments minus net reconciliation payments. We evaluated net savings to Medicare due to BPCI Advanced for selected clinical episodes based on the estimated impact of BPCI Advanced on Medicare FFS episode payments, adjusted by reconciliation payments made to or received from model participants. We calculated net Medicare savings (or losses) for all episodes pooled across the clinical episodes evaluated, for medical and surgical clinical episodes evaluated, for medical and surgical clinical episodes evaluated by episode initiator type, and for each clinical episode for which we calculated impact estimates.

4. Were there differences in outcomes under the BPCI Advanced Model for beneficiaries from populations that have been historically underserved?

We estimated changes in episode payments, readmission rates, and mortality rates and differences in patient care experiences and functional outcomes for BPCI Advanced Medicare FFS beneficiaries from select historically underserved populations relative to their counterparts in the comparison group. These populations include Black or African American beneficiaries, Hispanic beneficiaries, beneficiaries that are dually eligible for Medicare and Medicaid, beneficiaries that lived in areas with a high ADI, and beneficiaries that lived in rural areas. We also estimated the differences in outcomes between each underserved population and a reference population.

C. Data Sources and Outcomes

This evaluation relied on multiple primary and secondary data sources. Interviews and site visits with participants provided context for the COVID-19 PHE and quantitative findings. We also conducted a beneficiary survey to explore differences in patient care experiences and functional outcomes between Medicare FFS beneficiaries cared for by BPCI Advanced providers and similar beneficiaries whose providers did not participate in BPCI Advanced. Secondary data sources were used to construct samples, estimate outcomes, and supplement the quantitative results. We used provider-level data sources, including the CMS BPCI Advanced database and Provider of Services (POS) files to identify and describe BPCI Advanced-participating providers and select comparison providers. Medicare claims and enrollment data were used to construct episodes for beneficiaries at BPCI Advanced-participating episode initiators and at matched comparison providers. We also used claims data to create outcome measures and beneficiary risk factors associated with the outcomes. See **Appendix C** for more information on our primary and secondary data sources.

II. Results

This chapter presents results of descriptive analyses of the evaluation sample; the impact of BPCI Advanced on payments, utilization, and quality; changes in functional status, care experiences, and overall satisfaction with care as reported by beneficiaries; Medicare program savings under the model; and analyses of beneficiaries from populations that have been historically underserved. Analyses were performed by pooling episodes across all clinical episodes evaluated, by clinical episode type (medical and surgical), by episode initiator type (hospital and physician group practice (PGP)) (Exhibit 5). Results were grouped by medical and surgical clinical episodes and by hospital and PGP episode initiators because care redesign activities may vary by clinical episode type and episode initiator type, resulting in different impacts on outcomes. Some analyses were also performed by clinical episode or CESLG, and these results are presented in the appendices. For further details of our methodology, data, and sensitivity tests see **Appendix C**.

Exhibit 5: Analyses by Clinical Episode Groupings

Analysis	All Clinical Episodes	Clinical Episode Type	Episode Initiator Type	Clinical Episode by Episode Initiator Type
Impact of the Model on Payments, Utilization, and Quality	•	•		•
Changes in Patient-reported Functional Status, Care Experiences, and Satisfaction with Care			•	•
Medicare Program Savings	•	•		•
Historically Underserved Populations: Analysis of Payments and Quality		•		
Historically Underserved Populations: Analysis of Functional Status, Care Experiences, Satisfaction with Care			•	

***Note*:** Clinical episode types are medical and surgical. Episode initiator types are hospitals and PGPs. Clinical episode by episode initiator type includes medical and surgical episode types by hospitals and PGPs (e.g., medical clinical episodes initiated by hospital episode initiators). Impact analyses and Medicare program savings were also performed by clinical episode. Analyses of patient-reported functional status, care experiences, and satisfaction with care were also performed by CESLG. Results by clinical episode and CESLG are presented in the appendices.

A. Impact of BPCI Advanced

This section presents results of the analyses of the impact of BPCI Advanced on payments, utilization, and quality of care measured using claims data in Model Year 3 (2020), as well as the changes in patient-reported measures of functional status, care experiences, and overall satisfaction with care from a beneficiary survey in Model Year 4 (2021).

1. Payment, Utilization, and Claims-based Quality Outcomes

We estimated the impact of the model on claims-based outcomes using a difference-in-differences (DiD) approach in which we compare the change from baseline (April 1, 2013, through December 31, 2017) to intervention (January 1, 2020, through December 31, 2020) for BPCI Advanced episode initiators to the change for matched comparison providers, adjusting for

differences in patient mix, clinical severity, and provider characteristics. We also risk adjusted for confirmed COVID-19 diagnosis during the 90 days preceding the anchor stay or procedure and during the anchor stay or procedure to control for the direct effect of COVID-19 on evaluated outcomes. Because there were geographic and temporal differences in the spread and burden of the COVID-19 public health emergency (PHE), we included measures of COVID-19 incidence and deaths in the county of the anchor stay or procedure (see **Appendix C** for more information). It is possible that we did not fully capture the effects of the PHE through risk adjustment.

A key assumption to estimating the change in outcomes due to the model using a DiD approach is that outcomes for BPCI Advanced and comparison episodes had parallel trends during the baseline period. We tested the parallel trends assumption and found that some outcomes for some clinical episode groupings did not have parallel trends during the baseline. In those cases, the DiD estimates may be biased (see **Appendix C** for more information on our parallel trends testing methodology). In the text and exhibits, we note the results that failed tests of parallel trends. Though some of the individual DiD estimates may be biased, we maintain that the portfolio of results provides meaningful information about the impact of BPCI Advanced and how episode initiators achieved reductions in per-episode payments. For more information, see the discussion of parallel trends failures in Chapter III. Discussion and Conclusion. See **Appendix G** for the parallel trends results.

a. Key Findings

Impact of BPCI Advanced

- During Model Year 3, pooling across all clinical episodes analyzed, the BPCI Advanced Model reduced average standardized episode payments by $1,028 per episode, or 3.8% of the baseline mean, relative to the comparison group.
  - The reduction in per-episode payments was larger for surgical clinical episodes than medical clinical episodes (-$796 or -3.1% for medical clinical episodes vs. -$1,800 or -5.8% for surgical clinical episodes).
  - The reductions in per-episode payments for medical clinical episodes were similar for hospital episode initiators and physician group practice (PGP) episode initiators (-$756 or -2.9% for hospital episode initiators and -$667 or -2.7% for PGP episode initiators), but the reduction in payments for surgical episode initiators was over twice as large for PGP episode initiators than for hospital episode initiators (-$933 or -3.0% for hospital episode initiators compared to -$2,147 or -6.9% for PGP episode initiators).
- Reductions in payments were driven by reductions in skilled nursing facility (SNF) and inpatient rehabilitation facility (IRF) payments for both medical and surgical clinical episodes. For medical clinical episode payments, SNF was a larger driver (-$452 for SNF payments compared to -$109 for IRF payments). For surgical episode payments, IRF was the larger driver (-$533 for SNF payments compared to -$754 for IRF payments).
- For medical clinical episodes, hospital and PGP episode initiators reduced the share of episodes first discharged to an institutional post-acute care (PAC) setting by a similar amount (-0.90 pp for hospital medical clinical episodes compared to -0.86 pp for PGP medical clinical episodes). For surgical clinical episodes, the reduction in the share of institutional PAC discharges was larger for PGP episode initiators than hospital episode initiators, and the impact for hospital episode initiators was not statistically significant (-2.65 pp for hospital surgical clinical episodes and -5.71 pp for PGP surgical clinical episodes).
- Both hospital and PGP episode initiators reduced SNF days among episodes with at least one day of SNF use. For medical clinical episodes, the reduction was larger for hospital episode initiators than PGP episode initiators, and the impact for PGP episode initiators was not statistically significant (-2.24 days for hospital medical clinical episodes and -0.47 days for PGP medical clinical episodes). The decline was similar for hospital and PGP episode initiators for surgical clinical episodes (-3.28 days for hospital surgical clinical episodes and -3.19 days for PGP surgical clinical episodes.
- There was a reduction in the unplanned readmission rate (-0.83 pp or -2.8%) and the mortality rate (-0.64 pp or -3.4%) for PGP medical clinical episodes.

b. Sample

Participants and episode initiators voluntarily chose to participate in BPCI Advanced, and they could select one or more clinical episodes (out of 34) in which to participate in Model Year 3.[18] The evaluation included a subset of clinical episodes that had sufficient sample size for analysis. We constructed comparison groups for 17 hospital clinical episodes and for 17 PGP clinical episodes. (The clinical episodes evaluated are not the same for hospitals and PGPs.) Of the 17 hospital clinical episodes evaluated, 10 were medical and 7 were surgical. Of the 17 PGP clinical episodes evaluated, 12 were medical and 5 were surgical. In Model Year 3, there were 2,041 episode initiators, including 1,010 hospital episode initiators and 1,031 PGP episode initiators. To be included in our analysis, hospitals and PGPs had to meet the requirements for evaluation, such as having episode volume in the baseline and in Model Year 3, resulting in a sample of 1,539 episode initiators that were eligible for selecting comparison groups (Exhibit 6). Of these, 1,507 BPCI Advanced episode initiators were included in the final matched evaluation sample. See **Appendix E** for sample characteristics by clinical episode. See **Appendix C** for information on the methods used to determine the sample.

Exhibit 6: Matched BPCI Advanced Hospital and PGP Episode Initiators Included in the BPCI Advanced Impact Estimates, January 1, 2020 – December 31, 2020

Clinical Episodes	BPCI Advanced Participating Episode Initiators	Matched BPCI Advanced Episode Initiators	Intervention Episodes for Matched BPCI Advanced Episode Initiators
All Clinical Episodes	1,539	1,507	397,627
Medical	1,238	1,216	318,743
Hospitals	883	864	217,564
PGPs	355	352	101,179
Surgical	671	641	78,884
Hospitals	358	345	26,309
PGPs	314	296	52,575

Note: The numbers of BPCI Advanced participating episode initiators are limited to the BPCI Advanced hospitals and PGPs that met the requirements for evaluation, such as having episode volume in the baseline and in the intervention (Model Year 3, 2020). The number of matched BPCI Advanced episode initiators are limited to the BPCI Advanced hospitals and PGPs that were used to calculate the difference-in-differences results in the remainder of this section. The number of matched intervention episodes is based on the sample used to evaluate the impact of the model on total allowed standardized payments. The number of episode initiators in each category may not sum to the total because episode initiators can participate in multiple clinical episodes. See **Appendix C** for information on the methods used to determine the sample. See **Appendix E** for more detailed results. PGP = physician group practice.

Source: CMS BPCI Advanced Database, as of April 15, 2022, and the BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stay or procedure end dates beginning January 1, 2020 and ending on or before December 31, 2020 for BPCI Advanced hospitals and PGPs.

The clinical episodes evaluated represent 92.1% of total BPCI Advanced intervention volume, or, when grouped into medical and surgical clinical episode types, 96.4% of episodes initiated under medical clinical episodes and 78.9% of episodes initiated under surgical clinical episodes. After matching BPCI Advanced episode initiators to comparison hospitals and PGPs, our evaluation sample included 80.9% of BPCI Advanced episodes initiated in the clinical episodes evaluated. When grouped into medical and surgical clinical episode types, our evaluation sample included

[18] CMS required that hospital episode initiators have at least 41 episodes in the baseline period for each clinical episode in which they selected to participate.

82.2% of episodes initiated in the medical clinical episodes evaluated and 75.9% of episodes initiated in the surgical clinical episodes evaluated. See **Appendix C** for additional details on the sample and comparison group selection.

Because this report estimates the impact of the model after the onset of the COVID-19 PHE, we compared two key COVID-19 statistics for BPCI Advanced and comparison episodes, and we found that the two groups were similar on both measures. We examined the rates of confirmed COVID-19 diagnoses during the episode (including the inpatient stay or day of the outpatient procedure and 90 days after) and the seven-day moving average of the number of new COVID-19 cases per 100,000 residents on the day of the anchor hospital admission or procedure in the county where the hospital is located.[19] County-level average new COVID-19 cases were generally similar across groups, across hospitals and PGPs, and across clinical episode types, but there were differences in the rates of confirmed COVID-19 diagnoses during the episode (Exhibit 7). Approximately 21% of medical episodes had a confirmed COVID-19 diagnosis, while approximately 4% of surgical episodes had a confirmed COVID-19 diagnosis.

Exhibit 7: Share of Episodes with Confirmed COVID-19 Diagnosis and County-level Seven-day Moving Average of New COVID-19 Cases per 100,000, January 1, 2020 – December 31, 2020

Clinical Episodes	Episodes with Confirmed COVID-19 Diagnosis		County-level Average of New COVID-19 Cases per 100,000	
	BPCI Advanced Episodes	Comparison Group Episodes	BPCI Advanced Episodes	Comparison Group Episodes
All Clinical Episodes	17.7%	18.0%	16.86	17.20
Medical	21.1%	20.9%	17.19	17.66
Hospitals	21.9%	21.2%	17.39	17.52
PGPs	19.3%	20.1%	16.74	18.06
Surgical	4.3%	4.4%	15.56	15.08
Hospitals	4.7%	4.7%	15.83	15.43
PGPs	4.1%	4.1%	15.42	14.85

Note: The share of episodes with a confirmed COVID-19 diagnosis refers to a confirmed COVID-19 diagnosis using ICD-10 codes B97.29 from January 27, 2020 onward or U07.1 from April 1, 2020 onward during the episode, which includes the inpatient stay or day of the outpatient procedure and the 90 days after. The sample includes episodes used to evaluate the impact of the model on total allowed standardized payments. The county-level average of new COVID-19 cases per 100,000 refers to the seven-day moving average (using data from 3 days before and 3 days after) of number of new cases per 100,000 residents on the day of the anchor hospital admission or procedure in the county where the hospital is located. See **Appendix C** for information on the outcomes. ICD = international classification of diseases; PGP = physician group practice.

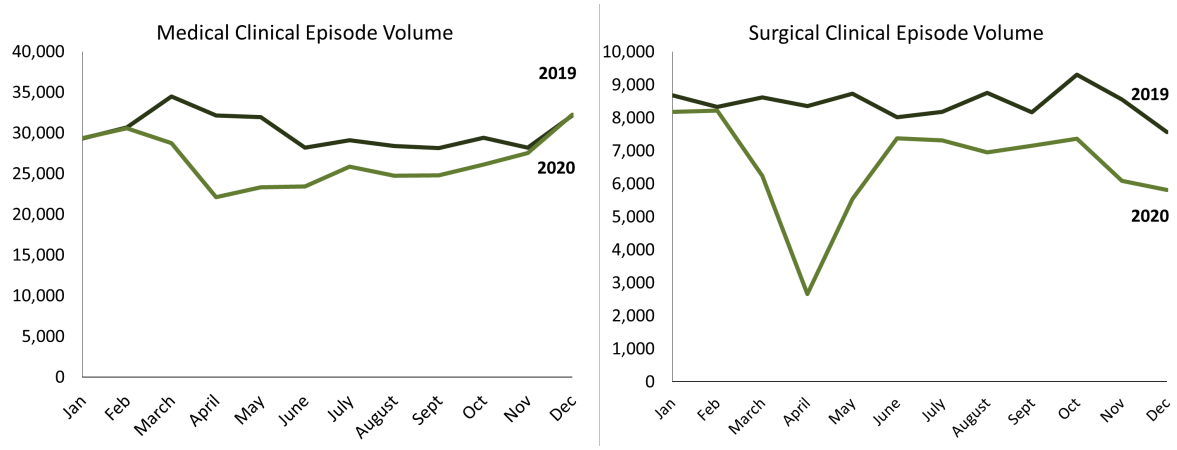
Source: USAFacts and the BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stay or procedure end dates beginning January 1, 2020 and ending on or before December 31, 2020 for BPCI Advanced hospitals and PGPs.

As part of the model evaluation, the team conducted site visits and key informant interviews to learn how participants are responding to the model and how exogenous factors, such as the COVID-19 PHE, impact participation in the model. After the onset of the COVID-19 PHE, episode initiators reported experiencing an overall decrease in patient volume, which they

[19] The seven-day moving average is calculated using data from 3 days before the date through 3 days after the date. For example, the seven-day moving average for March 5 is the average of the data for March 2 through 8.

attributed to patients delaying or avoiding care due to the COVID-19 PHE. BPCI Advanced episode volume was lower in 2020 (after the onset of the COVID-19 PHE) compared to the same months in 2019 for episode initiators that participated in Model Year 3. There were differences by episode type. Medical episode volume declined by 25% from January to April 2020 and recovered slowly, until it reached 2019 levels by December 2020. Surgical episode volume, on the other hand, declined sharply in March and April 2020 and nearly recovered by June, although 2020 levels remained below 2019 levels throughout the remainder of the year (Exhibit 8). The decline in surgical episode volume coincided with CMS recommendations that hospitals delay elective surgeries and non-essential medical and surgical procedures during the COVID-19 PHE.[20]

Exhibit 8: Volume Attributed to Model Year 3 BPCI Advanced Episode Initiators, January 2019 – December 2020



***Source**:* The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures that began on or after January 1, 2019 and ended by December 31, 2020 for BPCI Advanced hospitals and physician group practices that participated in the model during Model Year 3. Sample includes episodes where total allowed standardized payments had been calculated.

c. Have patient characteristics changed under the BPCI Advanced Model from the baseline period to Model Year 3 relative to the comparison group?

Descriptive analyses were conducted on claims-based data to assess changes in average patient mix and racial and ethnic composition between the baseline and intervention period (Model Year 3) for beneficiaries treated by BPCI Advanced hospitals and PGPs relative to the comparison group. The analyses did not reveal widespread changes in patient mix or in the distribution of race and ethnicity. The few changes in patient mix were generally small in magnitude and inconsistent. Some changes provided evidence of a shift toward a relatively more complex patient mix, such as the relatively higher percentage of patients who were disabled or dually eligible for Medicare and Medicaid, while other trends indicated a less complex patient mix, such as the relative decline in the share of beneficiaries with prior institutional post-acute care (PAC) use (Exhibit 9). The differences in racial and ethnic composition were also generally small in magnitude. Only two groups had a relative change that was greater than 1 percentage point (pp). The proportion of Black or African American beneficiaries increased for PGP medical clinical episodes by 2.19 pp while

[20] The recommendations were announced during the White House Task Force Press Briefing on March 18, 2020. The CMS press release is available for download at https://www.cms.gov/newsroom/press-releases/cms-releases-recommendations-adult-elective-surgeries-non-essential-medical-surgical-and-dental.

the share of Hispanic beneficiaries decreased by 1.79 pp (Exhibit 10).[21] Results by clinical episode are presented in **Appendix E**.

These analyses are unadjusted, and thus we cannot conclude whether changes in patient mix and racial and ethnic composition are the result of the BPCI Advanced Model or whether they are the result of other factors.

For the analyses presented below, we have a robust risk-adjustment methodology that includes patient mix characteristics. However, it is possible for patient characteristics to impact outcomes in ways that we are not able to account for in our model specifications.

[21] The analyses of the changes in race and ethnicity rely on the Research Triangle Institute (RTI) race codes from the Master Beneficiary Summary File (MBSF). The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm utilizing beneficiary names and geographical locations. Additional information can be found at https://resdac.org/cms-data/variables/research-triangle-institute-rti-race-code.

Exhibit 9: Patient Mix, Hospital and PGP Episode Initiators, Baseline (April 1, 2013 – October 31, 2017) and Intervention (January 1, 2020 – December 31, 2020)

Clinical Episodes	Patient Characteristics	BPCI Advanced		Comparison Group		Relative Change
		Baseline Mean	MY3 Mean	Baseline Mean	MY3 Mean	
Hospital Medical (N=221,323)	Institutional PAC Use	40.6%	38.1%	40.3%	37.8%	-0.05 pp
	Home Health	29.2%	26.6%	27.8%	25.9%	-0.71 pp
	Age: 80+ Years	45.1%	42.5%	44.7%	42.5%	-0.39 pp
	Male	43.2%	45.5%	43.3%	45.8%	-0.20 pp
	Disabled, No ESRD	26.8%	26.2%	26.9%	25.6%	0.73 pp
	Dual Eligibility	28.2%	27.5%	27.8%	26.7%	0.27 pp
	HCC Count	3.2	3.3	3.2	3.3	-0.02
	HCC Score	2.1	2.1	2.0	2.1	-0.01
PGP Medical (N=102,690)	Institutional PAC Use	36.1%	34.0%	38.1%	36.1%	-0.13 pp
	Home Health	26.1%	24.5%	26.3%	25.1%	-0.40 pp
	Age: 80+ Years	43.8%	41.4%	42.7%	41.4%	-1.14 pp
	Male	44.5%	45.8%	43.6%	45.7%	-0.82 pp
	Disabled, No ESRD	26.1%	25.5%	27.9%	25.6%	1.71 pp
	Dual Eligibility	25.4%	24.1%	25.7%	23.3%	1.13 pp
	HCC Count	3.0	3.1	3.1	3.2	-0.02
	HCC Score	1.9	2.0	2.0	2.0	-0.01
Hospital Surgical (N=26,801)	Institutional PAC Use	19.0%	16.6%	18.8%	17.5%	-0.97 pp
	Home Health	12.6%	11.6%	12.3%	11.6%	-0.29 pp
	Age: 80+ Years	31.1%	31.0%	31.0%	31.1%	-0.16 pp
	Male	47.7%	49.4%	48.6%	49.9%	0.47 pp
	Disabled, No ESRD	19.0%	49.4%	48.6%	49.9%	0.19 pp
	Dual Eligibility	14.0%	11.4%	15.5%	12.9%	0.01 pp
	HCC Count	1.9	2.0	1.9	2.0	-0.04
	HCC Score	1.2	1.2	1.2	1.3	-0.02
PGP Surgical (N=53,574)	Institutional PAC Use	13.9%	12.1%	13.9%	12.0%	0.12 pp
	Home Health	11.4%	10.5%	11.0%	11.0%	-0.95 pp
	Age: 80+ Years	27.6%	27.5%	27.1%	27.4%	-0.33 pp
	Male	36.3%	37.8%	36.9%	38.7%	-0.19 pp
	Disabled, No ESRD	17.4%	15.0%	17.8%	15.6%	-0.19 pp
	Dual Eligibility	11.4%	9.3%	11.3%	9.6%	-0.41 pp
	HCC Count	1.4	1.4	1.4	1.5	0.02
	HCC Score	1.0	1.0	1.0	1.1	0.00

Note: The sample size, N, is the number of episodes for the BPCI Advanced intervention (Model Year 3, 2020). These results are not risk adjusted. The institutional PAC setting and home health variables measure utilization of care in the six months prior to the anchor hospitalization or procedure. HCC count and HCC score are also based on the six months prior to the anchor hospitalization or procedure. See **Appendix C** for information on the patient mix methodology. See **Appendix E** for more detailed results. ESRD = end-stage renal disease. HCC = hierarchical condition categories; MY = model year; PAC = post-acute care; PGP = physician group practice; pp = percentage point(s).

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

Exhibit 10: Beneficiary Race and Ethnicity, Hospital and PGP Episode Initiators, Baseline (April 1, 2013 – October 31, 2017) and Intervention (January 1, 2020 – December 31, 2020)

Clinical Episodes	Beneficiary Race Categories	BPCI Advanced		Comparison Group		Relative Change
		Baseline Mean	MY3 Mean	Baseline Mean	MY3 Mean	
Hospital Medical (N=221,323)	Asian/Pacific Islander	1.8%	2.1%	2.2%	2.5%	-0.14 pp
	Black or African American	12.3%	12.2%	11.9%	11.4%	0.36 pp
	Hispanic	6.1%	6.3%	5.3%	6.1%	-0.66 pp
	American Indian/Alaska Native	0.4%	0.4%	0.4%	0.5%	-0.04 pp
	Other	0.5%	0.6%	0.6%	0.7%	-0.04 pp
	Non-Hispanic White	78.4%	77.5%	79.2%	77.8%	0.54 pp
PGP Medical (N=102,690)	Asian/Pacific Islander	1.9%	1.9%	1.1%	1.4%	-0.34 pp
	Black or African American	7.9%	9.1%	9.8%	8.9%	2.19 pp
	Hispanic	6.7%	5.7%	3.6%	4.4%	-1.79 pp
	American Indian/Alaska Native	0.5%	0.6%	0.7%	0.9%	-0.09 pp
	Other	0.5%	0.7%	0.5%	0.6%	0.03 pp
	Non-Hispanic White	82.0%	81.2%	83.9%	83.0%	0.07 pp
Hospital Surgical (N=26,801)	Asian/Pacific Islander	1.9%	1.8%	1.4%	1.4%	-0.05 pp
	Black or African American	6.7%	5.1%	6.4%	5.1%	-0.27 pp
	Hispanic	4.1%	3.7%	5.1%	4.7%	-0.07 pp
	American Indian/Alaska Native	0.5%	0.6%	0.6%	0.5%	0.14 pp
	Other	0.6%	0.6%	0.6%	0.6%	-0.11 pp
	Non-Hispanic White	85.2%	86.3%	85.1%	85.9%	0.33 pp
PGP Surgical (N=53,574)	Asian/Pacific Islander	0.8%	0.9%	0.9%	1.0%	0.00 pp
	Black or African American	4.7%	3.8%	4.9%	3.7%	0.29 pp
	Hispanic	3.5%	3.2%	2.7%	2.8%	-0.44 pp
	American Indian/Alaska Native	0.5%	0.5%	0.3%	0.3%	-0.04 pp
	Other	0.5%	0.5%	0.4%	0.5%	0.00 pp
	Non-Hispanic White	89.2%	89.4%	89.8%	89.8%	0.24 pp

Note: The sample size, N is the number of episodes for the BPCI Advanced intervention (Model Year 3, 2020). These results are not risk adjusted. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Master Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. MY = model year; PGP = physician group practice; pp = percentage point(s).

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

d. How have average standardized episode payments changed under BPCI Advanced?

We measured the impact of the BPCI Advanced Model on standardized allowed payments during the episode, calculated as Medicare Parts A and B payments, that included beneficiary cost sharing and were standardized to remove geographic and other payment adjustments. During Model Year 3, the BPCI Advanced Model reduced episode payments from the baseline by $1,028 per episode (90% confidence interval: -$1,205, -$852; $p < 0.01$) or 3.8% of the baseline mean relative to comparison hospitals and PGPs (Exhibit 11). The relative reduction in per-episode payments was twice as large for surgical clinical episodes than it was for medical clinical episodes. For medical clinical episodes, episode payments declined by $796 per episode (90% confidence interval: -$974, -$619; $p < 0.01$) or 3.1%. For surgical clinical episodes, episode payments declined by $1,800 per episode (90% confidence interval: -$2,149, -$1,451; $p < 0.01$) or 5.8%.[22] See **Appendix F** for detailed results and for results by clinical episode.

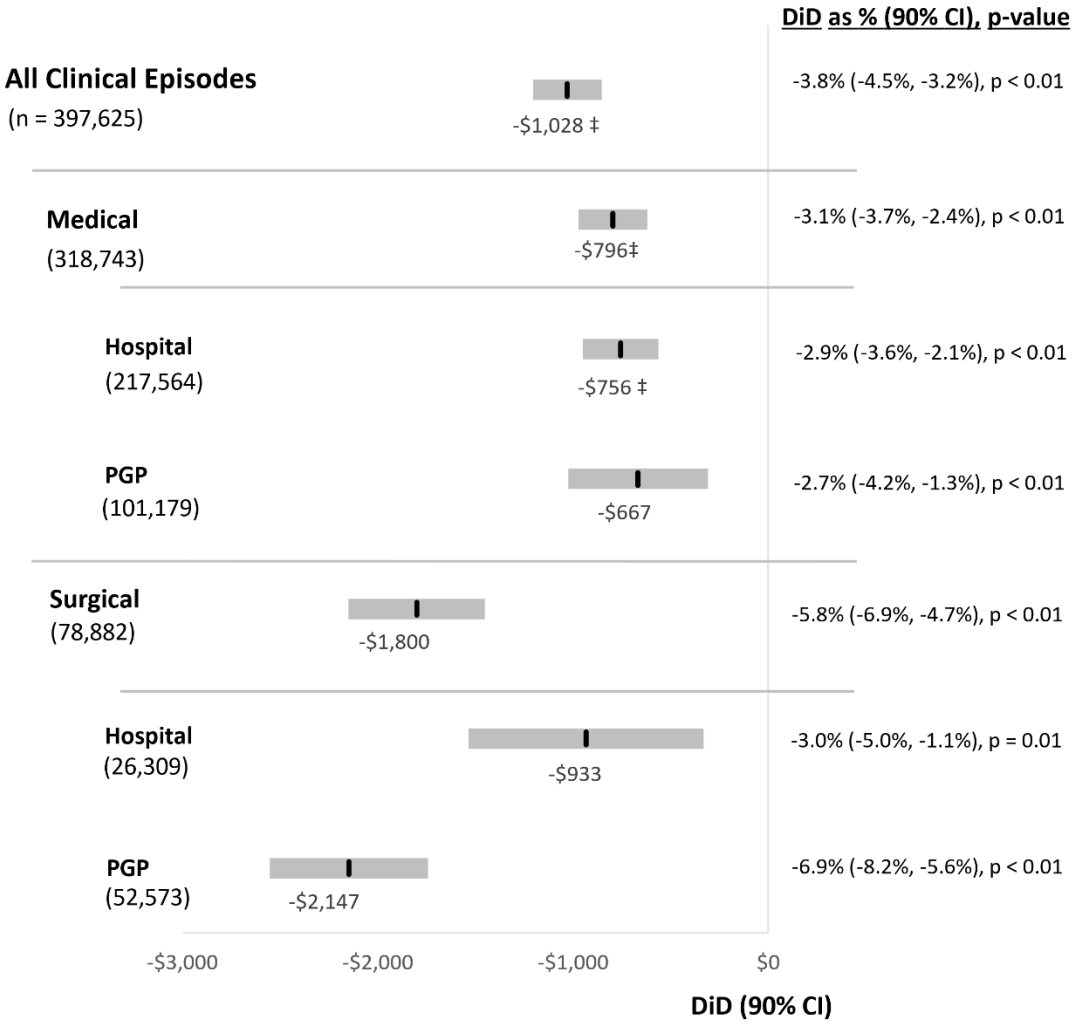
Both hospital and PGP episode initiators reduced per-episode payments for medical clinical episodes relative to comparison hospitals and PGPs, though hospital episode initiators reduced payments by a slightly larger amount. Hospital episode initiators reduced spending on medical clinical episodes by $756 per episode (90% confidence interval: -$950, -$563; $p < 0.01$) or 2.9% (though this outcome did not pass the parallel trends test for this grouping), while PGP episode initiators reduced per-episode payments for medical clinical episodes by $667 (90% confidence interval: -$1,025, -$310; $p<0.01$) or 2.7% (Exhibit 11).

For surgical clinical episodes, both hospital and PGP episode initiators reduced per-episode payments, but the decline was more than twice as large for PGP episode initiators as it was for hospital episode initiators. Hospital episode initiators reduced per-episode payments for surgical clinical episodes by $933 (90% confidence interval: -$1,534, -$331; $p=0.01$) or 3.0% of the baseline mean, while PGP episode initiators reduced per-episode payments for surgical clinical episodes by $2,147 (90% confidence interval: -$2,552, -$1,741; $p < 0.01$) or 6.9%.[23]

[22] Total episode payments for all episodes pooled across the clinical episodes evaluated and for medical clinical episodes did not pass the parallel trends test. A key assumption required for an unbiased DiD estimate is that outcomes for BPCI Advanced and the comparison group have the same trend during the baseline period. Additional discussion of the implications of the parallel trends test is in Chapter III. Discussion and Conclusion. Results of the parallel trends tests are reported in **Appendix G**. Additional details on the parallel trends test methodology are described in **Appendix C**.

[23] The estimated reductions in average standardized episode payments were generally robust across the specifications tested. For sensitivity test results see **Appendix H**.

Exhibit 11: Impact of BPCI Advanced on Average Episode Payments, Hospital and PGP Episode Initiators, January 1, 2020 – December 31, 2020



Note: Total payments represent Part A and B fee-for-service payments for the anchor stay or procedure and the 90-day PDP. The estimates in this exhibit are the results of a DiD model. The DiD estimates represent the relative change in dollars. Results are also presented as a percentage of average episode payments for BPCI Advanced during the baseline. The grey bars indicate the 90% confidence interval of the DiD estimate. This payment outcome is standardized to remove the effect of geographic and other payment adjustments. The number of episodes in each subgroup may not sum to the total, as episode-level weights were used to account for overlapping episodes. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period. PGP = physician group practice.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

e. How have average post-acute care payments changed under BPCI Advanced?

Reductions in payments to PAC providers during the 90-day post-discharge period (PDP) were a large driver of the declines in total episode payments. During Model Year 3, the BPCI Advanced Model reduced per-episode skilled nursing facility (SNF) payments by $494 (90% confidence interval: -$603, -$384; p <0.01) or 9.2% of the baseline mean SNF payment (this outcome did not pass the parallel trends test for this grouping). The BPCI Advanced Model reduced per-episode inpatient rehabilitation facility (IRF) payments by $234 (90% confidence interval -$294, -$173; p < 0.01) or 20.3% of the baseline mean IRF payment relative to the comparison group (Exhibit 12.1).

The model also reduced payments for home health (HH) services, although this represents a much smaller portion of the reduction in total episode payments. HH payments declined by $42 (90% confidence interval: -$68, -$17; p < 0.01) or 3.2% of the baseline mean HH payment (though this outcome did not pass the parallel trends test for this grouping).

For medical clinical episodes, the reduction in SNF payments was larger than the decline in IRF payments.[24] The BPCI Advanced Model reduced SNF payments by $452 (90% confidence interval: -$564, -$341; p < 0.01), or 8.6% of the baseline mean SNF payment, and the model reduced IRF payments by $109 (90% confidence interval: -$163, -$55; p < 0.01) or 10.8% of the baseline mean IRF payment, though SNF payments did not pass the parallel trends test for this grouping (Exhibit 12.2). In contrast, for surgical clinical episodes, the reduction in IRF payments was larger than the decline in SNF payments; the BPCI Advanced Model reduced SNF payments by $533 (90% confidence interval: -$756, -$309; p < 0.01), or 9.0% of the baseline mean SNF payment, and it reduced IRF payments by $754 (90% confidence interval: -$935, -$574; p < 0.01) or 44.1% of the baseline mean IRF payment (Exhibit 12.3).

The model had little impact on payments for HH services for medical clinical episodes; there was a small, not statistically significant relative increase in HH payments. For surgical clinical episodes, reductions in payments for HH services contributed to the overall decline in total episode payments, as payments for HH services declined by $228 (90% confidence interval: -$308, -$147; p < 0.01) or 13.0% of the baseline mean HH payment. See **Appendix F** for detailed results and for results by clinical episode.

The BPCI Advanced Model reduced payments for inpatient hospital readmissions during the 90-day PDP by $55 (90% confidence interval: -$100, -$10; p=0.04) or 1.5% of the baseline mean readmission payment, though this outcome did not pass the parallel trends test for this grouping. Both medical and surgical clinical episodes had reductions in readmission payments, though the reductions were not statistically significant, and the result for medical clinical episodes did not pass the parallel trends test.

[24] While the reduction in SNF payments was larger than the decline in IRF payments, the decline expressed as a percentage of the baseline mean is larger for IRF than for SNF because average SNF payments are larger than average IRF payments.

Results by Hospital and PGP Episode Initiators

Both hospital and PGP episode initiators reduced SNF payments for medical clinical episodes, though the reductions were slightly larger for hospital episode initiators. Hospital episode initiators reduced SNF payments by $481 (90% confidence interval: -$607, -$354; p < 0.01), or 8.9% of the baseline mean SNF payment relative to comparison hospitals, while PGP episode initiators reduced SNF payments by $282 (90% confidence interval: -$488, -$76; p=0.02), or 6.0% of the baseline mean SNF payment, relative to comparison PGPs.

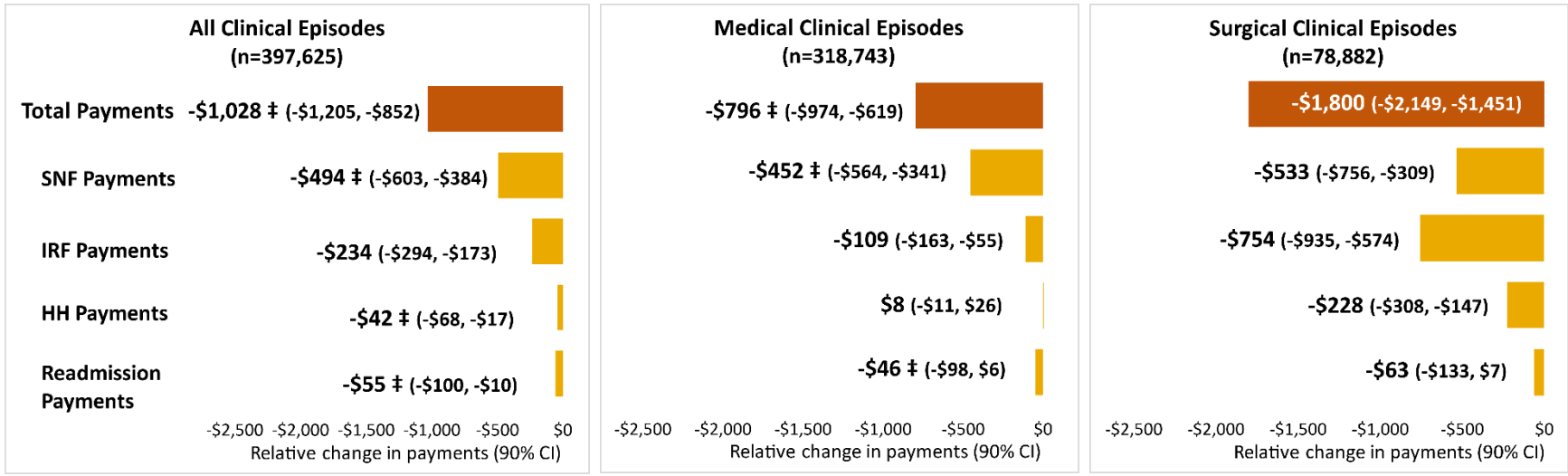
Similarly, hospital episode initiators had larger relative reductions in IRF payments than PGP episode initiators for medical clinical episodes. Hospital episode initiators reduced IRF payments by $136 (90% confidence interval: -$199, -$73; p < 0.01), or 12.6% of the baseline mean IRF payment, while PGP episode initiators reduced IRF payments by $33 (90% confidence interval: -$120, $54; p 0.54), or 4.1% of the baseline mean IRF payment. The reduction in IRF payments for PGP episode initiators was not statistically significant. Both hospital and PGP episode initiators had small, not statistically significant increases in payments for HH services for medical clinical episodes.

Hospital episode initiators had a small, not statistically significant increase in readmission payments for medical clinical episodes, while PGP episode initiators reduced medical clinical episode readmission payments by $142 (90% confidence interval: -$235, -$50; p=0.01), or 3.7% of the baseline mean readmission payment, though this outcome did not pass the parallel trends test for hospital or PGP medical clinical episodes.

Both hospital and PGP episode initiators reduced SNF and IRF payments for surgical clinical episodes. Hospital episode initiators reduced SNF payments by $328 (90% confidence interval: -$701, $44; p=0.15), or 5.9% of the baseline mean SNF payment, and IRF payments by $402 (90% confidence interval: -$652, -$151; p=0.01) or 22.8% of the baseline mean IRF payment. For hospital episode initiators, the reduction in SNF payments was not statistically significant. PGP episode initiators reduced SNF payments by $615 (90% confidence interval: -$898, -$331; p < 0.01), or 10.0% of the baseline mean SNF payment, and IRF payments by $911 (90% confidence interval: -$1,157, -$665; p < 0.01), or 53.5% of the baseline mean IRF payment. Hospital episode initiators also reduced HH payments, but the change was small and not statistically significant. For PGP surgical clinical episodes, reductions in payments for HH services contributed to the reduction in total episode payments as PGP episode initiators reduced payments for HH services by $332 (90% confidence interval: -$442, -$222; p < 0.01), or 17.4% of the baseline mean HH payment, relative to comparison PGPs.

For surgical clinical episodes, hospital episode initiators reduced readmission payments by $125 (90% confidence interval: -$264, $14; p=0.14), or 5.5% of the baseline mean readmission payment, though the estimate was not statistically significant. The model had little impact on readmission payments for PGP surgical clinical episodes; there was a small, not statistically significant reduction.

Exhibit 12.1: Impact of BPCI Advanced on SNF, IRF, HH and Readmission Payments in the 90-day PDP, Hospital and PGP Episode Initiators, Model Year 3, January 1, 2020 – December 31, 2020

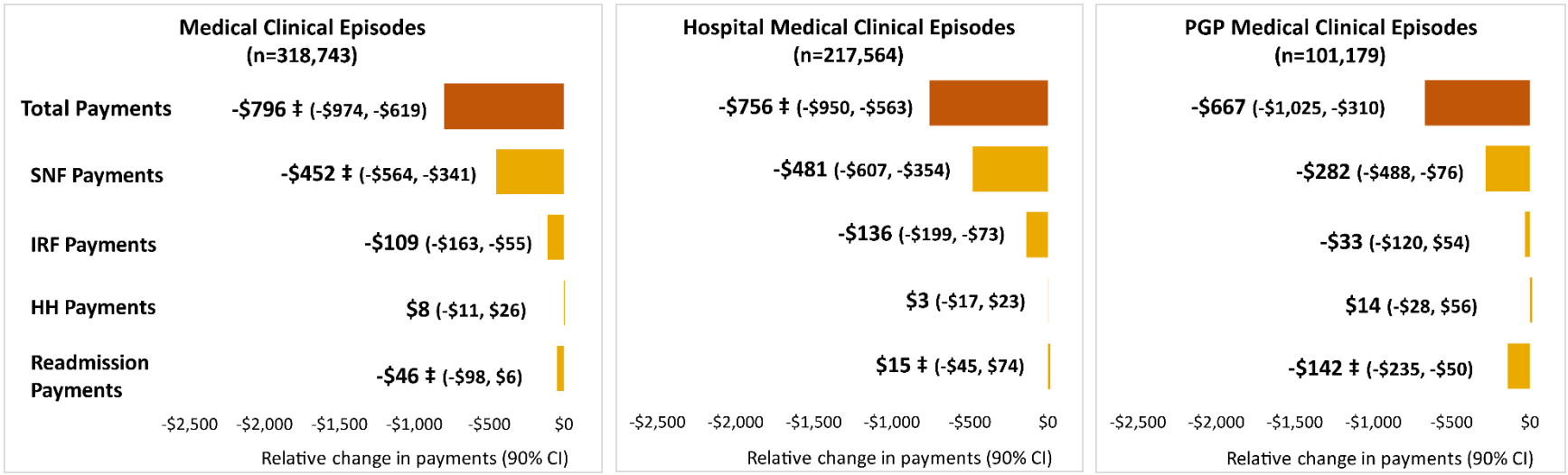


Note: The estimates in this exhibit are the results of DiD models. The DiD estimates represent the relative change in dollars. Payment outcomes were standardized to remove the effect of geographic and other payment adjustments. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; HH = home health; IRF = inpatient rehabilitation facility; PDP = post-discharge period; PGP = physician group practice; SNF = skilled nursing facility.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

Exhibit 12.2: Impact of BPCI Advanced on SNF, IRF, HH, and Readmission Payments in the 90-day PDP, Hospital and PGP Episode Initiators, Medical Clinical Episodes, Model Year 3, January 1, 2020 – December 31, 2020

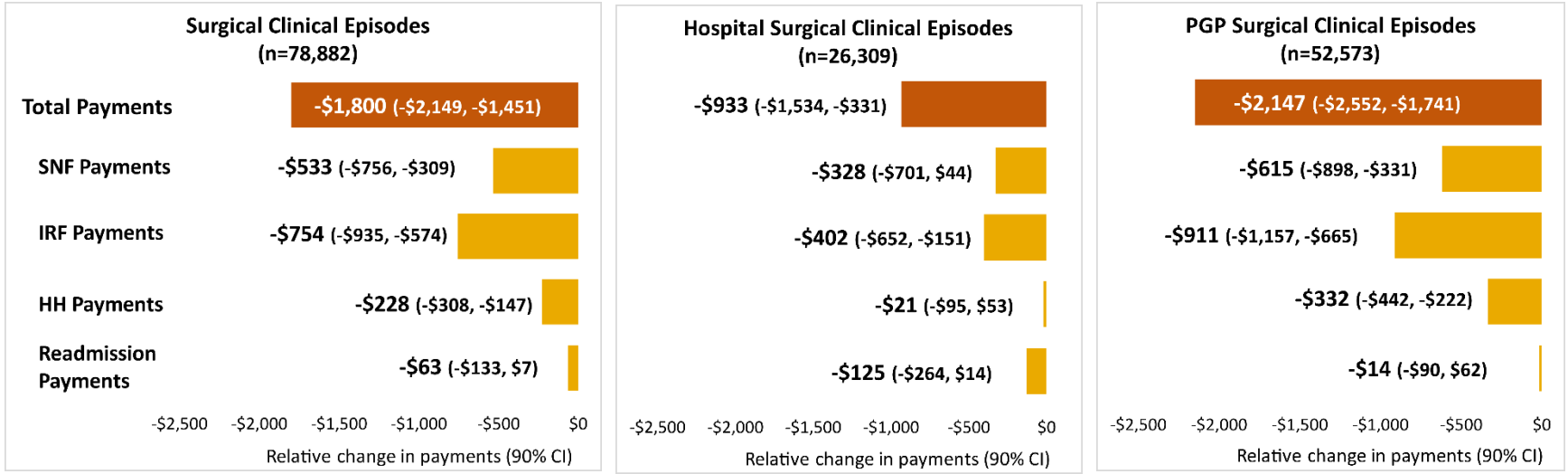


Note: The estimates in this exhibit are the results of DiD models. The DiD estimates represent the relative change in dollars. Payment outcomes were standardized to remove the effect of geographic and other payment adjustments. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; HH = home health; IRF = inpatient rehabilitation facility; PDP = post-discharge period; PGP = physician group practice; SNF = skilled nursing facility.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

Exhibit 12.3: Impact of BPCI Advanced on SNF, IRF, HH, and Readmission Payments in the 90-day PDP, Hospital and PGP Episode Initiators, Surgical Clinical Episodes, Model Year 3, January 1, 2020 – December 31, 2020



Note: The estimates in this exhibit are the results of DiD models. The DiD estimates represent the relative change in dollars. Payment outcomes were standardized to remove the effect of geographic and other payment adjustments. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; HH = home health; IRF = inpatient rehabilitation facility; PDP = post-discharge period; PGP = physician group practice; SNF = skilled nursing facility.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

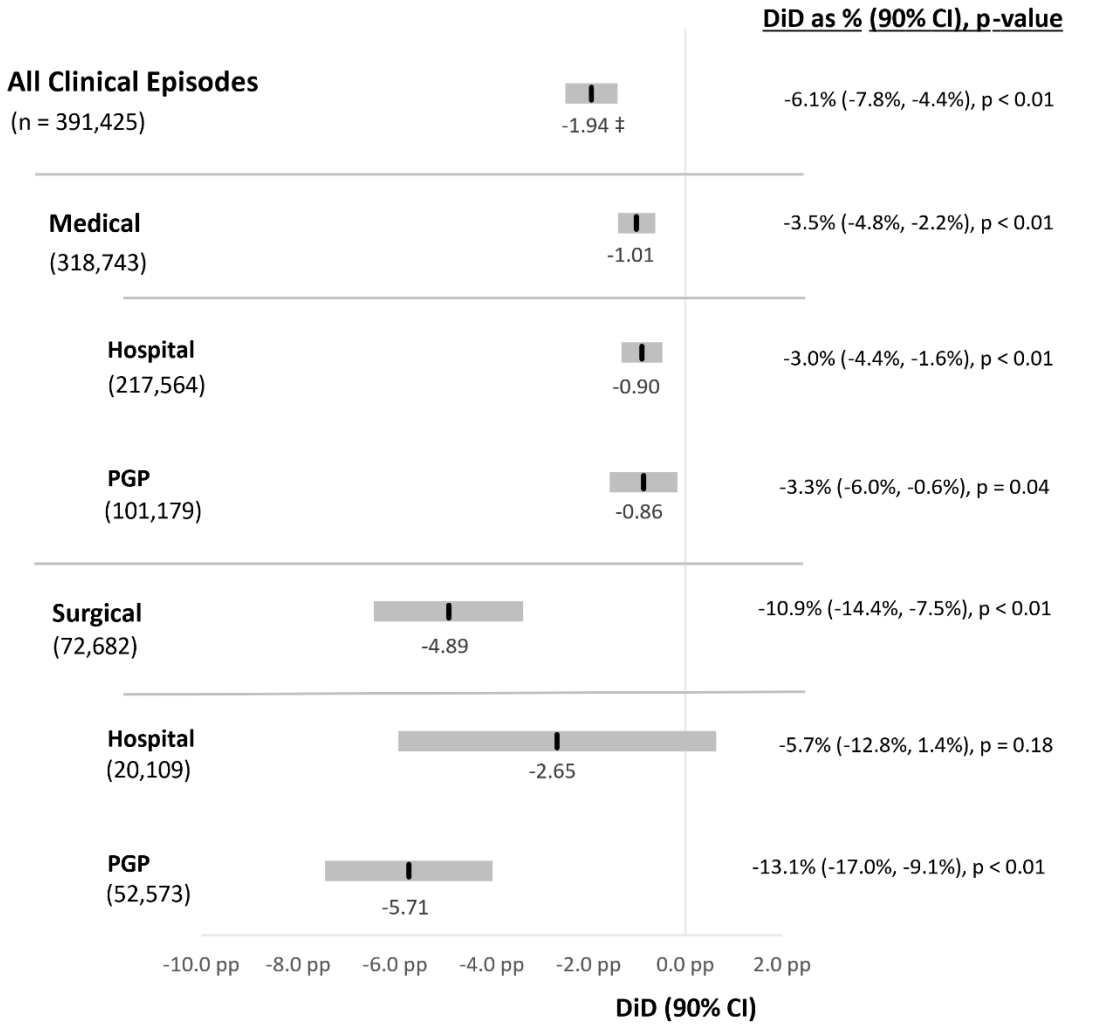
f. How has post-acute care service use changed under BPCI Advanced?

To help understand the changes in PAC payments above, we measured the impact of BPCI Advanced on two utilization outcomes for inpatient clinical episodes: the proportion of episodes with a first discharge to an institutional PAC setting (including SNF, IRF, and long-term care hospital) and the number of SNF days in the 90-day PDP (among beneficiaries with at least one SNF stay in the PDP, which account for about 31% of baseline episodes). These outcomes were chosen because evaluations of the BPCI Initiative Model 2 demonstrated that participants reduced episode payments primarily through these two mechanisms: reducing the proportion of episodes discharged to an institutional PAC setting and reducing the number of days in SNF.

During Model Year 3, the BPCI Advanced Model reduced the share of episodes discharged to an institutional PAC setting by 1.94 pp (90% confidence interval: -2.48, -1.40; $p < 0.01$), or 6.1% of the baseline mean discharge to an institutional PAC setting relative to comparison hospitals and PGPs, though this outcome did not pass the parallel trends test for this grouping (Exhibit 13). The reduction in the share of episodes discharged to institutional PAC settings was smaller for medical clinical episodes than surgical clinical episodes. For medical clinical episodes, the share of episodes discharged to institutional PAC settings declined 1.01 pp (90% confidence interval: -1.39, -0.63; $p < 0.01$), or 3.5%. For surgical clinical episodes, the share of episodes first discharged to an institutional PAC setting declined 4.89 pp (90% confidence interval: -6.43, -3.35; $p < 0.01$), or 10.9%. See **Appendix F** for detailed results and for results by clinical episode.

For medical clinical episodes, hospital and PGP episode initiators had similar reductions in the share of episodes first discharged to an institutional PAC setting. Hospital episode initiators reduced the share by 0.90 pp (90% confidence interval: -1.33, -0.47; $p < 0.01$), or 3.0% of the baseline mean discharge to an institutional PAC setting, while PGPs reduced the share by 0.86 pp (90% confidence interval: -1.57, -0.16; $p=0.04$), or 3.3%. For surgical clinical episodes, PGP episode initiators had a much larger reduction in the share of episodes first discharged to an institutional PAC setting compared to hospital episode initiators. Hospital episode initiators reduced the share of episodes first discharged to an institutional PAC setting by 2.65 pp (90% confidence interval: -5.93, 0.63; $p=0.18$), or 5.7%, and the result was not statistically significant, while PGP episode initiators reduced the share by 5.71 pp (90% confidence interval -7.44, -3.98; $p < 0.01$), or 13.1%.

Exhibit 13: Impact of BPCI Advanced on First Discharge to Institutional PAC Setting, Hospital and PGP Episode Initiators, Model Year 3, January 1, 2020 – December 31, 2020



Note: The estimates in this exhibit are the results of DiD models. The DiD estimates represent a percentage point change. Results are also presented as a percentage of the BPCI Advanced baseline average rate. The grey bars indicate the 90% confidence interval of the DiD estimate. The number of episodes in each subgroup may not sum to the total, as episode-level weights were used to account for overlapping episodes. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PAC = post-acute care; PGP = physician group practice; pp = percentage point(s).

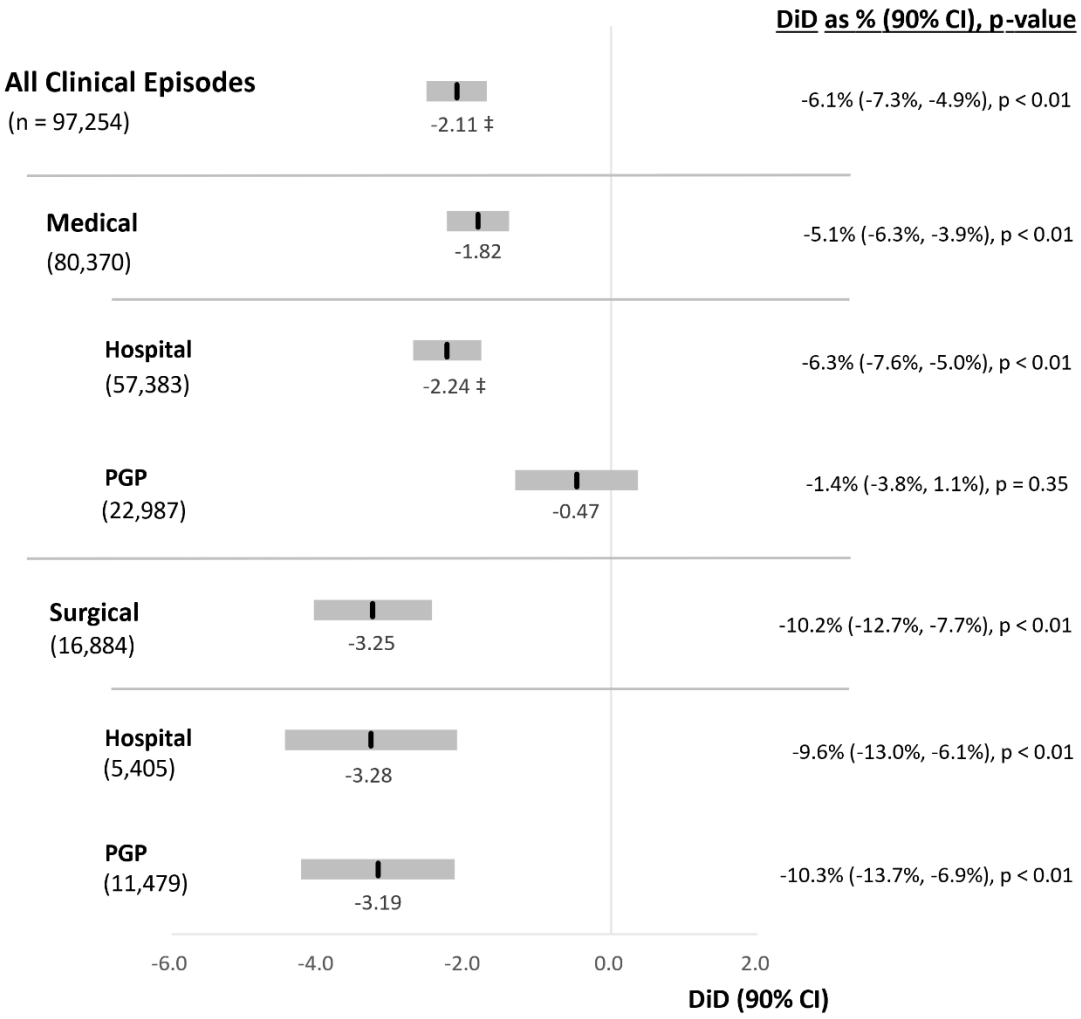
‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

The BPCI Advanced Model reduced the number of days in a SNF (referred to as SNF days) among episodes with at least one day of SNF use during the 90-day PDP by 2.11 days (90% confidence interval: -2.52, -1.70; $p < 0.01$), or 6.1% of the baseline mean number of SNF days, though this outcome did not pass the parallel trends test for this grouping (Exhibit 14). Reductions were larger in surgical clinical episodes than in medical clinical episodes. For medical clinical episodes, the reduction in SNF days was 1.82 days (90% confidence interval: -.24, -1.40; $p < 0.01$), or 5.1%. For surgical clinical episodes, the reduction was 3.25 days (90% confidence interval: -4.06, -2.45; $p < 0.01$), or 10.2%.

Both hospital and PGP episode initiators reduced the number of SNF days among SNF users for medical clinical episodes relative to comparison hospitals and PGPs; however, the reductions were larger for hospital episode initiators. Hospital episode initiators reduced SNF days by 2.24 days (90% confidence interval: -2.71, -1.77; $p < 0.01$), or 6.3% of the baseline mean number of SNF days, though this outcome did not pass the parallel trends test for this grouping. PGP episode initiators reduced SNF days by 0.47 days (90% confidence interval: -1.31, 0.37; $p=0.35$), or 1.4%, but this reduction was not statistically significant. For surgical clinical episodes, hospital and PGP episode initiators reduced SNF days by a similar amount. Hospital episode initiators reduced SNF days by 3.28 days (90% confidence interval: -4.45, -2.11; $p < 0.01$), or 9.6%, while PGP episode initiators reduced SNF days by 3.19 days (90% confidence interval: -4.24, -2.13; $p < 0.01$), or 10.3%.

Exhibit 14: Impact of BPCI Advanced on Number of SNF Days for SNF Users in the 90-day PDP, Hospital and PGP Episode Initiators, Model Year 3, January 1, 2020 – December 31, 2020



Note: The estimates in this exhibit are the results of DiD models. Results are presented as the relative change in days. Results are also presented as a percentage of the BPCI Advanced baseline mean number of SNF days. The grey bars indicate the 90% confidence interval of the DiD estimate. The number of episodes in each subgroup may not sum to the total, as episode-level weights were used to account for overlapping episodes. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; PGP = physician group practice; SNF = skilled nursing facility.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

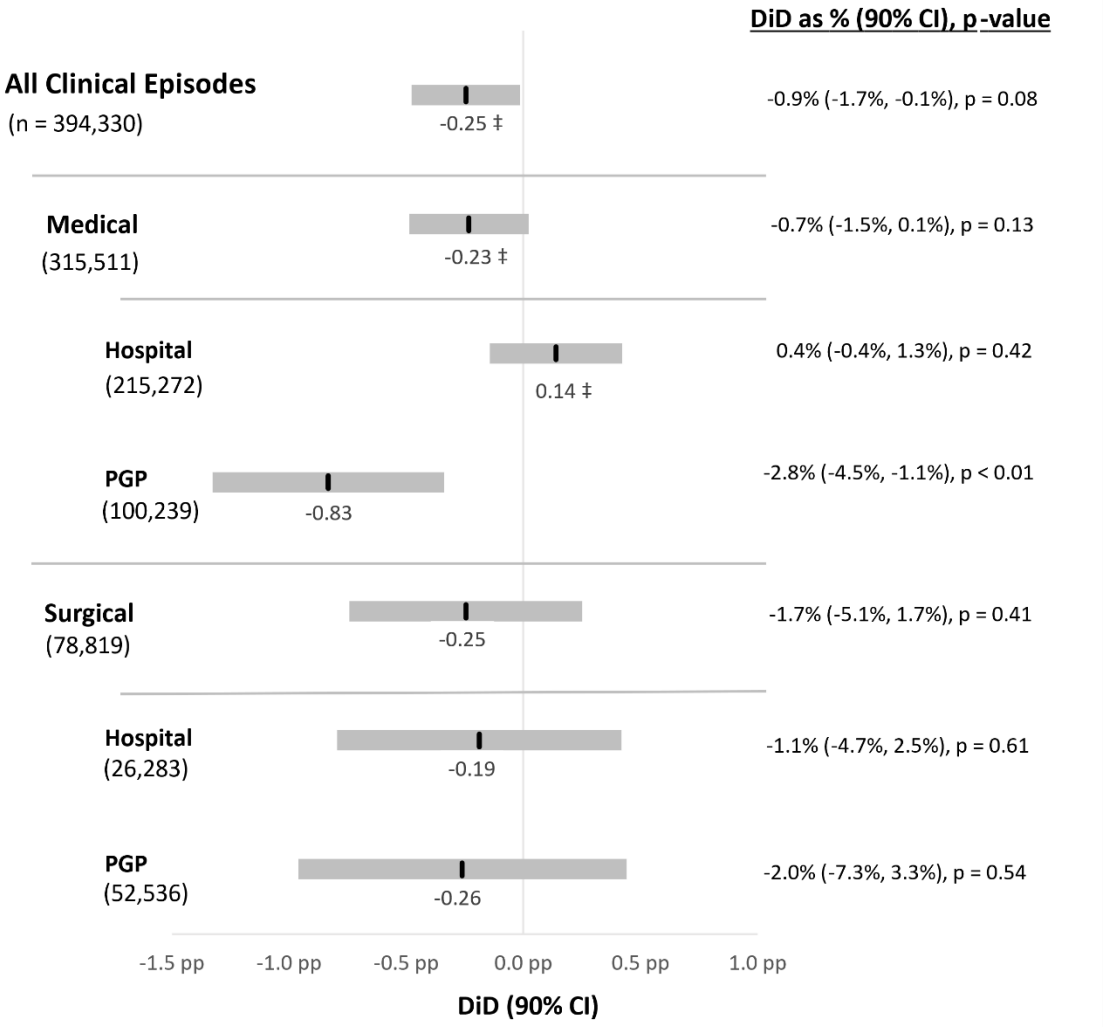
g. How has claims-based quality of care changed under BPCI Advanced?

To assess the quality of care received by beneficiaries treated by BPCI Advanced participants, we evaluated the impact of the model on two claims-based quality measures: the unplanned readmission rate during the 90-day PDP and the mortality rate during the anchor hospitalization through the 90-day PDP.[25]

During Model Year 3, there was a pattern of small reductions in the unplanned readmission rate for episodes pooled across all clinical episodes evaluated, for medical clinical episodes, and for surgical clinical episodes relative to the comparison group (Exhibit 15). The estimate for all clinical episodes was statistically significant: there was a decline of 0.25 pp (90% confidence interval: -0.48, -0.01; p=0.08), or 0.9% of the baseline mean readmission rate (though this outcome did not pass the parallel trends test). When analyzed by episode initiator type, PGP medical clinical episodes had a statistically significant decline in the readmission rate of 0.83 pp (90% confidence interval: -1.33, -0.34; p=0.01), or 2.8% of the baseline mean readmission rate for PGP medical clinical episodes. See **Appendix F** for detailed results and for results by clinical episode.

[25] We define the mortality rate measure to include mortality during both the anchor hospitalization and 90-day PDP. This is a comprehensive measure of the mortality rate since it includes those who do not survive the anchor hospitalization. Beneficiaries who do not survive the anchor hospitalization are not eligible to become a part of the BPCI Advanced Model. Other outcomes analyzed exclude episodes in which the beneficiary did not survive the anchor hospitalization.

Exhibit 15: Impact of BPCI Advanced on the Unplanned Readmission Rate During the 90-day PDP, Hospital and PGP Episode Initiators, Model Year 3, January 1, 2020 – December 31, 2020



Note: The estimates in this exhibit are the results of DiD models. The DiD estimates represent a percentage point change. Results are also presented as a percentage of the BPCI Advanced baseline average rate. The grey bars indicate the 90% confidence interval of the DiD estimate. The number of episodes in each subgroup may not sum to the total, as episode-level weights were used to account for overlapping episodes. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; PGP = physician group practice; pp = percentage point(s).

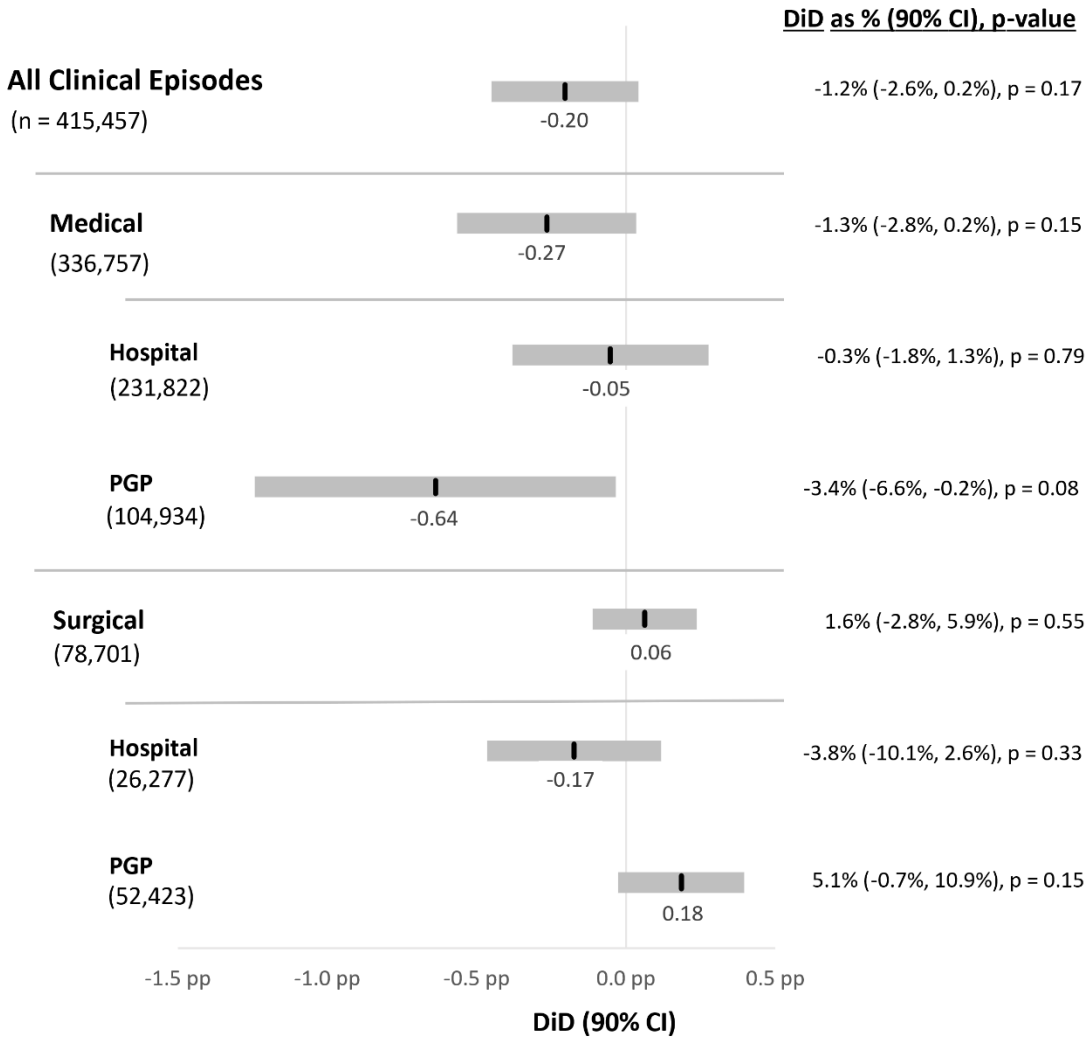
‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

BPCI Advanced did not have a statistically significant impact on the mortality rate for episodes pooled across all clinical episodes evaluated relative to the comparison group in Model Year 3, but

there was a statistically significant decline for PGP medical clinical episodes of 0.64 pp (90% confidence interval: -1.24, -0.03; p=0.08), or 3.4% of the baseline mean mortality rate.[26]

Exhibit 16: Impact of BPCI Advanced on the Mortality Rate During the Anchor Stay and 90-day PDP, Hospital and PGP Episode Initiators, Model Year 3, January 1, 2020 – December 31, 2020



Note: The estimates in this exhibit are the results of DiD models. The DiD estimates represent a percentage point change. Results are also presented as a percentage of the BPCI Advanced baseline average rate. The grey bars indicate the 90% confidence interval of the DiD estimate. The number of episodes in each subgroup may not sum to the total, as episode-level weights were used to account for overlapping episodes. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix F** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; PGP = physician group practice; pp = percentage point(s).

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or

[26] The findings for the unplanned readmission and mortality rates were generally robust across the specifications tested. For sensitivity test results see **Appendix H**.

procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

2. *Patient-reported functional status, care experiences, and overall satisfaction with care*

This section presents patient-reported changes in functional status, care experiences and satisfaction from the beneficiary survey for beneficiaries with episodes initiated in July and August 2021. We sampled beneficiaries in August and September 2021, and the survey was in the field from October 2021 until February 2022. On average, we received survey responses 115 days after a beneficiary's date of discharge or procedure. We sampled beneficiaries from all 34 BPCI Advanced clinical episodes separately for beneficiaries attributed to BPCI Advanced hospitals and beneficiaries attributed to BPCI Advanced PGPs. Comparison beneficiaries were matched to the BPCI Advanced beneficiaries using a nearest-neighbor propensity score approach, coupled with exact matching on clinical episode, age category (less than 65, 65-74, 75-84, 85 and older), and presence of major complication or comorbidity. We present results for respondents pooling across all 34 clinical episodes, by episode initiator type (hospital and PGP), and by clinical episode type (medical and surgical) (see Exhibit 2 above). We also discuss select findings for key clinical episode service line groups (CESLGs). Full results for all CESLGs that were powered for analysis are reported in **Appendix I**.

We used a cross-sectional regression approach to estimate differences in outcomes between respondents attributed to BPCI Advanced hospitals or PGPs and respondents attributed to a matched comparison group. Survey outcomes included seven measures of patient-reported change in functional status (improved, declined, or stayed the same) from before to after hospitalization, eight measures of care experiences, and two measures of satisfaction with care. All survey measures are based on categorical items such that differences in measures represent substantive differences in outcomes (such as, "agree" versus "disagree", "some help needed" versus "complete help needed"). Therefore, small differences between BPCI Advanced and comparison respondents should be interpreted as "a small probability of substantive differences" rather than "small differences in outcomes." Additional detail about the measures can be found in **Appendix C**.

The response rate averaged 26.5% among BPCI Advanced beneficiaries attributed to hospitals and 27.6% among comparison group beneficiaries, resulting in a response rate that was 1.1 pp lower (90% confidence interval: -1.9, -0.2; p=0.04) for BPCI Advanced beneficiaries attributed to hospitals. The response rate averaged 28.7% among BPCI Advanced beneficiaries attributed to PGPs and 25.3% among comparison group beneficiaries, resulting in a response rate that was 3.4 pp greater (90% confidence interval: 2.2, 4.6; p<0.01) for BPCI Advanced beneficiaries attributed to PGPs. The survey results for beneficiaries whose episodes were attributed to hospitals are based on 4,128 BPCI Advanced responses and 4,291 comparison group responses. The survey results for beneficiaries whose episodes were attributed to PGPs are based on 2,159 BPCI Advanced responses and 1,900 comparison group responses.

We applied non-response and sampling weights to all observations. Estimated differences between the BPCI Advanced and comparison respondents were risk-adjusted for beneficiary-, hospital-, and neighborhood-level characteristics. Because survey data were only collected during the intervention period, we cannot distinguish whether differences between BPCI Advanced and comparison respondents existed prior to the model or whether they were caused by BPCI

Advanced. We present more detail on the survey measures, sample selection, weighting, risk-adjustment, and strata-level results in **Appendix C**.

a. Key Findings

Patient-reported Outcomes Under BPCI Advanced

- For both hospital- and PGP-initiated episodes, BPCI Advanced respondents were slightly more likely to report unfavorable changes in functional status than comparison respondents during Model Year 4 (2021).
- For hospital-initiated episodes, BPCI Advanced respondents were slightly less likely to report favorable care experiences and the highest levels of satisfaction with care than comparison respondents.
- For PGP-initiated episodes, there was a mix of favorable and unfavorable results for care experiences and satisfaction with care across clinical episode service line groups.

b. Hospital Strata

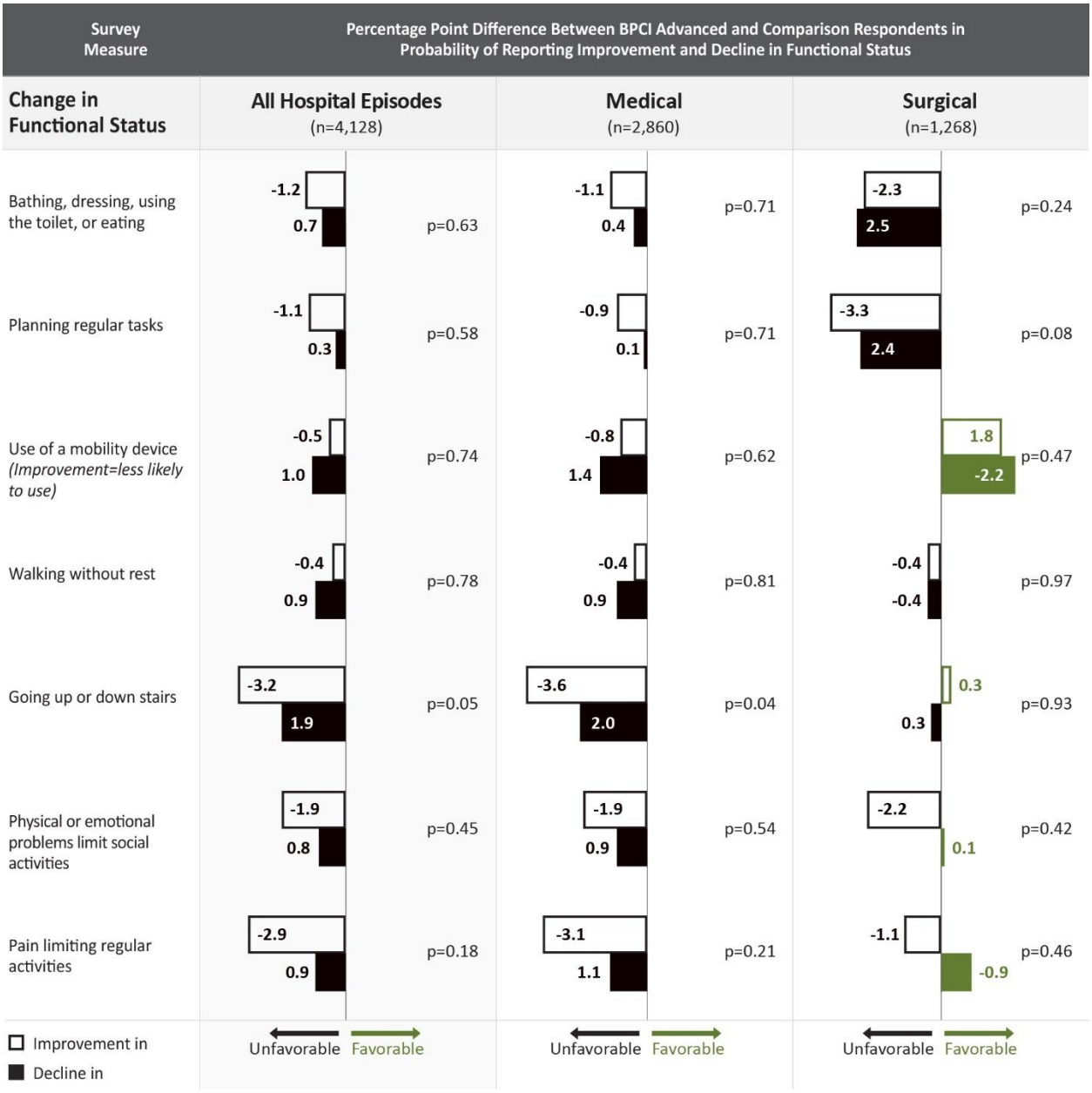
How did patient-reported functional status change for BPCI Advanced respondents with hospital-initiated episodes relative to comparison respondents?

BPCI Advanced respondents with hospital-initiated episodes in Model Year 4 were less likely to report favorable changes in functional status from before the anchor stay or procedure to the time of the survey than comparison respondents across all seven measures, and one result was statistically significant (Exhibit 17). BPCI Advanced respondents were 3.2 pp less likely to report improvement in going up or down stairs and 1.9 pp more likely to report decline (p=0.05). This unfavorable result for going up or down stairs was driven by hospital medical clinical episodes, particularly in the *medical and critical care* CESLG (Exhibit 19). Furthermore, for hospital medical clinical episodes, results were consistently unfavorable across the seven functional status outcomes.

For hospital surgical clinical episodes, results were mixed across functional status outcomes because of variation in results across hospital surgical CESLGs. For *cardiac procedures*, BPCI Advanced respondents were more likely to report favorable changes across the functional status measures, but for *orthopedics*, BPCI Advanced respondents were more likely to report unfavorable changes. Results for *spinal procedures* were mixed (Exhibit 19).

Across both medical and surgical episodes, we found generally unfavorable results in each of the three highest-volume CESLGs analyzed, and we found favorable results in only one CESLG. This suggests that our pooled findings capture a broad pattern of outcomes across the model and are not driven solely by less favorable outcomes for one type of clinical condition.

Exhibit 17: Differences in Patient-Reported Change in Functional Status Between BPCI Advanced and Comparison Respondents, Hospitals, July 2021-August 2021



***Note*:** The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. All responses were weighted for non-response and sampling design. Estimates were based on 4,128 BPCI Advanced survey respondents and 4,291 comparison survey respondents across all 34 clinical episodes. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results are reported in percentage point terms. Values to the left, represented in black, indicate unfavorable results. Values to the right, represented in green, indicate favorable results. Thus, increases in decline are displayed on the left, and decreases in decline are displayed on the right. The p-values for functional status results indicate joint significance for differences in the proportion of respondents indicating one of three categories: improvement (or maintained highest function); stayed the same; or declined (or maintained lowest function). See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results.

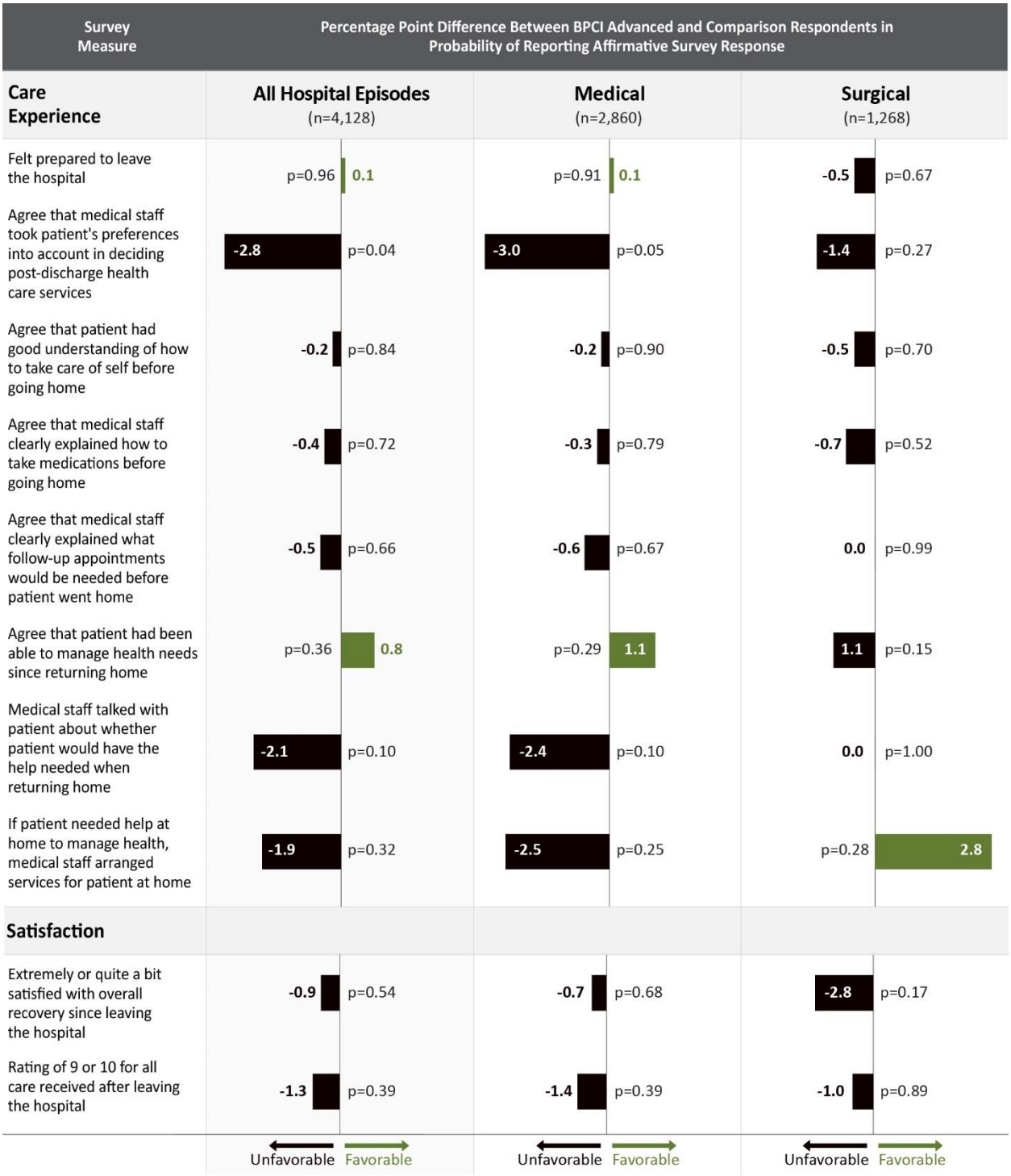
Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Were patient-reported care experiences and satisfaction similar between BPCI Advanced respondents with hospital-initiated episodes and comparison respondents?

BPCI Advanced respondents with hospital-initiated episodes in Model Year 4 were less likely to report favorable care experience outcomes for six of eight measures, and two results were statistically significant at the 10 percent level or less (Exhibit 18). BPCI Advanced respondents were 2.8 pp less likely (90% confidence interval: -5.0, -0.6; p=0.04) to agree that their preferences were taken into account in deciding health care services received after leaving the hospital, and 2.1 pp less likely (90% confidence interval: -4.1, 0.0; p=0.10) to agree that medical staff discussed with them whether they would have the necessary help when they got home. These two unfavorable results were driven almost entirely by the *medical and critical care* CESLG, which represents the majority of episode volume under the model. The results for medical clinical episodes alone were similar to the results for hospital clinical episodes pooled across medical and surgical clinical episodes, with BPCI Advanced respondents less likely to report favorable care experiences for the same six of eight measures. The results for hospital surgical clinical episodes were mixed, with unfavorable outcomes for five of eight measures, neutral outcomes for two measures, and a favorable outcome for one measure. There was also a mix of favorable and unfavorable results within each of the three hospital surgical CESLGs analyzed (*cardiac procedures*, *orthopedics* and *spinal procedures*) (Exhibit 20). This pattern of results suggests that care experiences for BPCI Advanced respondents were not consistently worse across clinical conditions. However, the lone CESLG with a pattern suggesting adverse results (*medical and critical care*) accounted for the majority of hospital episode volume under the model, which suggests that most BPCI Advanced respondents may have faced a slightly higher probability of less favorable care experiences, even if such outcomes were not widespread across clinical conditions.

For all hospital clinical episodes, although estimated differences were small and not statistically significant, BPCI Advanced respondents were slightly less likely to report the highest levels of satisfaction relative to comparison respondents. Additionally, four out of the five CESLGs analyzed had unfavorable satisfaction outcomes, and one outcome was statistically significant (Exhibit 20). For the *medical and critical care* CESLG, BPCI Advanced respondents were 4.3 pp less likely to report the highest levels of satisfaction with post-acute care relative to comparison respondents (p=0.09) (Exhibit 20). In contrast, the *cardiac care* CESLG was the one CESLG with favorable satisfaction outcomes: BPCI Advanced respondents were 7.7 pp more likely to report the highest levels of satisfaction with post-acute care relative to comparison respondents (p=0.03).

Exhibit 18: Differences in Patient-Reported Change in Care Experiences and Satisfaction with Care Between BPCI Advanced and Comparison Respondents, Hospitals, July 2021 – August 2021



***Note*:** The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for binary indicators. All responses were weighted for non-response and sampling design. Estimates were based on 4,128 BPCI Advanced survey respondents and 4,291 comparison survey respondents across all 34 clinical episodes. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results reported in percentage point terms. Values to the left, represented in black,

indicate unfavorable results. Values to the right, represented in green, indicate favorable results. The p-value for satisfaction with all care received after leaving the hospital indicates joint significance for differences in proportion of respondents indicating one of three categories: 9-10 rating; 7-8 rating; 0-6 rating. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results.

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Exhibit 19: Differences in Patient-Reported Change in Functional Status Between BPCI Advanced and Comparison Respondents for Individual Clinical Episode Service Line Groups, Hospitals, July 2021-August 2021

Domain	Measures	Response if indicator = 1	Medical Episodes		Surgical Episodes		
			Medical and Critical Care (n=1,448)	Cardiac Care (n=890)	Cardiac Procedures (n=351)	Orthopedics (n=577)	Spinal Procedures (n=301)
Functional Status and Activities of Daily Living	Bathing, dressing, using the toilet, or eating	Improvement	-1.6	-2.5	2.7	-2.9	-3.7
		Decline	1.1	0.2	-2.1	3.2	4.2
	Planning regular tasks	Improvement	-0.9	-2.8	-0.3	-4.1	-4.7
		Decline	-0.1	2.1	0.8	2.8	3.3
	Use of mobility device	Improvement	0.2	-2.7	**-1.0***	3.0	3.4
		Decline	-0.3	4.8	**-3.5**	-3.5	-2.7
	Walking without rest	Improvement	-0.7	-3.5	2.1	-1.3	1.9
		Decline	1.0	3.3	-0.1	2.0	-5.4
	Going up or down stairs	Improvement	**-5.1****	-0.6	0.7	-0.1	**-0.4***
		Decline	**4.1**	-3.2	-1.3	3.1	**-6.2**
	Physical/emotional problems limiting social activities	Improvement	-3.3	-1.1	-0.3	-3.0	-1.6
		Decline	1.6	0.4	-3.2	1.7	-1.5
	Pain limiting regular activities	Improvement	-4.5	0.2	5.9	-3.4	-1.3
		Decline	2.2	1.6	-4.0	-0.1	3.6

Note: The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. All responses were weighted for non-response and sampling design. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results are reported in percentage point terms. Green shading indicates favorable results. Pink shading indicates unfavorable results. Grey shading indicates neutral results. If differences in improvement and decline indicated opposite effects (e.g., more improvement *and* more decline) and the net difference was less than 1.0 pp, differences were considered neutral. If the net difference exceeded 1.0 pp, then it was considered favorable (if more improvement/less decline) or unfavorable (if less improvement/more decline) on net. All decline measures are unshaded to help visually distinguish outcomes. We analyzed all service-line groups for which we had a minimum detectable difference of 10.0 pp or less. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results. Pp = percentage point(s). *p<0.10 **p<0.05 ***p<0.01

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Exhibit 20: Differences in Care Experiences and Satisfaction Between BPCI Advanced and Comparison Respondents for Individual Clinical Episode Service Line Groups, Hospitals, July 2021-August 2021

Domain	Measures	Response if indicator = 1	Medical Episodes		Surgical Episodes		
			Medical & Critical Care (n=1,448)	Cardiac Care (n=890)	Cardiac Procedures (n=351)	Orthopedics (n=577)	Spinal Procedures (n=301)
Care Experiences	How prepared did you feel to leave the hospital?	Very/Somewhat Prepared	-0.3	-0.2	-0.1	-1.1	1.2
	Medical staff took your preferences into account in deciding what health care services you should have after you left the hospital	Agree/Strongly Agree	**-4.4****	-0.3	0.4	-1.1	-2.1
	Good understanding of how to take care of self before going home	Agree/Strongly Agree	0.2	-2.7	-0.6	0.1	1.8
	Medical staff clearly explained how to take medications before going home	Agree/Strongly Agree	-1.0	0.1	2.6	-1.3	-1.9
	Medical staff clearly follow-up appointments or treatments would be needed	Agree/Strongly Agree	-1.0	-0.2	-0.2	0.5	1.3
	Able to manage your health needs since returning home	Agree/Strongly Agree	1.4	0.6	-0.9	-0.1	-2.1
	Medical staff talked with you about whether you would have the help you needed when you got home	Yes	**-3.6****	1.4	1.9	-1.1	-3.2
	If you needed help at home to manage your health, medical staff arranged services for you at home	Yes	-2.5	0.0	2.6	5.0	-1.0
Satisfaction with Care	Overall satisfaction with recovery since leaving hospital	Extremely Satisfied/Quite a bit Satisfied	-2.4	2.0	-2.0	-4.3	-3.7
	Rating of all care received after leaving the hospital	Rating 9-10	**-4.3***	**7.7****	-2.7	1.8	-5.8
		Rating 0-6	**4.0**	**-2.0**	-0.6	0.0	3.8

Note: The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for binary indicators. All responses were weighted for non-response and sampling design. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results are reported in percentage point terms. Green shading indicates favorable results. Pink shading indicates unfavorable results. Grey shading indicates neutral results. Results were judged neutral if the point estimate rounded to 0.0, indicating a difference less than 0.05 pp. We analyzed all service-line groups for which we had a minimum detectable difference of 10.0 pp or less. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results. Pp = percentage point(s). *p<0.10 **p<0.05 ***p<0.01

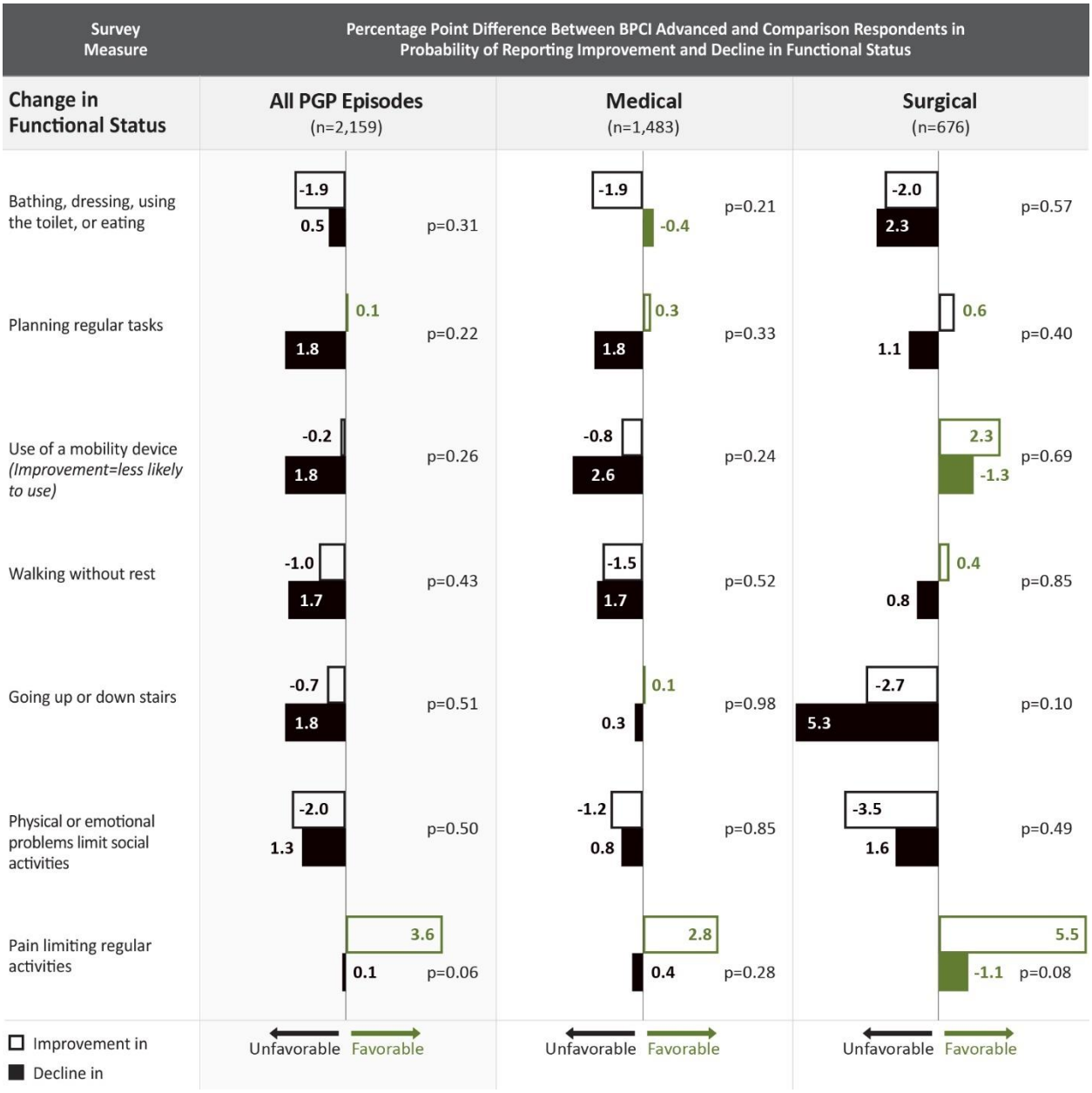
Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

c. PGP Strata

How did patient-reported functional status change for BPCI Advanced respondents with PGP-initiated episodes relative to comparison respondents?

BPCI Advanced respondents with PGP-initiated episodes in Model Year 4 were slightly less likely than comparison respondents to report favorable changes in functional status from before the anchor stay or procedure to the time of the survey for six of seven measures (Exhibit 21). The unfavorable functional status outcomes were primarily driven by medical CESLGs (Exhibit 23). Results were mixed for *orthopedics*, the one surgical CESLG analyzed, and the only statistically significant difference was favorable and showed that BPCI Advanced respondents were 8.8 pp *more* likely to report improvement in the degree to which pain limited their regular activities (p=0.01).

Exhibit 21: Differences in Patient-Reported Change in Functional Status Between BPCI Advanced and Comparison Respondents, PGPs, July 2021-August 2021



***Note*:** The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. All responses were weighted for non-response and sampling design. Estimates were based on 2,159 BPCI Advanced survey respondents and 1,900 comparison survey respondents across all 34 clinical episodes. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results reported in percentage point terms. Values to the left, represented in black, indicate unfavorable results. Values to the right, represented in green, indicate favorable results. Thus, increases in decline are displayed on the left, and decreases in decline are displayed on the right. The p-value for functional status results indicates joint significance for differences in proportion of respondents indicating one of three categories: improvement (or maintained highest function); stayed the same; or declined (or maintained lowest function). See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results. PGP = physician group practice; pp = percentage point(s).

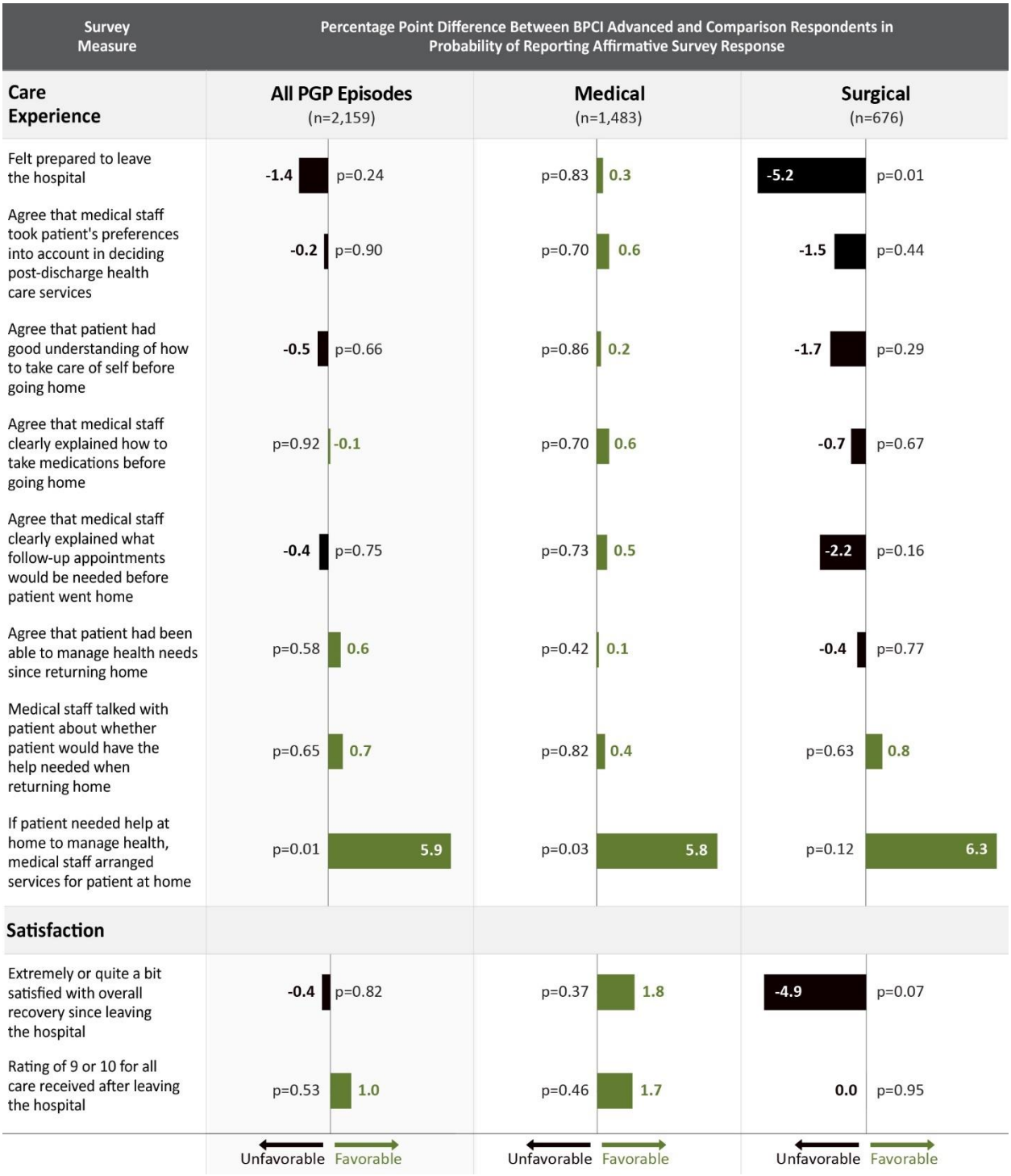
Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Were patient-reported care experiences and satisfaction similar between BPCI Advanced respondents with PGP-initiated episodes and comparison respondents?

For all PGP episodes, differences between BPCI Advanced and comparison respondents for most measures of care experiences were small (less than or equal to 1.0 pp) with roughly half favorable and half unfavorable during Model Year 4 (Exhibit 22). However, one outcome was favorable and statistically significant: BPCI Advanced respondents with PGP-initiated episodes were 5.9 pp more likely (90% confidence interval: 2.4, 9.5 pp; p=0.01) than comparison respondents to report that medical staff arranged services to help manage their health at home, if they needed such help. This outcome was favorable and of similar magnitude for respondents with PGP medical episodes and for respondents with PGP surgical episodes. For PGP medical episodes, differences for the remaining seven care experience outcomes were small, and all were favorable. However, results vary for the two PGP medical CESLGs analyzed (Exhibit 24). For the *medical and critical care* CESLG, BPCI Advanced respondents were more likely than comparison respondents to report favorable care experience outcomes for all eight measures and one was statistically significant: BPCI Advanced respondents were 6.3 pp more likely to say that if they needed help at home managing their health, medical staff arranged services to help (90% confidence interval: 0.8, 11.7; p=0.06). For the *cardiac care* CESLG, BPCI Advanced respondents were less likely than comparison respondents to report favorable care experience outcomes for all eight measures, although none were statistically significant. For PGP surgical episodes, the results for the other seven care experience measures suggest that BPCI Advanced respondents had slightly less favorable outcomes, with one statistically significant result: BPCI Advanced respondents were 5.2 pp less likely to feel prepared to leave the hospital than were comparison respondents (90% confidence interval: -8.6, -1.8; p=0.01).

BPCI Advanced and comparison respondents with PGP-initiated episodes were equally likely to report the highest levels of satisfaction with care. The lack of differences in the probability of reporting the highest levels of satisfaction with overall recovery was due to offsetting favorable and unfavorable results in medical and surgical CESLGs. For example, BPCI Advanced respondents in the *medical and critical care* CESLG, the highest-volume CESLG overall, were 4.7 pp more likely (90% confidence interval: 0.6, 8.8; p=0.06) than comparison respondents to report the highest levels of satisfaction with recovery. In contrast, BPCI Advanced respondents in the *orthopedics* CESLGs were 5.7 pp less likely (90% confidence interval: -11.0, -0.3; p=0.08) than comparison respondents to report the highest levels of satisfaction with recovery (Exhibit 24).

Exhibit 22: Differences in Patient-Reported Care Experiences and Satisfaction with Care Between BPCI Advanced and Comparison Respondents, PGPs, July 2021 – August 2021



***Note*:** The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for binary indicators. All responses were weighted for non-response and sampling design. Estimates were based on 2,159 BPCI Advanced survey respondents and 1,900 comparison survey respondents across all 34 clinical episodes. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results reported in percentage point terms. Values to the left, represented in black, indicate unfavorable results. Values to the right, represented in green, indicate favorable results. The p-value for satisfaction with post-discharge care indicates joint significance for differences in proportion of respondents indicating one of three categories: 9-10

rating; 7-8 rating; 0-6 rating. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results. PGP = physician group practice; pp = percentage point(s).

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Exhibit 23: Differences in Patient-Reported Change in Functional Status Between BPCI Advanced and Comparison Respondents for Individual Clinical Episode Service Line Groups, PGPs, July 2021-August 2021

Domain	Measures	Response if indicator = 1	Medical Episodes		Surgical Episodes
			Medical & Critical Care (n=805)	Cardiac Care (n=414)	Orthopedics (n=457)
Functional Status and Activities of Daily Living	Bathing, dressing, using the toilet, or eating	Improvement	**-2.3****	3.7	-1.9
		Decline	**-2.5**	0.5	1.5
	Planning regular tasks	Improvement	-0.1	**2.9***	0.9
		Decline	0.1	**2.7**	1.7
	Use of mobility device	Improvement	-0.7	-1.7	2.8
		Decline	2.9	1.9	-1.7
	Walking without rest	Improvement	-1.2	-5.9	1.5
		Decline	0.4	2.9	-0.5
	Going up or down stairs	Improvement	-0.2	**-7.3***	-1.2
		Decline	-1.6	**2.7**	4.9
	Physical/emotional problems limiting social activities	Improvement	-1.2	0.4	-4.1
		Decline	-0.9	3.0	3.1
	Pain limiting regular activities	Improvement	2.0	0.7	**8.8*****
		Decline	1.3	-2.6	**-2.5**

Note: The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. All responses were weighted for non-response and sampling design. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results are reported in percentage point terms. Green shading indicates favorable results. Pink shading indicates unfavorable results. Grey shading indicates neutral results. If differences in improvement and decline indicated opposite effects (e.g., more improvement *and* more decline) and the net difference was less than 1.0 pp, differences were considered neutral. If the net difference exceeded 1.0 pp, then it was considered favorable (if more improvement/less decline) or unfavorable (if less improvement/more decline) on net. All decline measures are unshaded to help visually distinguish outcomes. We analyzed all service-line groups for which we had a minimum detectable difference of 10.0 pp or less. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results. PGP = physician group practice; pp = percentage point(s). *p<0.10 **p<0.05 ***p<0.01

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Exhibit 24: Differences in Patient-Reported Care Experiences and Satisfaction with Care Between BPCI Advanced and Comparison Respondents for Individual Clinical Episode Service Line Groups, PGPs, July 2021-August 2021

Domain	Measures	Response if indicator = 1	Medical Episodes		Surgical Episodes
			Medical & Critical Care (n=805)	Cardiac Care (n=414)	Orthopedics (n=457)
Care Experiences	How prepared did you feel to leave the hospital?	Very/Somewhat Prepared	1.2	-1.4	**-8.6*** **
	Medical staff took your preferences into account in deciding what health care services you should have after you left the hospital	Agree/Strongly Agree	1.3	-1.7	-1.9
	Good understanding of how to take care of self before going home	Agree/Strongly Agree	1.0	-2.5	**-3.3***
	Medical staff clearly explained how to take medications before going home	Agree/Strongly Agree	1.3	-3.6	-2.6
	Medical staff clearly follow-up appointments or treatments would be needed	Agree/Strongly Agree	1.5	-3.0	-2.8
	Able to manage your health needs since returning home	Agree/Strongly Agree	1.6	-0.3	-1.8
	Medical staff talked with you about whether you would have the help you needed when you got home	Yes	0.5	-1.6	1.1
	If you needed help at home to manage your health, medical staff arranged services for you at home	Yes	**6.3***	-1.9	**8.9***
Satisfaction with Care	Overall satisfaction with recovery since leaving hospital	Extremely Satisfied/Quite a bit Satisfied	**4.7***	-2.0	**-5.7***
	Rating of all care received after leaving the hospital	Rating 9-10	2.4	-3.0	3.4
		Rating 0-6	1.9	0.1	-0.8

Note: The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for binary indicators. All responses were weighted for non-response and sampling design. Reported sample sizes reflect the number of BPCI Advanced survey respondents. Results are reported in percentage point terms. Green shading indicates favorable results. Pink shading indicates unfavorable results. Grey shading indicates neutral results. Results were judged neutral if the point estimate rounded to 0.0, indicating a difference less than 0.05 pp. We analyzed all service-line groups for which we had a minimum detectable difference of 10.0 pp or less. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix I** for more detailed results. PGP = physician group practice; pp = percentage point(s). *p<0.10 **p<0.05 ***p<0.01

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

B. Medicare Program Savings

This section presents estimates of Medicare program savings for Model Year 3 (2020) of the BPCI Advanced Model. We calculated net Medicare savings and losses for episodes pooled across the clinical episodes evaluated, for medical and surgical clinical episodes, and for medical and surgical episodes by episode initiator type (i.e., hospital medical clinical episodes, PGP medical clinical episodes, hospital surgical clinical episodes, and PGP surgical clinical episodes). We also calculated savings and losses for each clinical episode and episode initiator type for which we evaluated impact estimates.[27] Medicare savings estimates differ from the episode payment impact estimates because they account not only for Medicare FFS payments to providers but also for the Net Payment Reconciliation Amounts (NPRA) or repayment amounts, collectively referred to as reconciliation payments made to or received from participants.

Net Medicare savings due to the BPCI Advanced Model equals the reduction in non-standardized episode payments minus reconciliation payments paid to or received from participants. We calculated the reduction in non-standardized payments by converting the difference-in-differences (DiD) impact estimates based on standardized Medicare paid amounts to non-standardized payments.[28] Total reconciliation payments for relevant clinical episodes during the period were then subtracted to obtain net Medicare savings expressed in dollars. To calculate per-episode savings, we divided net savings, expressed in dollars, by the count of BPCI Advanced episodes in the Model Year 3 intervention period used in the DiD impact estimates.

The count of BPCI Advanced intervention episodes from the evaluation is larger than the count of BPCI Advanced intervention episodes from reconciliation for two main reasons. First, when faced with overlapping models or overlapping episodes, the reconciliation methodology excludes episodes to avoid paying out the same savings to participants of multiple models. In the evaluation, however, we retain episodes that overlap and calculate the incremental impact of the BPCI Advanced Model over and above other models' impacts, or in the case of overlapping BPCI Advanced episodes, we prorate the episode across multiple participants and clinical episodes. Second, in response to the COVID-19 PHE, CMS allowed BPCI Advanced Model Year 3 participants to select one of two participation agreement amendments, 1) to forgo reconciliation for all episodes in Model Year 3, or 2) to only exclude episodes with a COVID-19 diagnosis from reconciliation. Participants could also choose not to elect either amendment. Episodes were excluded from reconciliation based on participants' choice of amendments, but the evaluation retained all episodes. These two factors, plus a few additional factors that differ between the reconciliation and evaluation samples,[29] result in a count of BPCI Advanced intervention episodes in the evaluation that is approximately 55% larger than the count of BPCI Advanced episodes used to calculate reconciliation payments. In this way, the evaluation

[27] There was sufficient sample size to evaluate 17 out of 34 clinical episodes for hospital episode initiators and 17 out of 34 clinical episodes for PGP episode initiators.

[28] Non-standardized Medicare paid amounts reflect actual FFS payments from Medicare to providers for services furnished, as they include adjustments for wages, practice expenses, and other initiatives (e.g., medical education), and they exclude beneficiary cost sharing. See **Appendix C** for more details.

[29] For example, the reconciliation sample excludes episodes for beneficiaries aligned to certain Accountable Care Organizations (ACOs) and it excludes episodes from participants after they withdrew, while the evaluation retains both types of episodes.

captures the impact of the model on a broader set of episodes than the set of episodes considered in reconciliation.

The BPCI Advanced Model was designed to save the Medicare program 3% of payments under the counterfactual, what payments would have been if the BPCI Advanced Model had not occurred. The target prices for Model Years 1 and 2 (2018 and 2019) were based on historical payments from January 2013 through December 2016 (target price baseline period). In Model Year 3, the target price baseline period was October 1, 2014 through September 30, 2018. To assess how well the model achieved its goals, we express Medicare program savings estimates and the components of savings (reduction in non-standardized payments and reconciliation payments) as a percentage of the evaluation's estimate of the counterfactual, which is calculated as the average BPCI Advanced non-standardized episode payment in the baseline plus the change in the average non-standardized episode payment for the comparison group from baseline to intervention. See **Appendix C** for additional details on the definitions and calculations of savings.

1. Key Findings

Medicare Program Savings Under BPCI Advanced

- During Model Year 3 (2020), the BPCI Advanced Model resulted in an estimated net loss to the Medicare program of $113.7 million, or 0.8% of Medicare payments under the counterfactual (what Medicare payments would have been if the BPCI Advanced Model had not occurred), ranging from a loss of $20.1 million to $207.4 million (loss of 0.1% to a loss of 1.5%) based on a 90% confidence interval.
- Overall, for both hospital and physician group practice (PGP) episode initiators, the BPCI Advanced Model generally resulted in estimated net losses for medical clinical episodes and estimated net savings for surgical clinical episodes.
- For medical clinical episodes, the model resulted in an estimated net loss of $200.5 million, or 1.9%.
- For surgical clinical episodes, the model resulted in an estimated net savings of $71.3 million, or 2.3%.
- For both hospitals and PGPs, the evidence suggests that target prices may have been too high for medical clinical episodes but were generally more appropriate for surgical clinical episodes.

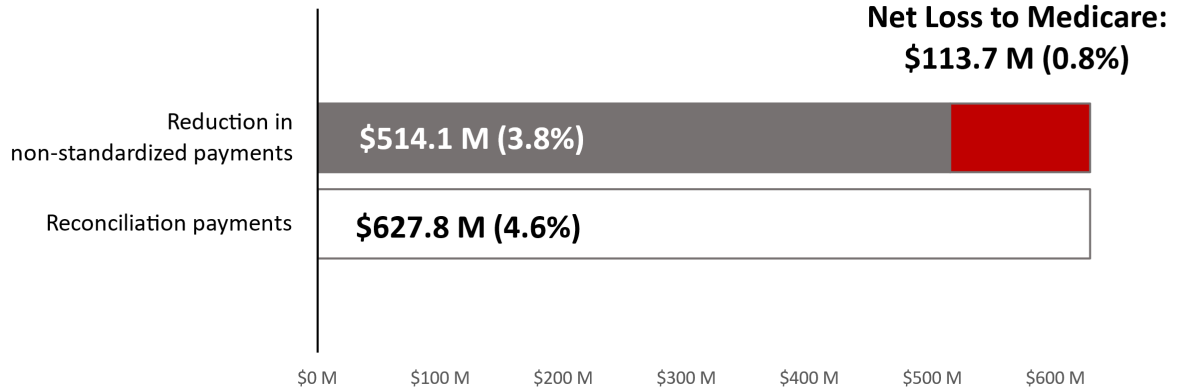
2. Did BPCI Advanced result in savings to Medicare during Model Year 3 (2020)?

a. Pooled Clinical Episodes

During Model Year 3, the BPCI Advanced Model reduced non-standardized episode payments by an estimated $514.1 million, or about 3.8% of payments under the counterfactual (that is, what Medicare payments would have been if the BPCI Advanced Model had not occurred) (Exhibit 25). After accounting for $627.8 million in reconciliation payments, or 4.6% of

payments under the counterfactual, the model resulted in an estimated net loss of $113.7 million to Medicare, or 0.8%. That is, Medicare spending increased by an estimated $113.7 million due to the BPCI Advanced Model, equivalent to an increase of $227 per episode. When considering the 90% confidence interval of our DiD impact estimate, net Medicare spending ranged from a loss of $20.1 million to $207.4 million, or 0.1% to 1.5%. See **Appendix J** for detailed results of the Medicare savings analysis.

Exhibit 25: Medicare Savings due to BPCI Advanced, Model Year 3 (January 1, 2020 – December 31, 2020)

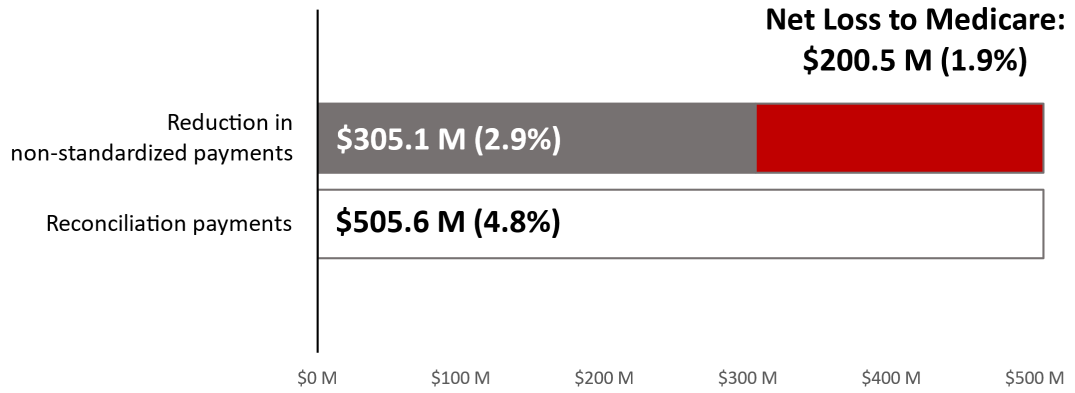


***Note*:** Medicare savings and its components are calculated only for the clinical episodes evaluated, which account for 91.1% of all episodes initiated by hospital episode initiators and 94.0% of all episodes initiated by physician group practice episode initiators. The estimated reduction in non-standardized payments is based on difference-in-differences (DiD) models of standardized Medicare paid amounts. Net loss to Medicare is the estimated reduction in non-standardized payments minus reconciliation payments. The estimates are also presented as a percentage of payments under the counterfactual, or what payments would have been if the BPCI Advanced Model had not occurred, which is estimated as the average BPCI Advanced baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix J** for detailed results of net Medicare savings. M = million.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period), and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers and CMS reconciliation data from the same period.

For medical clinical episodes, the BPCI Advanced Model reduced non-standardized episode payments by an estimated $305.1 million, or 2.9% of payments under the counterfactual (Exhibit 26). After accounting for $505.6 million in reconciliation payments, or 4.8%, the model resulted in an estimated net loss of $200.5 million, or 1.9%, for medical clinical episodes. That is, Medicare spending on medical clinical episodes increased by an estimated $200.5 million due to the BPCI Advanced Model, equivalent to an increase of $506 per episode. When considering the 90% confidence interval of our DiD impact estimate, net Medicare spending for medical clinical episodes ranged from a loss of $128.2 million to $272.8 million, or 1.2% to 2.6%.

Exhibit 26: Medicare Savings due to BPCI Advanced, Medical Clinical Episodes, Model Year 3 (January 1, 2020 – December 31, 2020)

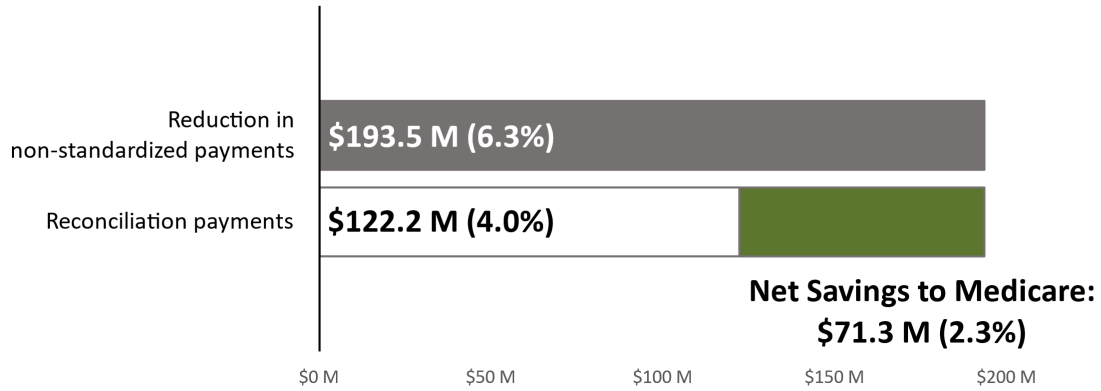


Note: The estimated reduction in non-standardized payments is based on difference-in-differences (DiD) models of standardized Medicare paid amounts for evaluated clinical episodes. Net loss to Medicare is the estimated reduction in non-standardized payments minus reconciliation payments. The estimates are also presented as a percentage of payments under the counterfactual, or what payments would have been if the BPCI Advanced Model had not occurred, which is estimated as the average BPCI Advanced baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. The estimates of Medicare savings for the subgroups may not sum to the estimate of Medicare savings for the pooled clinical episodes because each estimate is derived from a different DiD model. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix J** for detailed results of net Medicare savings. M = million.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period), and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers and CMS reconciliation data from the same period.

For surgical clinical episodes, the BPCI Advanced Model reduced non-standardized episode payments by an estimated $193.5 million, or about 6.3% of payments under the counterfactual (Exhibit 27). After accounting for $122.2 million in reconciliation payments, or 4.0%, the model resulted in an estimated net savings of $71.3 million, or 2.3%, for surgical clinical episodes. That is, Medicare spending on surgical clinical episodes decreased by an estimated $71.3 million due to the BPCI Advanced Model, which is equivalent to a decrease of $678 per episode. When considering the 90% confidence interval of our DiD impact estimate, net savings for surgical clinical episodes ranged from a savings of $31.4 million to $111.3 million, or 1.0% to 3.6%.

Exhibit 27: Medicare Savings due to BPCI Advanced, Surgical Clinical Episodes, Model Year 3 (January 1, 2020 – December 31, 2020)



***Note*:** The estimated reduction in non-standardized payments is based on difference-in-differences (DiD) models of standardized Medicare paid amounts for evaluated clinical episodes. Net savings to Medicare is the estimated reduction in non-standardized payments minus reconciliation payments. The percentage estimates are a percentage of payments under the counterfactual, or what payments would have been if the BPCI Advanced Model had not occurred, which is estimated as the average BPCI Advanced baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. The estimates of Medicare savings for the subgroups may not sum to the estimate of Medicare savings for the pooled clinical episodes because each estimate is derived from a different DiD model. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix J** for detailed results of net Medicare savings. M = million.

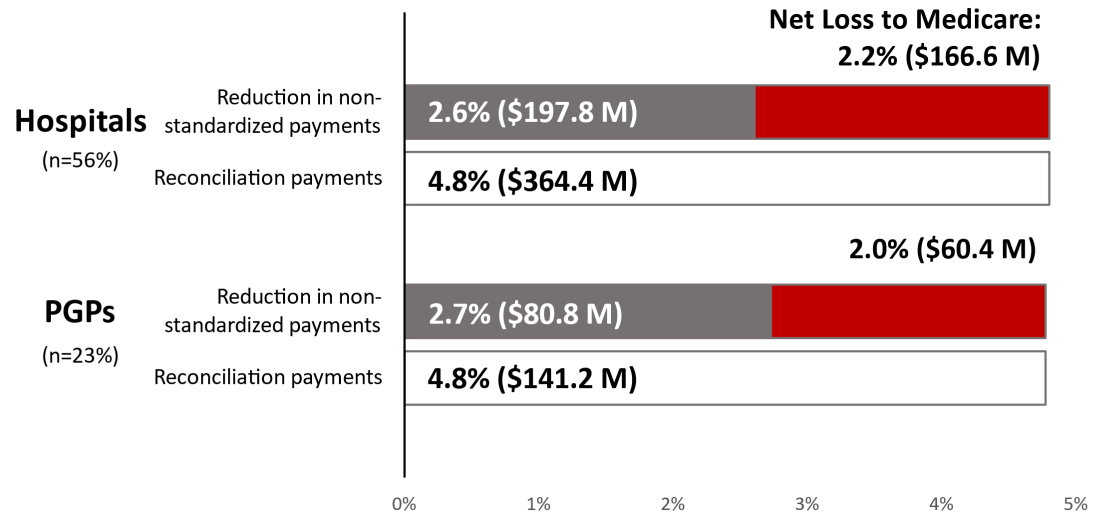
Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period), and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers and CMS reconciliation data from the same period.

b. Medical Clinical Episodes for Hospitals and Physician Group Practices

For hospital medical clinical episodes, the BPCI Advanced Model reduced non-standardized episode payments by an estimated $197.8 million, or about 2.6% of payments under the counterfactual (Exhibit 28). After accounting for $364.4 million in reconciliation payments, or 4.8%, the model resulted in an estimated net loss of $166.6 million, or 2.2%, for hospital medical clinical episodes. That is, Medicare spending on hospital medical clinical episodes increased by an estimated $166.6 million due to the BPCI Advanced Model, which is equivalent to an increase of $597 per episode. When considering the 90% confidence interval of our DiD impact estimate, net Medicare spending for hospital medical clinical episodes ranged from a loss of $111.1 million to $222.0 million, or 1.5% to 2.9%.

For PGP medical clinical episodes, the BPCI Advanced Model reduced non-standardized episode payments by an estimated $80.8 million, or about 2.7% of payments under the counterfactual. After accounting for $141.2 million in reconciliation payments, or 4.8%, the model resulted in an estimated net loss of $60.4 million, or 2.0%, for PGP medical clinical episodes. That is, Medicare spending on PGP medical clinical episodes increased by an estimated $60.4 million due to the BPCI Advanced Model, which is equivalent to an increase of $513 per episode. When considering the 90% confidence interval of our DiD impact estimate, net Medicare spending for PGP medical clinical episodes ranged from a loss of $17.5 million to $103.4 million, or a loss of 0.6% to 3.5%.

Exhibit 28: Medicare Savings due to BPCI Advanced, Medical Clinical Episodes for Hospital and PGP Episode Initiators, Model Year 3 (January 1, 2020 – December 31, 2020)



***Note*:** The estimated reduction in non-standardized payments is based on difference-in-differences (DiD) models of standardized Medicare paid amounts for evaluated clinical episodes. Net loss to Medicare is the estimated reduction in non-standardized payments minus reconciliation payments. The percentage estimates are a percentage of payments under the counterfactual, or what payments would have been if the BPCI Advanced Model had not occurred, which is estimated as the average BPCI Advanced baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. The sample size (n=) refers to the percentage of total episode volume for all clinical episodes used to calculate the reduction in non-standardized payments. The estimates of Medicare savings for the subgroups may not sum to the estimate of Medicare savings for the pooled clinical episodes because each estimate is derived from a different DiD model. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix J** for detailed results of net Medicare savings. M = million; PGP = physician practice group.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period), and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers and CMS reconciliation data from the same period.

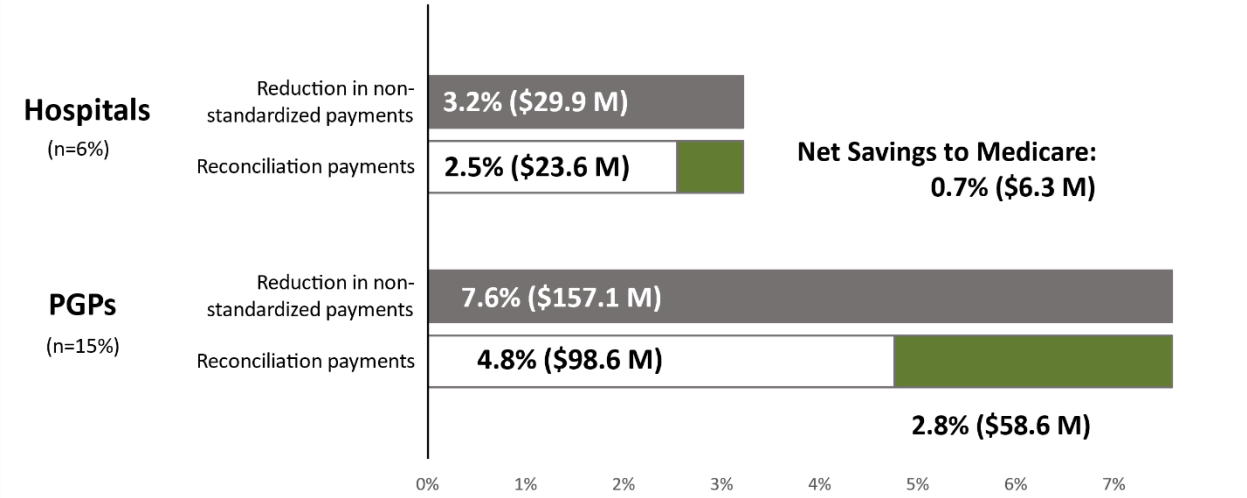
c. Surgical Clinical Episodes for Hospitals and Physician Group Practices

For hospital surgical clinical episodes, the BPCI Advanced Model reduced non-standardized episode payments by an estimated $29.9 million, or about 3.2% of payments under the counterfactual (Exhibit 29). After accounting for $23.6 million in reconciliation payments, or 2.5%, the model resulted in an estimated net savings of $6.3 million, or 0.7%, for hospital surgical clinical episodes. That is, Medicare spending on hospital surgical clinical episodes decreased by an estimated $6.3 million due to the BPCI Advanced Model, which is equivalent to a decrease of $200 per episode. When considering the 90% confidence interval of our DiD impact estimate, net Medicare spending for hospital surgical clinical episodes ranged from a loss of $14.1 million to a savings of $26.6 million, or a loss of 1.5% to a savings of 2.9%.

For PGP surgical clinical episodes, the BPCI Advanced Model reduced non-standardized episode payments by an estimated $157.1 million, or about 7.6% of payments under the counterfactual. After accounting for $98.6 million in reconciliation payments, or 4.8%, the model resulted in an estimated net savings of $58.6 million, or 2.8%, for PGP surgical clinical episodes. That is, Medicare spending on PGP surgical clinical episodes decreased by an estimated $58.6 million due to the BPCI Advanced Model, which is equivalent to a decrease of $793 per episode. When

considering the 90% confidence interval of our DiD impact estimate, net savings for PGP surgical clinical episodes ranged from a savings of $26.7 million to a savings of $90.5 million, or a savings of 1.3% to a savings of 4.4%.

Exhibit 29: Medicare Savings due to BPCI Advanced, Surgical Clinical Episodes for Hospital and PGP Episode Initiators, Model Year 3 (January 1, 2020 – December 31, 2020)



Note: The estimated reduction in non-standardized payments is based on difference-in-differences (DiD) models of standardized Medicare paid amounts for evaluated clinical episodes. Net savings to Medicare is the estimated reduction in non-standardized payments minus reconciliation payments. The percentage estimates are a percentage of payments under the counterfactual, or what payments would have been if the BPCI Advanced Model had not occurred, which is estimated as the average BPCI Advanced baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. The sample size (n=) refers to the percentage of total episode volume for all clinical episodes used to calculate the reduction in non-standardized payments. The estimates of Medicare savings for the subgroups may not sum to the estimate of Medicare savings from the pooled clinical episodes because each estimate is derived from a different DiD model. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix J** for detailed results of net Medicare savings. M = million; PGP = physician group practice.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period), and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers and CMS reconciliation data from the same period.

3. Did estimates of Medicare program savings in Model Year 3 (2020) differ from Model Years 1 and 2 (2018-2019)?

The BPCI Advanced Model resulted in small net losses to Medicare in both Model Year 3 and Model Years 1 and 2, but compared to Model Years 1 and 2, Medicare losses were slightly larger in Model Year 3 (Exhibit 30). Losses were $113.7 million, or 0.8% of payments under the counterfactual for Model Year 3 compared to $65.7 million, or 0.4% for Model Years 1 and 2. See **Appendix J** for detailed results of the Medicare savings analysis.

Although the declines in average episode payments were larger in Model Year 3, episode volume was lower, resulting in a slightly smaller decline in non-standardized payments in Model Year 3 than in Model Years 1 and 2. In addition, reconciliation payments per episode to participants were larger in Model Year 3 compared to Model Years 1 and 2, and despite lower episode volume, total reconciliation payments were larger in Model Year 3.

For medical clinical episodes, the model resulted in an estimated net loss of $200.5 million, or 1.9% of payments under the counterfactual, which was smaller than the estimated net loss in Model Years 1 and 2 of $275.0 million, or 2.2%. The declines in average episode payments were larger in Model Year 3 than in Model Years 1 and 2 for medical clinical episodes, and although episode volume was smaller, there was a larger decline in non-standardized payments in Model Year 3 than in Model Years 1 and 2. In addition, reconciliation payments to participants in medical clinical episodes were slightly smaller in Model Year 3 compared to Model Years 1 and 2.

For surgical clinical episodes, the model resulted in an estimated net savings of $71.3 million, or 2.3%, which was smaller than the estimated net savings in Model Years 1 and 2 of $204.4 million, or 3.6%. Although the declines in average episode payments were larger in Model Year 3 than in Model Years 1 and 2 for surgical clinical episodes, episode volume was lower, resulting in a smaller decline in non-standardized payments in Model Year 3 than in Model Years 1 and 2. In addition, reconciliation payments to participants were larger in Model Year 3 compared to Model Years 1 and 2 for surgical clinical episodes, which was primarily driven by reconciliation payments to hospital participants for *major joint replacement of the lower extremity* episodes. On net, CMS made reconciliation payments to hospital participants in Model Year 3 for *major joint replacement of the lower extremity* episodes, whereas hospital participants owed amounts to CMS in Model Years 1 and 2.

Exhibit 30: Medicare Savings, BPCI Advanced Hospitals and PGPs, Model Years 1-2 vs. Model Year 3, October 1, 2018 – December 31, 2020

Clinical Episodes	Model Years 1 and 2 [a]			Model Year 3		
	Reduction in Non-standardized Payments	Reconciliation Payments	Savings to Medicare	Reduction in Non-standardized Payments	Reconciliation Payments	Savings to Medicare
All Clinical Episodes	$550.7 M	$616.5 M	-$65.7 M	$514.1 M	$627.8 M	-$113.7 M
Medical	$252.5 M	$527.5 M	-$275.0 M	$305.1 M	$505.6 M	-$200.5 M
Hospitals	$164.2 M	$406.1 M	-$241.9 M	$197.8 M	$364.4 M	-$166.6 M
PGPs	$73.0 M	$121.3 M	-$48.3 M	$80.8 M	$141.2 M	-$60.4 M
Surgical	$293.4 M	$89.0 M	$204.4 M	$193.5 M	$122.2 M	$71.3 M
Hospitals	$43.8 M	$2.3 M	$41.5 M	$29.9 M	$23.6 M	$6.3 M
PGPs	$243.8 M	$86.7 M	$157.1 M	$157.1 M	$98.6 M	$58.6 M

***Note*:** The estimated reduction in non-standardized payments is based on difference-in-differences (DiD) models of standardized Medicare paid amounts for evaluated clinical episodes. The estimates for the two episode initiator types (hospitals and PGPs) do not sum to the pooled estimate for a given clinical episode type (medical and surgical), and the estimates for medical and surgical do not sum to the estimate for all clinical episodes because each estimate is derived from a different DiD model. Net savings to Medicare is the estimated reduction in non-standardized payments minus reconciliation payments. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix J** for detailed results of net Medicare savings. M = million; PGP = physician group practice.

[a] The BPCI Advanced Third Evaluation Report is available for download at https://innovation.cms.gov/innovation-models/bpci-advanced.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period), episodes with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2019 (intervention period for Model Years 1 and 2), and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period for Model Year 3) for BPCI Advanced episode initiators and matched comparison providers and CMS reconciliation data from the same period.

4. Were target prices in the BPCI Advanced Model during Model Year 3 (2020) appropriate?

Target prices in the BPCI Advanced Model were constructed such that the Medicare program would save 3% of payments under the counterfactual, or what payments would have been if the BPCI Advanced Model had not occurred. However, as explained by a former CMMI Director in a 2021 letter to the *New England Journal of Medicine*, constructing accurate target prices can be difficult because it requires projecting future costs for the episodes, and costs can change over time and vary across geographic regions.[30] It is also more challenging to construct accurate target prices *prospectively*—using data from the baseline period to forecast future peer group trends which can be disrupted by unforeseen practice or policy changes—rather than *retrospectively*—using realized peer group trends which account for changes during the performance period. To improve target pricing accuracy, CMS made changes to the target price construction beginning in Model Year 4 (2021), such as a change from prospectively trended target prices in Model Years 1 through 3 (2018 through 2020) to include retrospective trend adjustments in Model Year 4.

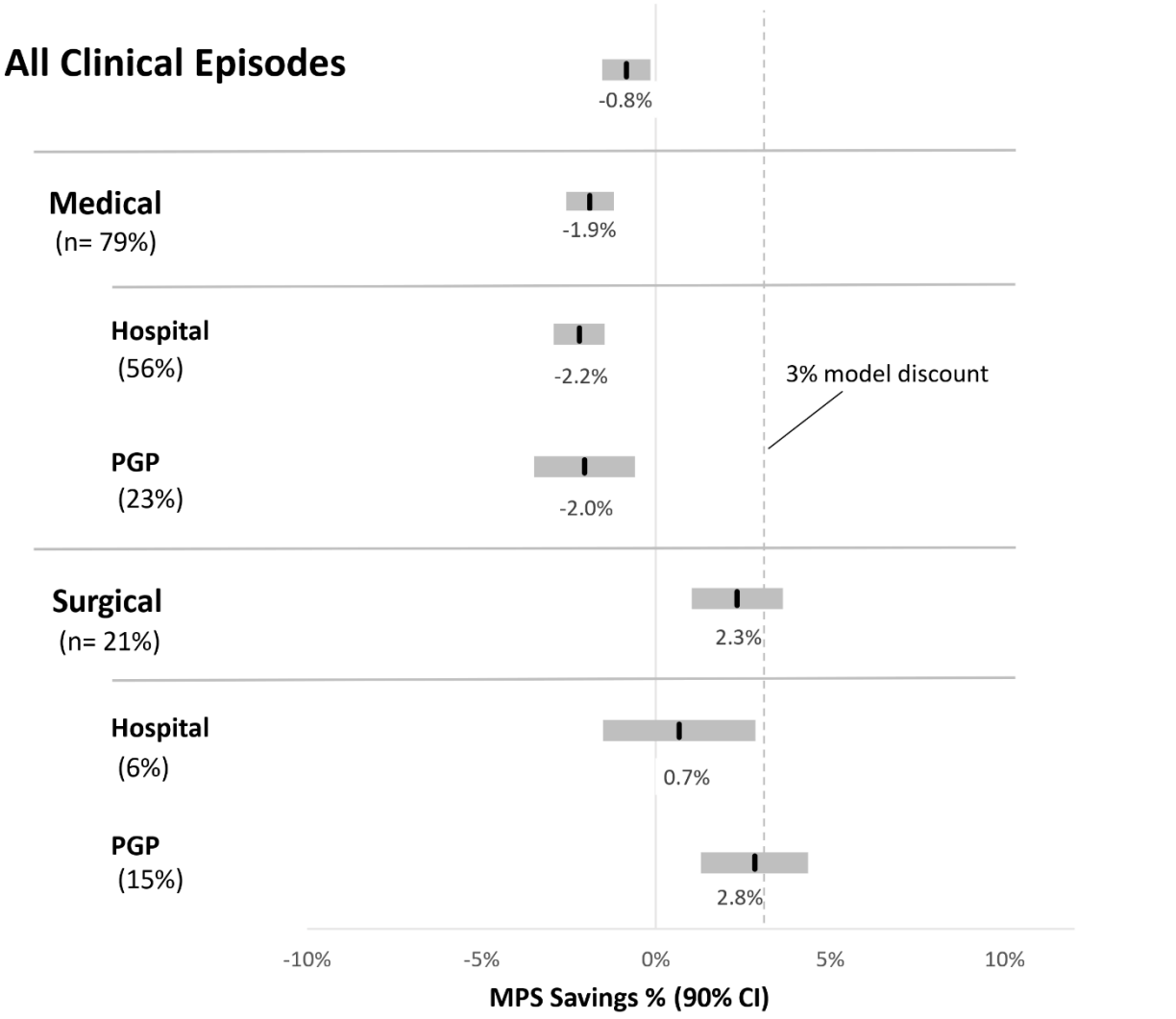
Exhibit 31 displays estimates of net savings to the Medicare program as a percentage of payments under the counterfactual and compares those estimates to the 3% savings goal. If 3% falls within the 90% confidence intervals of the net savings estimates, then we conclude there is evidence that the target prices were appropriate. If the 90% confidence intervals fall below 3%, then the evidence suggests target prices may have been too high. If the 90% confidence intervals fall above 3%, then the evidence suggests target prices may have been too low.

The evidence suggests that target prices may have been too high for hospital and PGP medical clinical episodes and for hospital surgical clinical episodes, because the BPCI Advanced Model resulted in estimated losses with 90% confidence intervals that fall below the 3% savings goal. In Model Years 1 and 2, target prices may have been too high for hospital and PGP medical clinical episodes.

The evidence suggests that target prices may have been reasonably accurate for PGP surgical clinical episodes. We estimate that the BPCI Advanced Model resulted in savings greater than 3% for PGP surgical clinical episodes, but the 90% confidence intervals around the estimates include 3%.

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

Exhibit 31: Medicare Savings Compared to the 3% Model Discount, Model Year 3 (January 1, 2020 – December 31, 2020)



Note: Net savings to Medicare is the estimated reduction in non-standardized payments minus reconciliation payments. The estimates are presented as a percentage of payments under the counterfactual, or what payments would have been if the BPCI Advanced Model had not occurred, which is estimated as the average BPCI Advanced baseline payment amount plus the average change in the episode payment amount for the comparison group from baseline to intervention. The grey bars indicate the 90% confidence interval. The confidence intervals associated with the estimates of net Medicare program savings are based on the estimates of the change in non-standardized payments from the difference-in-differences (DiD) models. The grey dashed line at the 3% mark indicates the 3% model discount. The sample size (n=) refers to the percentage of total episode volume for all clinical episodes used to calculate the reduction in non-standardized payments. See **Appendix C** for details of the DiD methodology, outcome definitions, and additional information on methods. See **Appendix J** for detailed results of net Medicare savings.
CI = confidence interval; MPS = Medicare program savings; PGP = physician group practice.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period), and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers and CMS reconciliation data from the same period.

C. Beneficiaries from Populations that have been Historically Underserved

The BPCI Advanced Model was designed to reduce spending while improving the quality of care. More recently, CMMI launched a strategy refresh, introduced in 2021, which seeks to promote equitable outcomes through high-quality, affordable, person-centered care, with a special focus on underserved communities.[31,32] One of the five strategic objectives is to advance health equity, and new models will be designed to reduce inequities in health care outcomes. We have begun to assess health equity in the BPCI Advanced Model for the first time in this evaluation report. Because the model was not designed to address health equity, we assess health equity as a potential unintended consequence of the model.

We conducted analyses on subgroups of the population to determine whether the sample of BPCI Advanced beneficiaries is representative of historically underserved populations and to evaluate whether outcomes under BPCI Advanced differed for beneficiaries from historically underserved populations. For our analyses, we consider underserved populations that are identifiable in the data and have sufficient sample size. For the analyses of the representativeness of the BPCI Advanced sample and of claims-based outcomes, we examined outcomes in Model Years 1 and 2 and Model Year 3 for Black or African American beneficiaries and beneficiaries dually eligible for Medicare and Medicaid. For the beneficiary survey, we examined outcomes in Model Year 4 for the following subsets of beneficiaries: Black or African American, Hispanic, dually eligible, living in a ZIP code in the top 20% of the Area Deprivation Index (ADI),[33] and living in a rural ZIP code. Differences between BPCI Advanced and comparison respondents within each underserved population were compared to differences between BPCI Advanced and comparison respondents within a corresponding reference population. Non-Hispanic White beneficiaries are the reference population for Black or African American beneficiaries; nondual-eligible beneficiaries are the reference population for dual-eligible beneficiaries; beneficiaries living in a ZIP code in the bottom 80% of the ADI are the reference population for beneficiaries living in a ZIP code in the top 20% of the ADI, non-rural beneficiaries are the reference population for rural beneficiaries. This allowed us to start understanding the extent to which BPCI Advanced may have exacerbated or mitigated differences between underserved populations relative to other groups.

[31] See https://innovation.cms.gov/strategic-direction for more information about CMMI's strategy refresh.

[32] The Racial Equity and Support for Underserved Communities Through the Federal Government Executive Order 13985, states: "The term 'equity' means the consistent and systematic fair, just, and impartial treatment of all individuals, including individuals who belong to underserved communities that have been denied such treatment, such as Black, Latino, and Indigenous and Native American persons, Asian Americans and Pacific Islanders and other persons of color; members of religious minorities; lesbian, gay, bisexual, transgender, and queer (LGBTQ+) persons; persons with disabilities; persons who live in rural areas; and persons otherwise adversely affected by persistent poverty or inequality. The term 'underserved communities' refers to populations sharing a particular characteristic, as well as geographic communities, that have been systematically denied a full opportunity to participate in aspects of economic, social, and civic life, as exemplified by the list in the preceding definition of 'equity.'" The executive order is available for download at https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/.

[33] The ADI is a measure of neighborhood socioeconomic disadvantage. Neighborhoods with a higher ADI have lower neighborhood measures of income, education, employment, and housing quality. The ADI score, measured at the nine-digit zip code level, is defined in percentiles of the national population. Our subpopulation defined by ADI>=80 represents beneficiaries who reside in the top fifth most deprived neighborhoods in the country.

Our analysis of outcomes for beneficiaries of certain race and ethnicity categories is limited by data availability and sample size. We categorize beneficiaries based on the RTI race code available through the CMS' enrollment files. See **Appendix C** for further detail on this measure.

1. What was the representation under BPCI Advanced of beneficiaries from populations that have been historically underserved?

To understand whether the BPCI Advanced Model reached historically underserved populations, we measured the composition of the FFS beneficiaries with medical and surgical stays as well as whether imposing different model rules changed the composition of the sample. Specifically, we looked at the set of beneficiaries who had a Medicare-paid hospital discharge or outpatient procedure with a BPCI Advanced triggering Healthcare Common Procedure Coding System (HCPCS) code during Model Years 1, 2, or 3 (2018-2020).[34] In a stepwise fashion, we limited this sample to include only beneficiaries that had hospital discharges with a BPCI Advanced Medicare Severity-Diagnosis Related Group (MS-DRG) as the primary diagnosis or an outpatient procedure with BPCI Advanced HCPCS codes, then to include only eligible discharges based on model criteria, and then only episodes attributed to BPCI Advanced participants in financial reconciliation.

[34] Given the vast number of Medicare outpatient procedures that occur annually, we limit the outpatient procedures used in this analysis to only those with a BPCI Advanced HCPCS code for data processing purposes. All Medicare inpatient discharges are included in the analysis.

a. Key Findings

Representation Under BPCI Advanced for Underserved Populations

- During Model Years 1 through 3, the representation of Black or African American beneficiaries among BPCI Advanced episodes was slightly lower than the share in the Medicare fee-for-service (FFS) population (8.7% compared to 9.3%), while the representation of dual-eligible beneficiaries among BPCI Advanced episodes was higher than the share in the FFS population (21.2% compared to 17.6%).
- There were no substantial differences in representation based on restrictions in the design of the model (such as requiring beneficiary Part A and B coverage, excluding Maryland providers), suggesting differences in representation stem from the underlying FFS system. Compared to their relative shares in the population of FFS beneficiaries who had a hospital discharge or procedure with BPCI Advanced triggering MS-DRGs or HCPCS codes, the share of beneficiaries with BPCI Advanced episodes that were Black or African American was slightly lower (8.7% versus 9.0%), and the share of beneficiaries with BPCI Advanced episodes who were dual eligible was slightly higher (21.2% versus 19.0%).
- There were large differences in representation for both Black or African American beneficiaries and dual-eligible beneficiaries when the sample was grouped by medical and surgical clinical episodes. For Black or African American beneficiaries, the representation among all FFS beneficiaries with any discharge or procedure was 10.1%. While representation among beneficiaries who had a medical hospital discharge with a BPCI Advanced triggering MS-DRG was slightly higher, at 11.2%, the representation among beneficiaries who had a surgical discharge or procedure with a BPCI Advanced triggering MS-DRG or HCPCS code was much lower, at 5.6%. For dual-eligible beneficiaries, the representation among all FFS beneficiaries with any discharge or procedure was 20.2%. While the representation among beneficiaries who had a medical hospital discharge with a BPCI Advanced triggering MS-DRG was higher, at 24.3%, the representation among beneficiaries who had a surgical discharge or procedure with a BPCI Advanced triggering MS-DRG or HCPCS code was much lower, at 10.9%.

b. Black or African American Beneficiaries

The analysis of the representativeness of Black or African American beneficiaries is included in Exhibit 32. Approximately 9.3% of the FFS Medicare population was Black or African American during Model Years 1, 2, and 3 (October 1, 2018 through December 31, 2020). In the sample of beneficiaries with a Medicare FFS discharge or outpatient procedure during this time, we find that

10.1% were Black or African American.[35] After further restricting the sample to beneficiary discharges with BPCI Advanced MS-DRGs or HCPCS codes, the sample includes a smaller fraction of Black or African American beneficiaries (9.0%). The share of discharges and outpatient procedures for Black or African American beneficiaries was much larger for medical clinical episodes than for surgical clinical episodes (11.2% versus 5.6%).

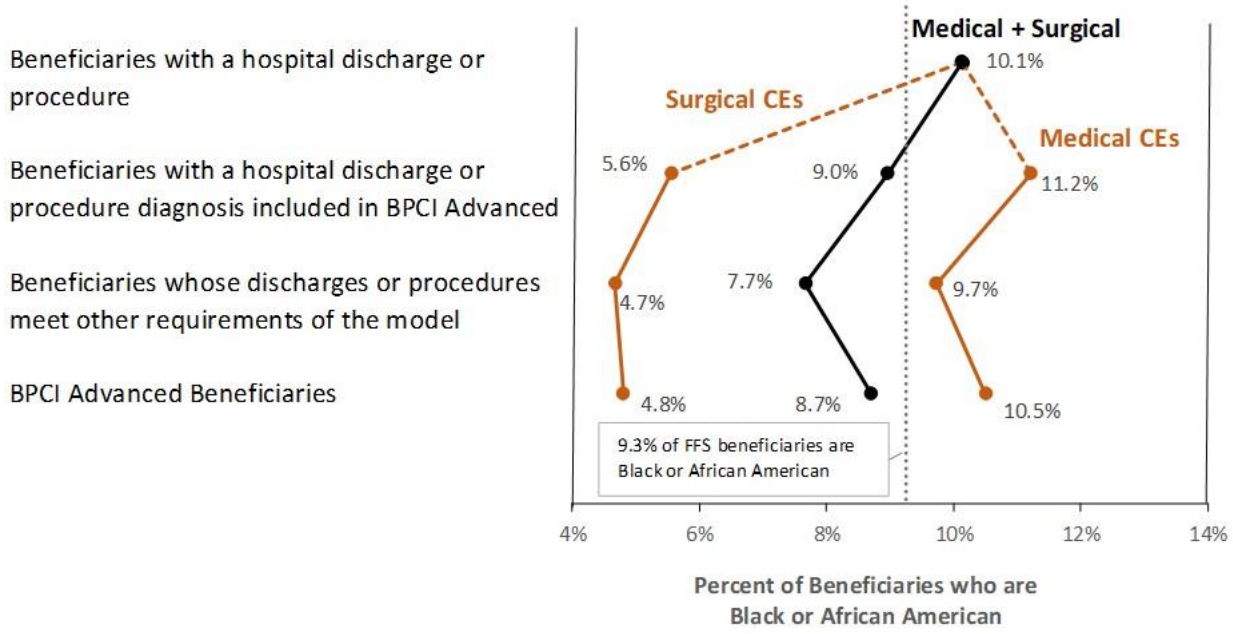
The sample included a smaller share of discharges and procedures for Black or African American beneficiaries after reducing the sample based on model criteria.[36] The decrease is driven by some of the beneficiary-level exclusions, such as excluding beneficiaries with end-stage renal disease (ESRD) or excluding beneficiaries without continuous enrollment in Medicare FFS Parts A & B. The decrease is also the result of provider-level exclusions, such as excluding all discharges and procedures which occur at hospitals in Maryland. While these criteria limit the reach of the BPCI Advanced Model to those groups of beneficiaries, many of the exclusions are in place because the beneficiaries are attributed to other CMMI Models or because the hospitals are ineligible during this period. For example, beneficiaries with ESRD may have been attributed to the Comprehensive ESRD Care Model, and beneficiaries treated at Maryland hospitals may have been attributed to the Maryland Total Cost of Care Model.

In the final sample, in which the set of beneficiaries was limited to those with an episode initiated by a BPCI Advanced participant and included in financial reconciliation, we find that 8.7% of these discharges were for Black or African American beneficiaries. The divergence in representation by episode type from the broader FFS discharge sample remained in this group of discharges, with Black or African American beneficiaries accounting for 10.5% of medical episodes compared to only 4.8% of surgical episodes.

[35] This sample includes all FFS discharges with any Medicare-covered MS-DRG, but it only includes outpatient procedures with a BPCI Advanced HCPCS code.

[36] These restrictions exclude any discharge or outpatient procedure: at a non-acute care hospital or at a cancer or critical access hospital; at a Maryland acute care hospital; with conflicting dates (e.g., discharge date prior to admission date); where the beneficiary is receiving end-stage renal disease (ESRD) care (e.g., dialysis in the 90 days prior to the episode, a kidney transplant in the 3 years prior to the episode); where Medicare is not the primary payer in the 90 days prior through the end of the episode; where the beneficiary is not continuously enrolled in Medicare FFS Part A & Part B in the 90 days prior through the end of the episode; where the beneficiary is aligned to an Accountable Care Organization (ACO), the Rural Health Demonstration, or the Pennsylvania Rural Health Model; where the anchor stay lasts more than 60 days; or if the outpatient procedure has a J1 status indicator and is not the highest ranked revenue line on the claim.

Exhibit 32: Impact of Exclusion Criteria on Black or African American Beneficiary Representation in the BPCI Advanced Model, Model Years 1-3, October 1, 2018 – December 31, 2020



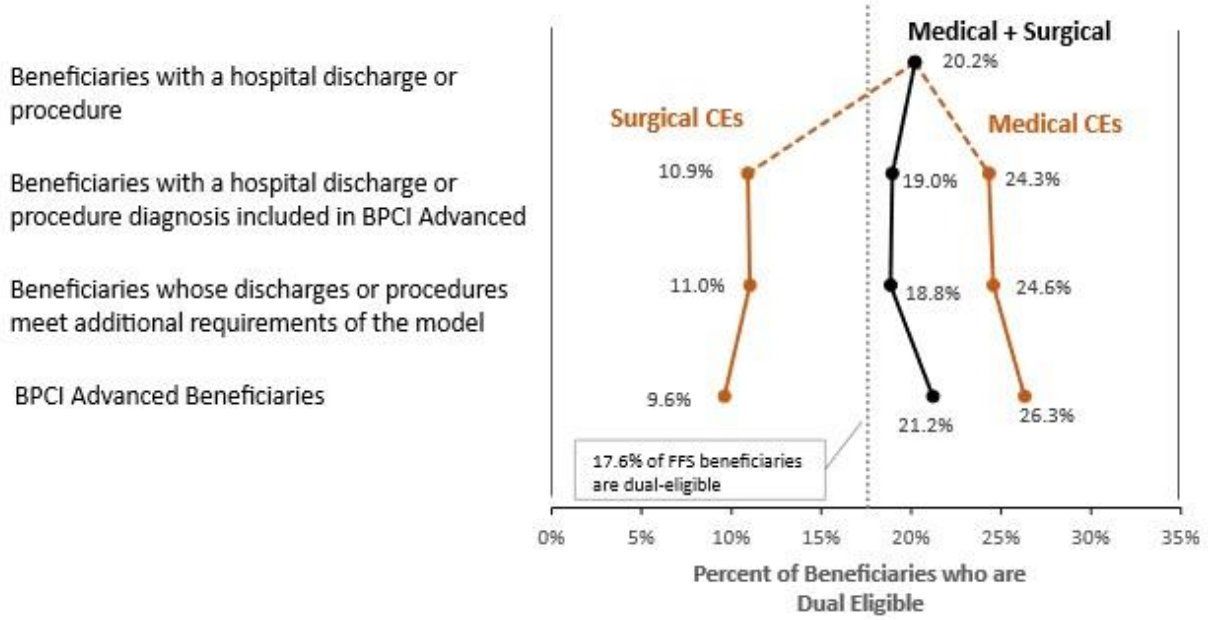
Note: The medical and surgical sample includes all inpatient acute hospitalizations (acute stay and acute transfer stay, when applicable) and outpatient procedures for BPCI Advanced HCPCS (Healthcare Common Procedure Coding System) with positive standardized allowed payments. The sample includes only the first discharge for a beneficiary within the Model Year 1 through 3 intervention period (September 1, 2018 through December 31, 2020). The BPCI Advanced Episodes sample is based on reconciliation data from the second true-up for all three performance periods. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Medicare Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. See **Appendix C** for details of the outcome definitions, data sources, and additional information on methods. CE = clinical episode; DRG = diagnosis-related group; FFS = fee-for-service.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for discharges with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2020. The BPCI Advanced Episodes sample is based on the implementation contractor's reconciliation data from the second true-up for Performance Periods 1 through 3.

c. *Dual-eligible Beneficiaries*

A similar analysis was performed for beneficiaries dually eligible for Medicare and Medicaid (Exhibit 33). Approximately 17.6% of all FFS enrollees were dual eligible, but dual-eligible beneficiaries accounted for approximately 20.2% of beneficiaries with a Medicare FFS discharge or outpatient procedure with a BPCI Advanced HCPCS code during Model Years 1, 2, or 3. The share of beneficiaries that were dual eligible was lower after we restricted the sample to beneficiaries that had discharges with BPCI-Advanced MS-DRGs or HCPCS codes. We saw differences in the representation of dual-eligible beneficiaries by episode type, with a larger share represented in medical clinical episodes (24.3%) than surgical clinical episodes (10.9%). The representation of dual-eligible beneficiaries remained stable even after excluding discharges based on model criteria. However, when we limited the sample to beneficiaries with episodes initiated by BPCI Advanced participants and included in reconciliation, there was a slight increase in the share within medical and a slight decrease in the share within surgical. Dual-eligible beneficiaries represented 26.3% of the share for medical clinical episodes and 9.6% of the share for surgical clinical episodes in the reconciliation sample.

Exhibit 33: Impact of Exclusion Criteria on Dual-Eligible Beneficiary Representation in the BPCI Advanced Model, Model Years 1-3, October 1, 2018 – December 31, 2020



Note: The medical and surgical sample includes inpatient acute hospitalizations (acute stay and acute transfer stay, when applicable) and outpatient procedures for BPCI Advanced CE triggering HCPCS (Healthcare Common Procedure Coding System) with positive standardized allowed payments. The sample includes only the first discharge for a beneficiary within the Model Years 1 through 3 intervention period (September 1, 2018 through December 31, 2020). The BPCI Advanced Episodes sample is based on reconciliation data from the second true-up for all three performance periods. See **Appendix C** for details of the outcome definitions, data sources, and additional information on methods. CE = clinical episode; DRG = diagnosis-related group; FFS = fee-for-service.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for discharges with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2020. The BPCI Advanced Episodes sample is based on the implementation contractor's reconciliation data from the second true-up for Performance Periods 1 through 3.

2. *Were there differences in outcomes under the BPCI Advanced Model for beneficiaries from populations that have been historically underserved?*

We estimated the change from the baseline to the intervention period for beneficiaries with BPCI Advanced episodes relative to comparison group episodes (referred to as the relative change) for each historically underserved population and a reference population. Additionally, we calculated the difference between these relative changes for each historically underserved population and the reference population (referred to as the differential change) to understand whether the model's impacts on underserved population differed from its impact on the reference populations. We conducted the analysis for two populations: Black or African American beneficiaries (relative to Non-Hispanic White beneficiaries) and beneficiaries dually eligible for Medicare and Medicaid (relative to beneficiaries not dually eligible), and for two time periods: Model Years 1 and 2 and Model Year 3.

This methodology relies on the assumption that our treatment and comparison groups are balanced on key characteristics for each underserved population and reference population. However, we constructed the comparison groups such that the treatment and comparison groups would be balanced across the full sample of beneficiaries and not necessarily for each underserved population and reference population (see **Appendix C** for more information on our comparison group methodology).

The analyses included the following outcomes: total standardized allowed episode payments (referred to as total episode payments), the unplanned readmission rate during the 90-day post-discharge period (PDP), the mortality rate during the anchor stay and 90-day PDP (for Model Year 3 analyses), and the mortality rate during the 90-day PDP (for Model Years 1 and 2). We conducted the analysis for episodes pooled across all medical clinical episodes evaluated and for episodes pooled across all surgical clinical episodes evaluated. Results were risk adjusted to account for differences in patient mix, clinical severity, and provider characteristics.

a. Key Findings

Changes in Outcomes Under BPCI Advanced for Underserved Populations

- During Model Years 1 and 2 and Model Year 3, total episode payments declined for all populations evaluated (Black or African American, Non-Hispanic White, dual eligible and nondual eligible) relative to their comparison groups, and the declines were larger for populations that have been historically underserved than the reference populations.
- There were no relative changes in readmission rates for the historically underserved populations analyzed in Model Year 3, but in Model Years 1 and 2, readmission rates declined for dual-eligible beneficiaries.
- For medical clinical episodes in Model Year 3, we found a statistically significant differential increase in the mortality rate for Black or African American beneficiaries compared to Non-Hispanic White beneficiaries and for dual-eligible beneficiaries compared to nondual-eligible beneficiaries. These differential increases were driven by relative declines in the mortality rate for the respective reference groups, suggesting that historically underserved populations were less likely to benefit under the model than the reference groups in Model Year 3. We did not find any statistically significant differential results for the mortality rate during Model Years 1 and 2.

b. Model Year 3 Findings

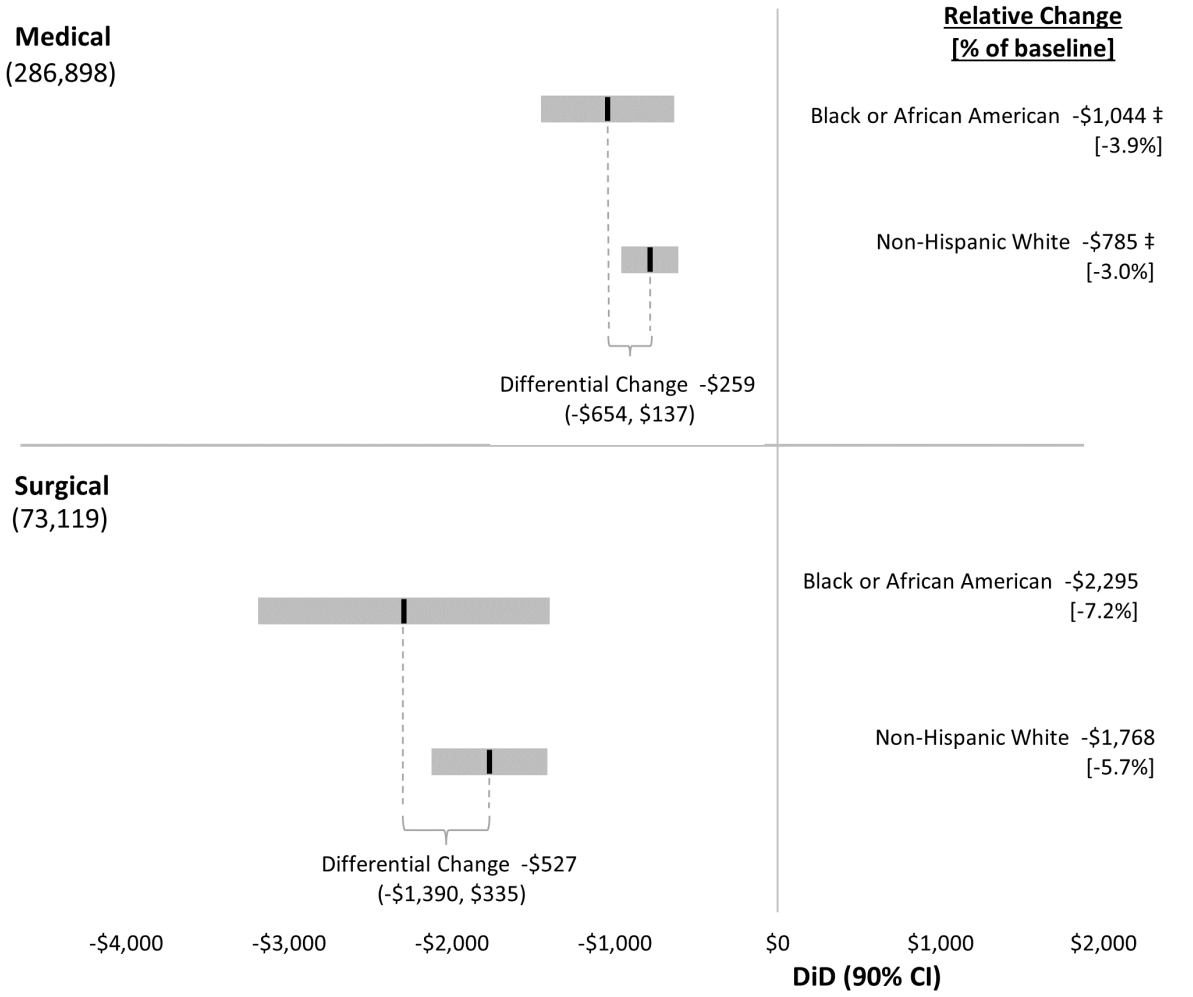
For this set of analyses, we estimated the change from the baseline to the Model Year 3 intervention period for beneficiaries with BPCI Advanced episodes relative to comparison group episodes used in the analyses presented in **Section II.A** for each historically underserved population. These analyses include the same set of clinical episodes, participants, and comparison group providers, limited to the subset of the beneficiaries from the population of interest and their reference population.

Black or African American Beneficiaries, Model Year 3

In both medical and surgical clinical episodes, total episode payments decreased for both Black or African American beneficiaries and for Non-Hispanic White beneficiaries relative to their comparison groups in Model Year 3 (Exhibit 34).[37] Reductions in total episode payments were over twice as large for surgical clinical episodes as they were for medical clinical episodes for both Black or African American beneficiaries and for Non-Hispanic White beneficiaries relative to their comparison groups. For Black or African American beneficiaries in medical clinical episodes, total episode payments declined $1,044 (90% confidence interval: -$1,452, -$635; p<0.01), or 3.9% of the baseline mean, relative to their comparison group, while payments for surgical clinical episodes declined $2,295 (90% confidence interval: -$3,189, -$1,401; p<0.01), or 7.2%. For Non-Hispanic White beneficiaries in medical clinical episodes, total episode payments declined $785 (90% confidence interval: -$961, -$609; p<0.01), or 3.0%, relative to their comparison group, while payments for surgical clinical episodes declined $1,768 (90% confidence interval: -$2,123, -$1,413; p<0.01), or 5.7%. The differential reduction for surgical clinical episodes was larger than the differential reduction for medical clinical episodes (-$527 vs. -$259), though neither result was statistically significant.

[37] The results for medical clinical episodes did not pass the parallel trends test. A key assumption required for an unbiased DiD estimate is that outcomes for BPCI Advanced and the comparison group had the same trend during the baseline period. Results of the parallel trends tests are reported in **Appendix G**. Additional details on the parallel trends test methodology are described in **Appendix C**.

Exhibit 34: Change in Total Episode Payments Under BPCI Advanced for Black or African American Beneficiaries and Non-Hispanic White Beneficiaries, Model Year 3, January – December 2020



Note: Total payments represent fee-for-service Parts A and B payments for the anchor stay or procedure and the 90-day post-discharge period. The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in dollars. Results are also presented as a percentage of the BPCI Advanced baseline mean. The grey bars indicate the 90% confidence interval of the DiD estimates. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Medicare Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, data sources, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences.

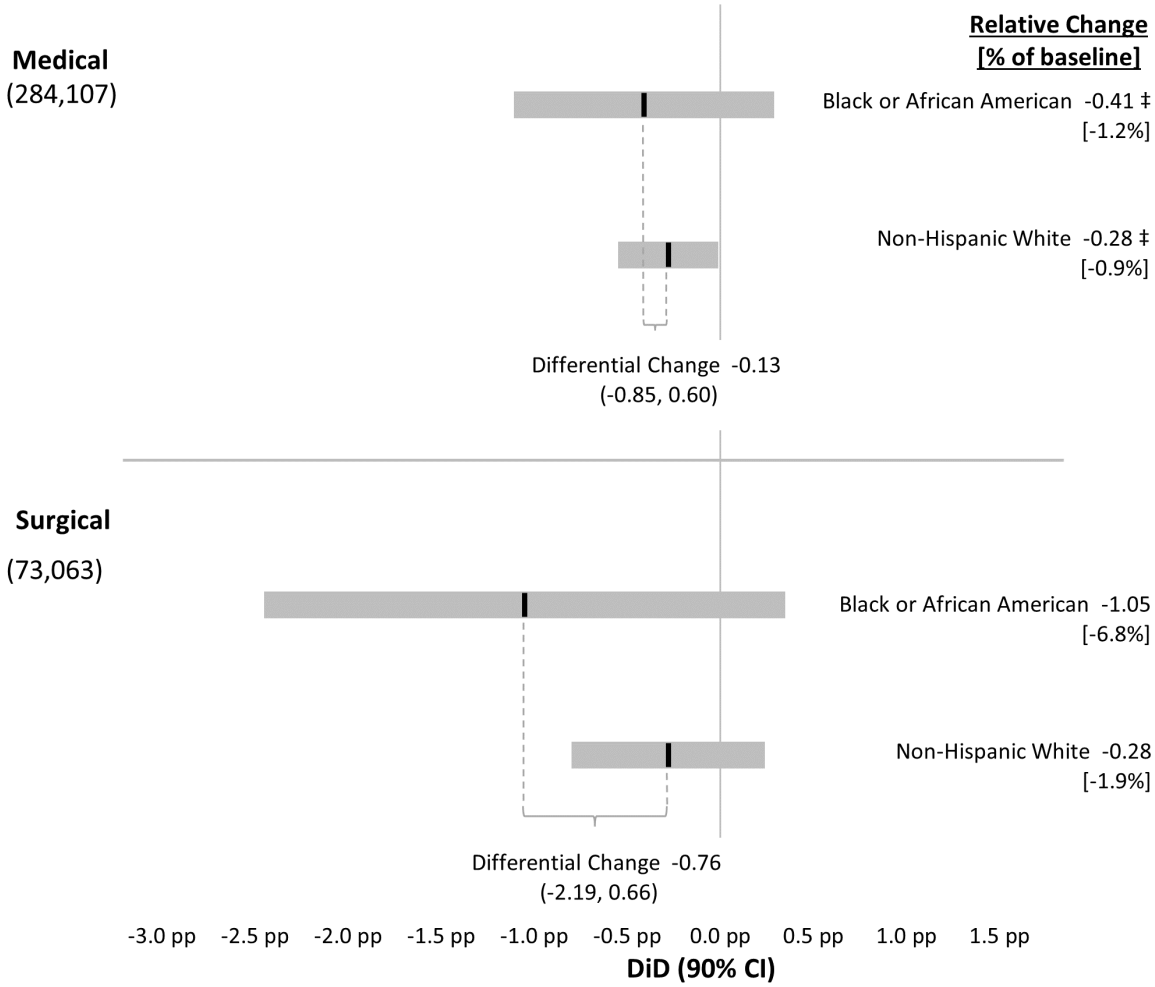
‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

There were relative declines in the unplanned readmission rate during the 90-day PDP for both Black or African American beneficiaries and for Non-Hispanic White beneficiaries in the BPCI

Advanced Model relative to their comparison groups for both medical and surgical clinical episodes during Model Year 3, though only the estimated decline for Non-Hispanic White beneficiaries in medical clinical episodes was statistically significant, and the results for medical clinical episodes failed the parallel trends test (Exhibit 35).

Exhibit 35: Change in the Unplanned Readmission Rate During the 90-day PDP for Black or African American Beneficiaries and Non-Hispanic White Beneficiaries, Model Year 3, January – December 2020



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean readmission rates. The grey bars indicate the 90% confidence interval of the DiD estimates. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Medicare Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, data sources, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or

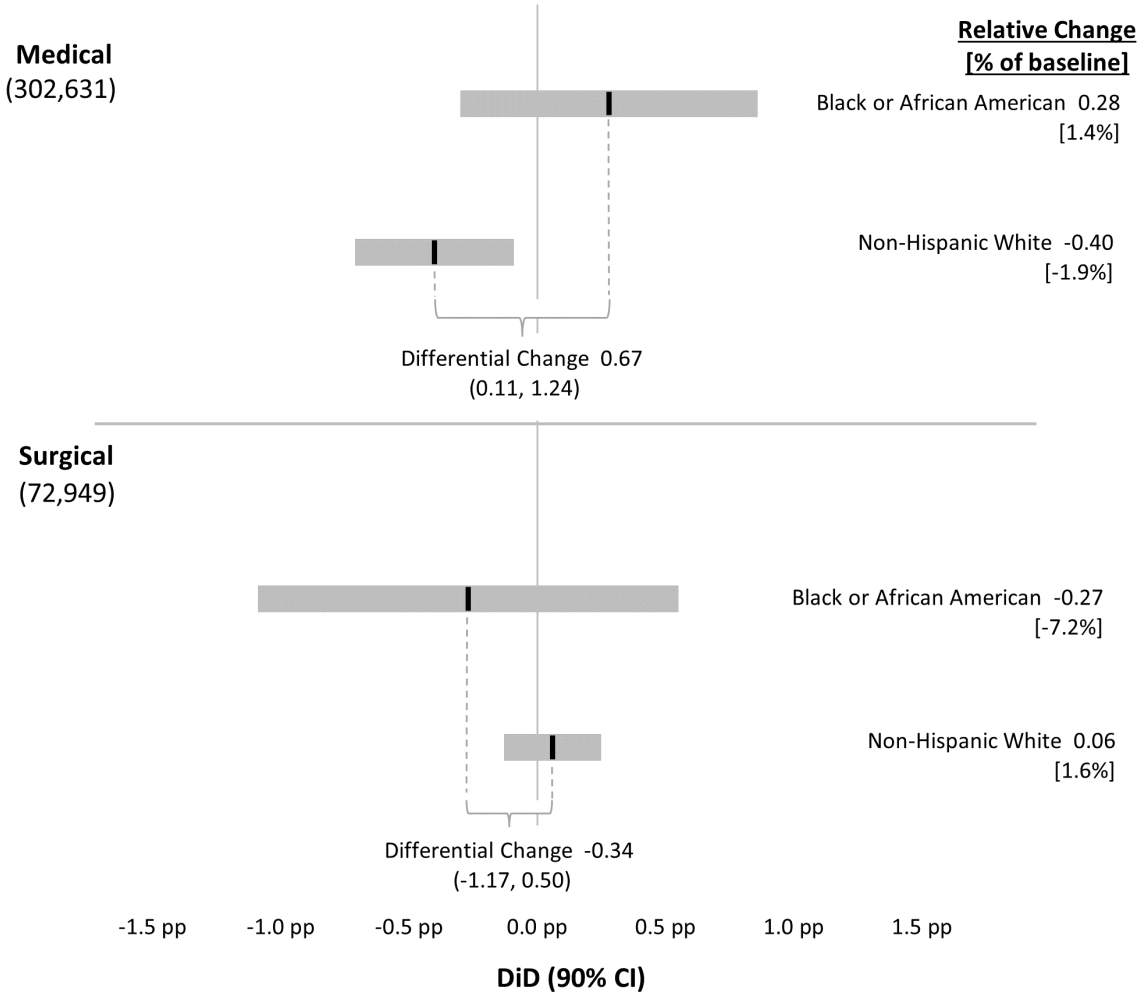
procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

We found a statistically significant differential increase of 0.67 pp (90% confidence interval: 0.11, 1.24; p=0.05) in the mortality rate for Black or African American beneficiaries compared to Non-Hispanic White beneficiaries for medical clinical episodes in Model Year 3 (Exhibit 36). For Black or African American beneficiaries in medical clinical episodes, although not statistically significant, the mortality rate increased by 0.28 pp (90% confidence interval: -0.31, 0.86; p=0.44), or 1.4% of the baseline readmission rate, relative to their comparison group. For Non-Hispanic White beneficiaries in medical clinical episodes, the mortality rate decreased by 0.40 pp (90% confidence interval: -0.71, -0.09; p=0.03), or 1.9%, relative to their comparison group. These results suggest that Black or African American beneficiaries were less likely to benefit from the BPCI Advanced Model than Non-Hispanic White beneficiaries.

For surgical clinical episodes, we found a non-statistically significant differential decrease of 0.34 pp (90% confidence interval: -1.17, 0.50; p=0.51) in the mortality rate for Black or African American beneficiaries compared to Non-Hispanic White beneficiaries in Model Year 3. For Black or African American beneficiaries in surgical clinical episodes, although not statistically significant, the mortality rate decreased by 0.27 pp (90% confidence interval: -1.09, 0.55; p=0.59) or 7.2%, relative to their comparison group. For Non-Hispanic White beneficiaries in surgical clinical episodes, although not statistically significant, the mortality rate increased by 0.06 pp (90% confidence interval: -0.12, 0.25; p=0.57), or 1.6%, relative to their comparison group.

The differential changes in mortality rates occurred in Model Year 3, during the COVID-19 PHE, and the differential rates of COVID-19 diagnosis experienced by these populations provide the broader context for these findings. For Black or African American beneficiaries in medical clinical episodes, the share of BPCI Advanced episodes with a COVID-19 diagnosis was 1.6 pp higher than the share for Black or African American beneficiaries in the comparison group. This difference was only 0.2 pp for Non-Hispanic White beneficiaries (see Exhibit K.13 in **Appendix K**). While we control for COVID-19 diagnosis, the risk adjustment may not perfectly control for the impact of COVID-19 on mortality rates.

Exhibit 36: Change in the Mortality Rate During the Anchor Stay or Procedure and the 90-Day PDP for Black or African American Beneficiaries and Non-Hispanic White Beneficiaries, Model Year 3, January – December 2020



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean mortality rates. The grey bars indicate the 90% confidence interval of the DiD estimates. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Medicare Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

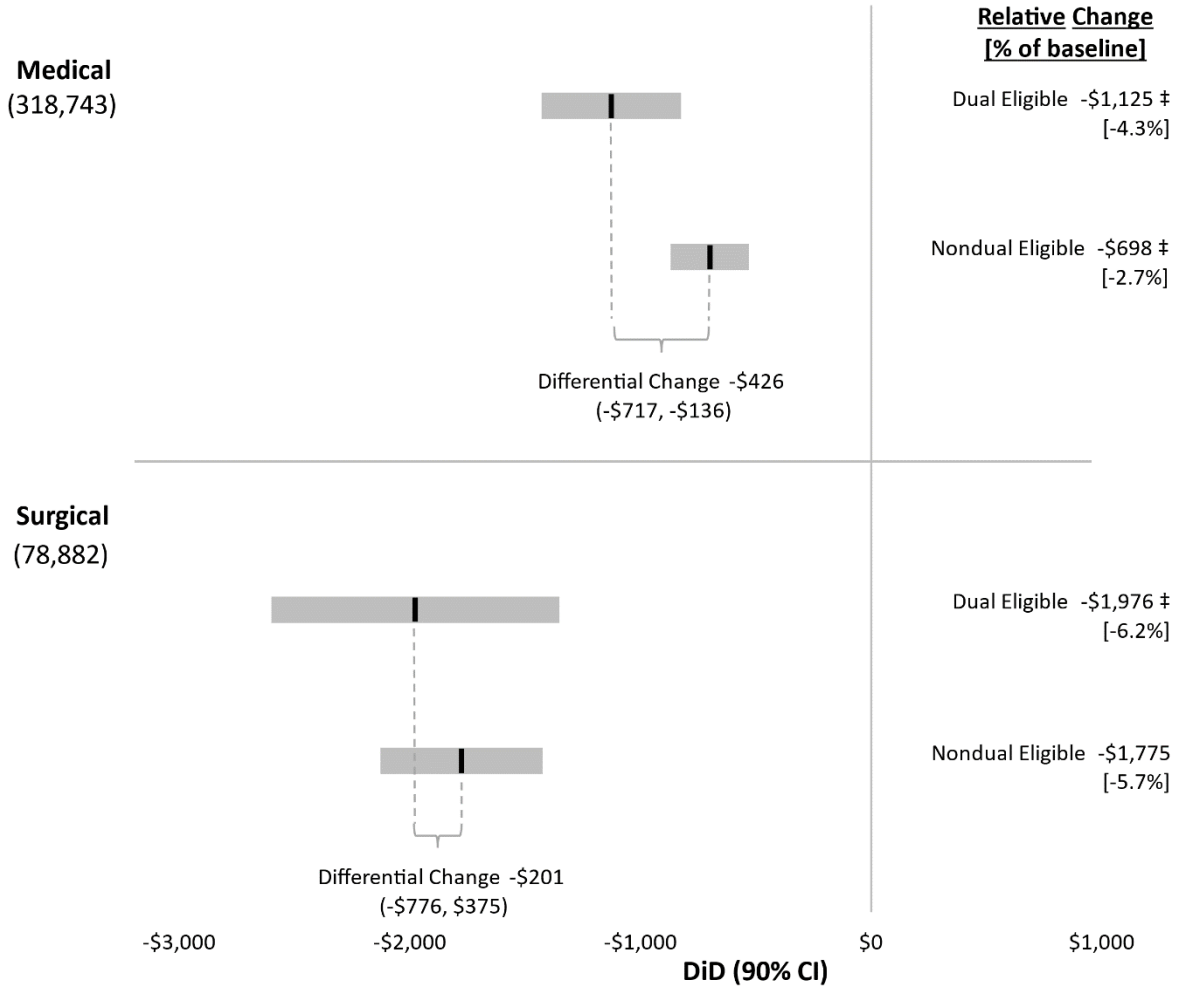
Dual-Eligible Beneficiaries, Model Year 3

For medical clinical episodes, total episode payments decreased for BPCI Advanced dual-eligible beneficiaries and for BPCI Advanced nondual-eligible beneficiaries relative to their comparison

groups in Model Year 3, but the relative decrease was larger for dual-eligible beneficiaries compared to nondual-eligible beneficiaries (Exhibit 37). The reduction in average episode payments was $1,125 (90% confidence interval: -$1,427, -$823; p<0.01), or 4.3% of the baseline mean, for dual-eligible beneficiaries and $698 (90% confidence interval: -$868, -$529; p<0.01), or 2.7%, for nondual-eligible beneficiaries. As a result, total episode payments decreased by $426 (90% confidence interval: -$717, -$136; p=0.02) more for dual-eligible beneficiaries compared to nondual-eligible beneficiaries in medical clinical episodes.

For surgical clinical episodes, total episode payments declined by $1,976 (90% confidence interval: -$2,600, -$1,352; p<0.01), or 6.2%, for dual-eligible beneficiaries relative to their comparison group in Model Year 3 and by $1,775 (90% confidence interval: -$2,128, -$1,423; p<0.01), or 5.7%, for nondual-eligible beneficiaries. As a result, total episode payments decreased by $201 (90% confidence interval: -$776, $375; p=0.57) more for dual-eligible beneficiaries compared to nondual-eligible beneficiaries in surgical clinical episodes, though the difference was not statistically significant.

Exhibit 37: Change in Total Episode Payments for Dual-Eligible Beneficiaries and Nondual-Eligible Beneficiaries, Model Year 3, January – December 2020



Note: Total payments represent fee-for-service Parts A and B payments for the anchor stay or procedure and the 90-day post-discharge period. The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in dollars. Results are also presented as a percentage of the BPCI Advanced baseline mean. The grey bars indicate the 90% confidence interval of the DiD estimates. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

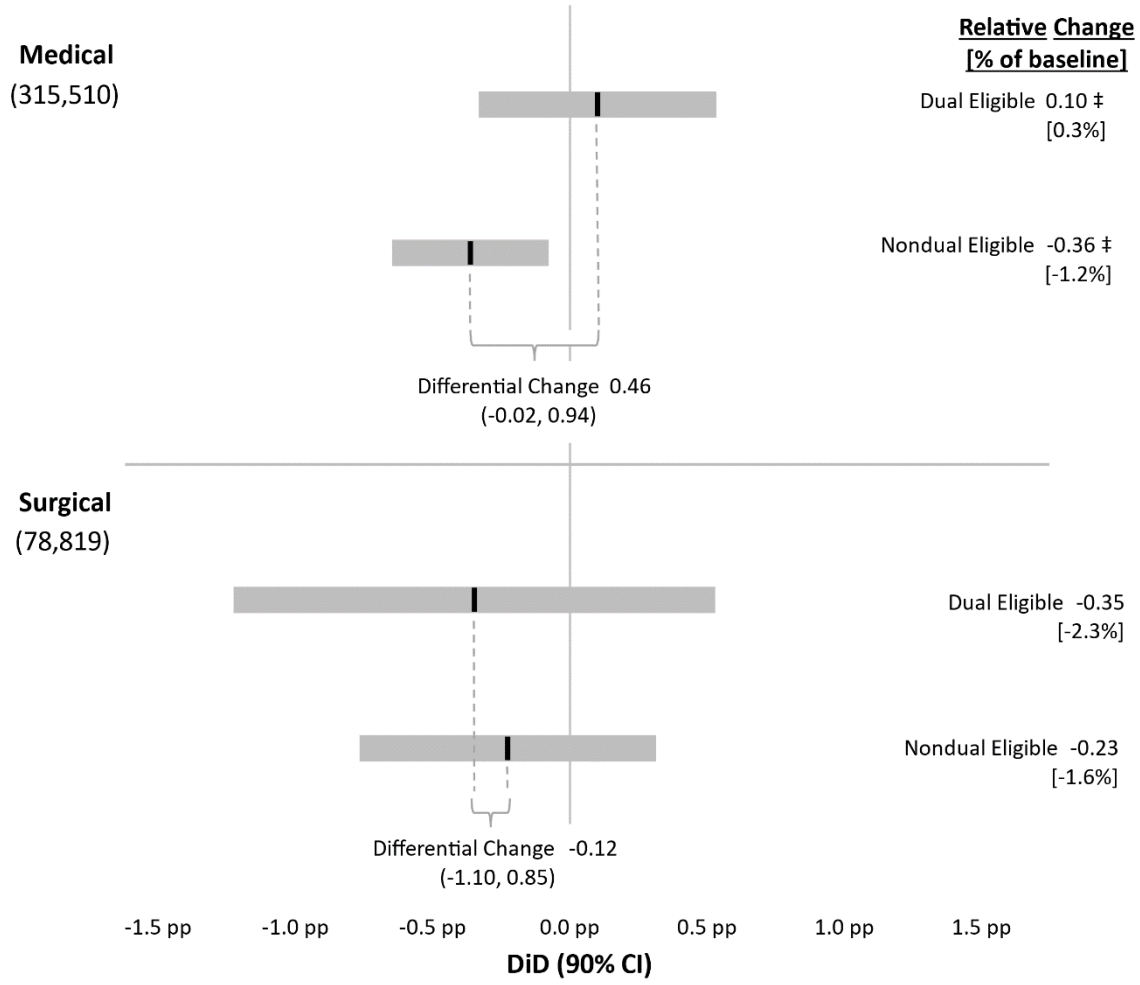
Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

For medical clinical episodes, although not statistically significant, the readmission rate increased by 0.10 pp (90% confidence interval: -0.33, 0.53; p=0.70), or 0.3% percent of the baseline mean readmission rate, for dual-eligible beneficiaries relative to their comparison group during Model Year 3 (Exhibit 38). The readmission rate declined by 0.36 pp (90% confidence interval: -0.65, -0.08; p=0.04), or 1.2%, for nondual-eligible beneficiaries with medical clinical episodes relative to their comparison group. The resulting differential change for dual-eligible beneficiaries

compared to nondual-eligible beneficiaries in medical clinical episodes was not statistically significant (estimate=0.46 pp; 90% confidence interval: -0.02, 0.94; p=0.11).

Readmission rates decreased for both dual-eligible and nondual-eligible beneficiaries in surgical clinical episodes relative to their comparison groups in Model Year 3, with a differential decrease of 0.12 pp (90% confidence interval: -1.10, 0.85; p=0.84) for dual-eligible beneficiaries compared to nondual-eligible beneficiaries in surgical clinical episodes. None of these estimates were statistically significant.

Exhibit 38: Change in the Unplanned Readmission Rate During the 90-Day PDP for Dual-Eligible Beneficiaries and Nondual-Eligible Beneficiaries, Model Year 3, January – December 2020



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean readmission rates. The grey bars indicate the 90% confidence interval of the DiD estimates. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

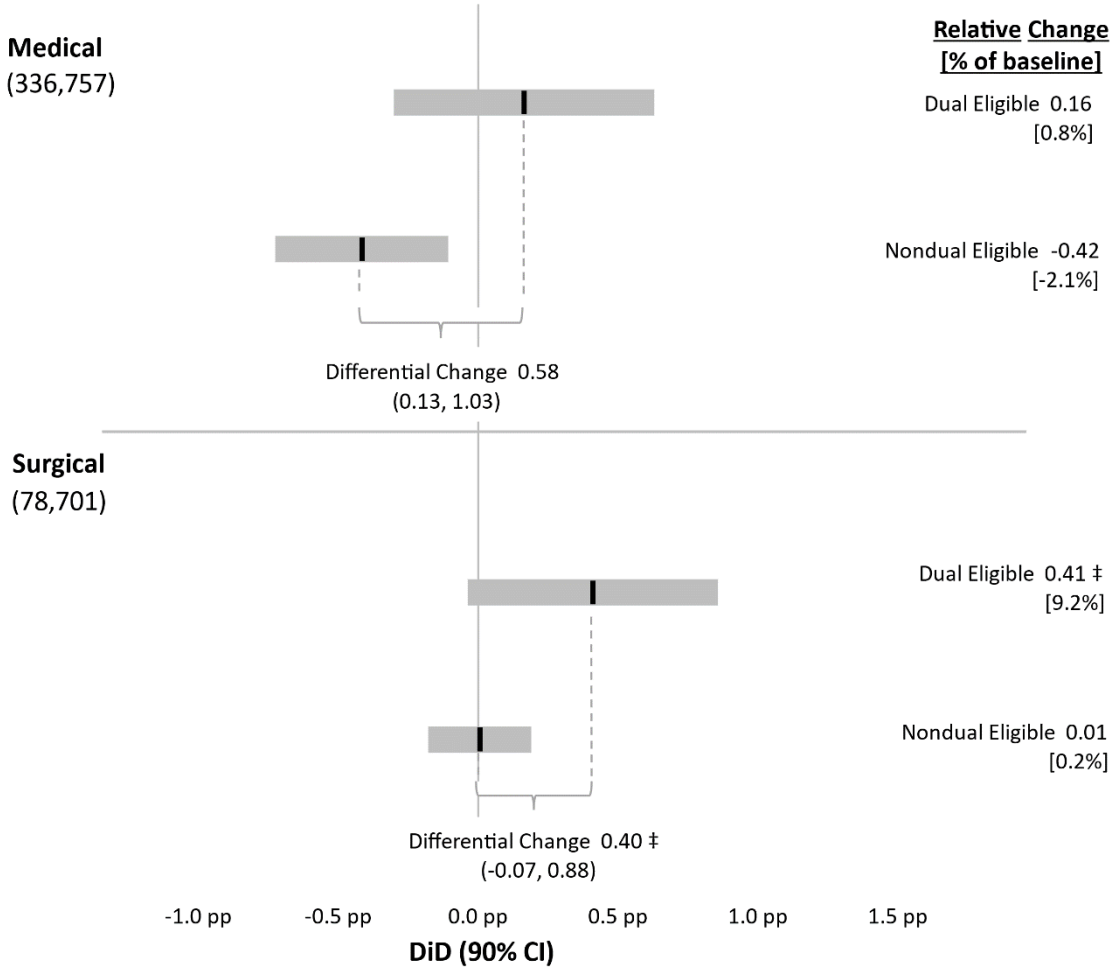
Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

For medical clinical episodes, we found a statistically significant differential increase of 0.58 pp (90% confidence interval: 0.13, 1.03; p=0.03) in the mortality rate for dual-eligible beneficiaries compared to nondual-eligible beneficiaries during Model Year 3 (Exhibit 39).[38] Although not statistically significant, the mortality rate for dual-eligible beneficiaries increased by 0.16 pp (90% confidence interval: -0.30, 0.63; p=0.56), or 0.8% of the baseline mortality rate, relative to their comparison group. The mortality rate for nondual-eligible beneficiaries in medical clinical episodes decreased 0.42 pp (90% confidence interval: -0.72, -0.11; p=0.03), or 2.1%, relative to their comparison group. These results suggest that dual-eligible beneficiaries were less likely to benefit from the BPCI Advanced Model than nondual-eligible beneficiaries.

For surgical clinical episodes, although not statistically significant, the mortality rate for dual-eligible beneficiaries increased by 0.41 pp (90% confidence interval: -0.04, 0.86; p=0.13), or 9.2%, relative to their comparison group in Model Year 3. There was no change in the mortality rate for nondual-eligible beneficiaries in surgical episodes relative to their comparison group (estimate=0.01 pp; 90% confidence interval: -0.18, 0.19; p=0.95). Although not statistically significant, there was a differential increase in mortality for dual-eligible beneficiaries in surgical clinical episodes compared to nondual-eligible beneficiaries (estimate=0.40; 90% confidence interval: -0.07, 0.88; p=0.16).

[38] We examined the differential rates of COVID diagnosis experienced by these populations. However, COVID rates for dual-eligible beneficiaries were fairly balanced between BPCI Advanced and the comparison group. (See Exhibit K.14 in **Appendix K**).

Exhibit 39: Change in the Mortality Rate During the Anchor Stay or Procedure and 90-Day PDP for Dual-Eligible Beneficiaries and Nondual-Eligible Beneficiaries, Model Year 3, January – December 2020



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean mortality rates. The grey bars indicate the 90% confidence interval of the DiD estimates. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

c. Model Years 1 and 2 Findings

Given the differential increases in mortality rates during Model Year 3 for Black or African American beneficiaries and for dual-eligible beneficiaries, we conducted additional analyses to determine whether there were differential increases in mortality rates in Model Years 1 and 2.

Between Model Years 1 and 2 and Model Year 3, there were differences in participation, the matched comparison group, and the methods used for our analyses, which create challenges in comparing the magnitude of estimates over time and in determining the mechanism behind any differential changes.[39] However, conducting the analysis on Model Years 1 and 2 in addition to Model Year 3 provides a more thorough understanding of the BPCI Advanced Model in its entirety.

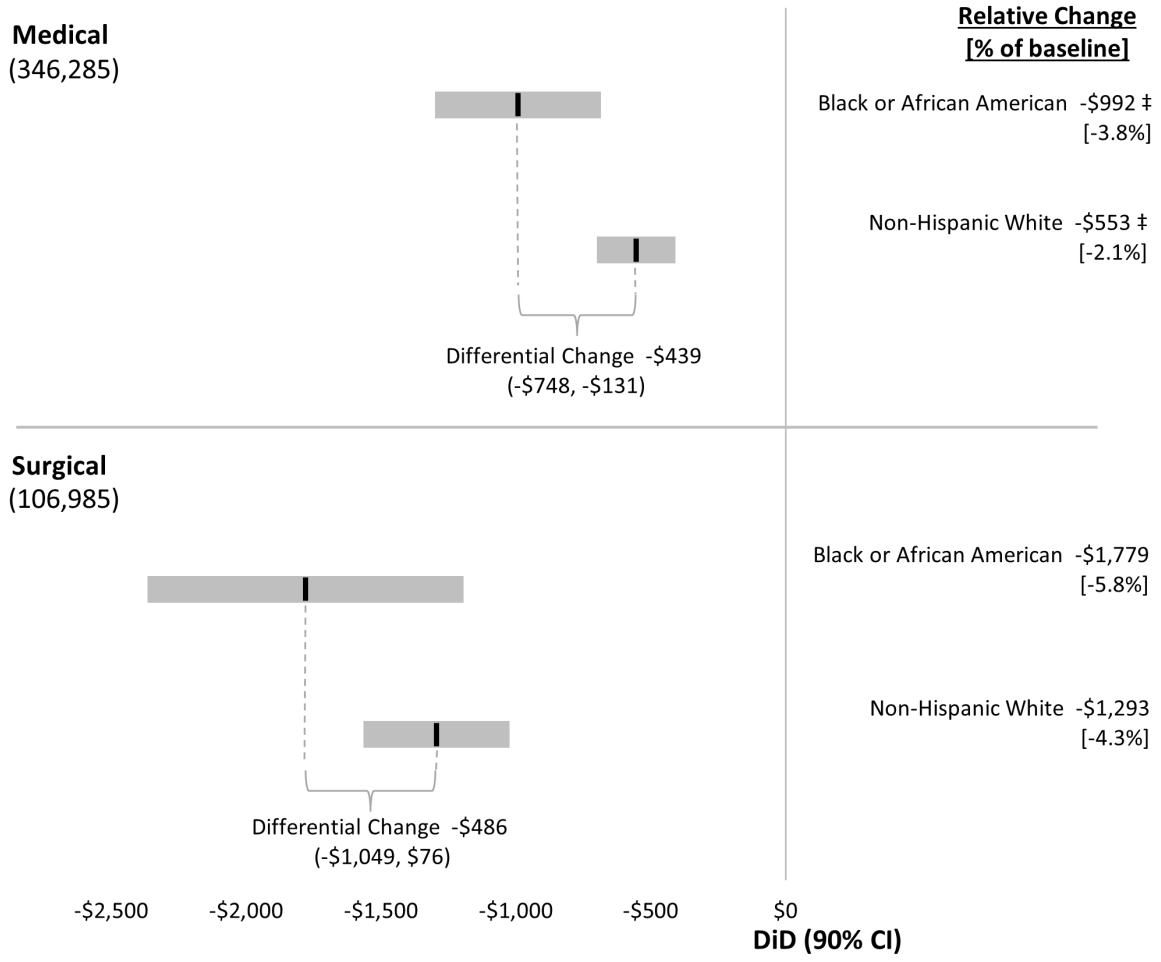
Black or African American Beneficiaries, Model Years 1 and 2

In medical and surgical clinical episodes, total episode payments decreased for both Black or African American beneficiaries and Non-Hispanic White beneficiaries relative to their comparison groups during Model Years 1 and 2, with larger reductions in surgical clinical episodes than in medical clinical episodes (Exhibit 40).[40] For Black or African American beneficiaries in medical clinical episodes, total episode payments declined $992 (90% confidence interval: -$1,300, -$685; $p<0.001$) or 3.8% of the baseline mean, relative to their comparison group, while payments for surgical clinical episodes declined $1,779 (90% confidence interval: -$2,364, -$1,194; $p<0.001$) or 5.8%. For Non-Hispanic White beneficiaries in medical clinical episodes, payments declined $553 (90% confidence interval: -$699, -$407; $p<0.001$) or 2.1%, while payments for surgical clinical episodes declined $1,293 (90% confidence interval: -$1,563, -$1,022; $p<0.001$) or 4.3%. The resulting differential reductions for medical and surgical clinical episodes were similar in magnitude (-$439 and -$486, respectively), though only the reduction for medical clinical episodes was statistically significant.

[39] The **Appendix C** of the BPCI Advanced Third Evaluation Report and the Fourth Evaluation Report include details of how the comparison groups were constructed, which clinical episodes were evaluated, and methods used for estimation. One notable difference is that the outcome definition for the mortality rate differed between the analyses performed for Model Years 1 and 2 and for Model Year 3. In our analyses of Model Years 1 and 2, we measured the mortality rate in the 90-day PDP, whereas in Model Year 3, we used a measure which includes mortality during the anchor stay and the 90-day PDP. Additionally, there were changes to participation (new BPCI Advanced participants joined in Model Year 3, existing participants made changes to clinical episode selection, and some participants withdrew from the model) which warranted construction of different comparison groups for the Model Year 3 analyses. For the analyses in this report, we evaluated additional clinical episodes (compared to the Third Evaluation Report), and we used a different set of variables in risk adjustment. Finally, Model Year 3 occurs in 2020, and thus, the impact of the model may be different due to the COVID-19 PHE. The BPCI Advanced Third Evaluation Report is available for download at https://innovation.cms.gov/innovation-models/bpci-advanced.

[40] The results for medical clinical episodes did not pass the parallel trends test. A key assumption required for an unbiased DiD estimate is that outcomes for BPCI Advanced and the comparison group had the same trend during the baseline period. Results of the parallel trends tests are reported in **Appendix G**. Additional details on the parallel trends test methodology are described in **Appendix C**.

Exhibit 40: Change in Total Episode Payments Under BPCI Advanced for Black or African American Beneficiaries and Non-Hispanic White Beneficiaries, Model Years 1-2, October 2018 – December 2019



Note: Total payments represent fee-for-service Parts A and B payments for the anchor stay or procedure and the 90-day post-discharge period. The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiD (relative changes) and DiDiD (differential change) estimates represent the relative change in dollars. Results are also presented as a percentage of the BPCI Advanced baseline mean. The grey bars indicate the 90% confidence interval of the DiD estimates. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Medicare Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, data sources, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences.

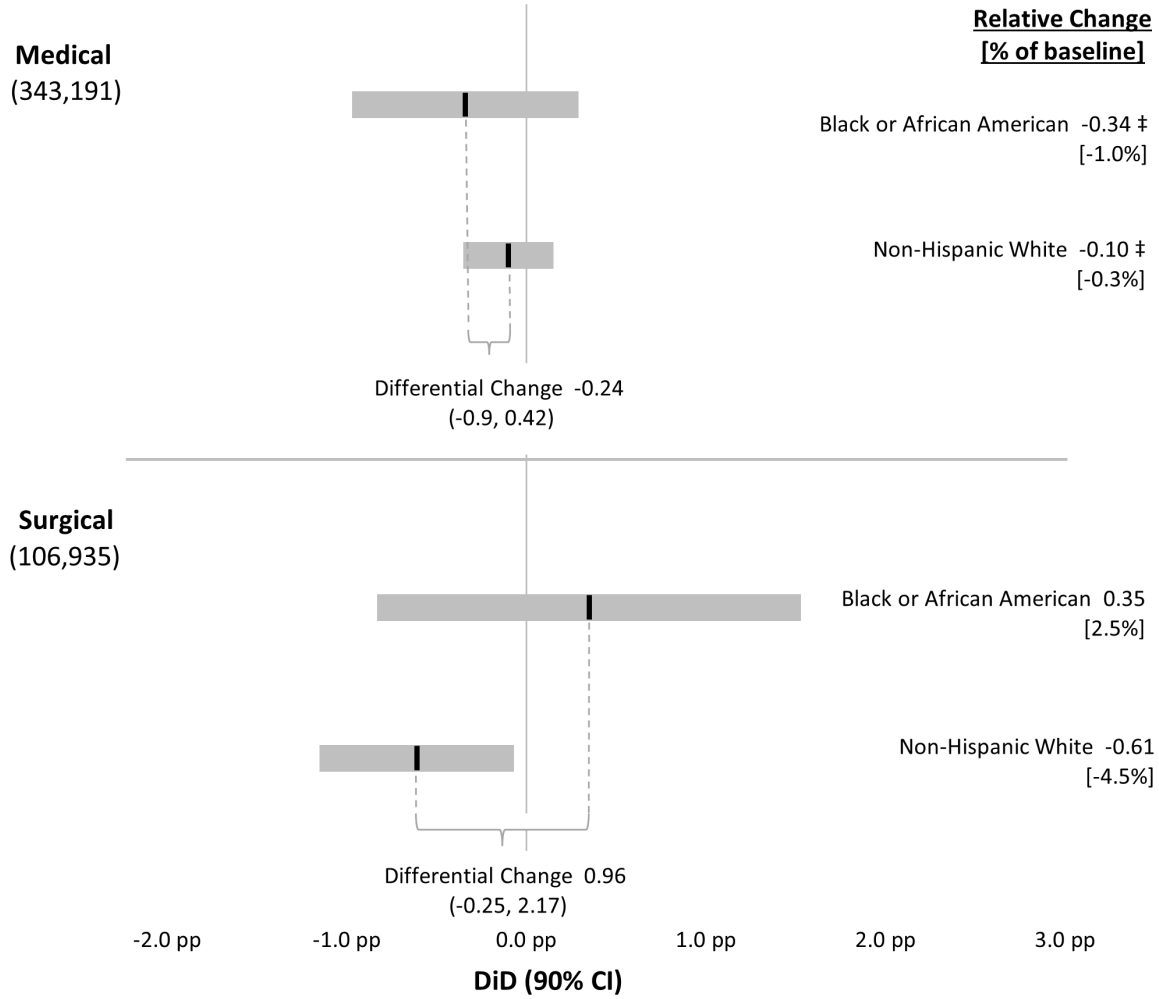
‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2019 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

There were declines in the unplanned readmission rate during the 90-day PDP for Black or African American beneficiaries in medical clinical episodes and for Non-Hispanic White beneficiaries in both medical and surgical clinical episodes relative to their comparison groups during Model Years

1 and 2, though only the estimate for Non-Hispanic White beneficiaries in surgical clinical episodes was statistically significant. The results for both populations in medical clinical episodes failed the parallel trends test (Exhibit 41). The differential changes for beneficiaries in the two groupings were not statistically significant. For medical clinical episodes, there was a small differential decrease in the readmission rate of 0.24 pp (90% confidence interval: -0.90, 0.42; p=0.56), while for surgical clinical episodes, there was a differential increase in the readmission rate of 0.96 pp (90% confidence interval: -0.25, 2.17; p=0.19).

Exhibit 41: Change in the Unplanned Readmission Rate During the 90-day PDP Under BPCI Advanced for Black or African American Beneficiaries and Non-Hispanic White Beneficiaries, Model Years 1-2, October 2018 – December 2019



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean readmission rates. The grey bars indicate the 90% confidence interval of the DiD estimates. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Medicare Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, data sources, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

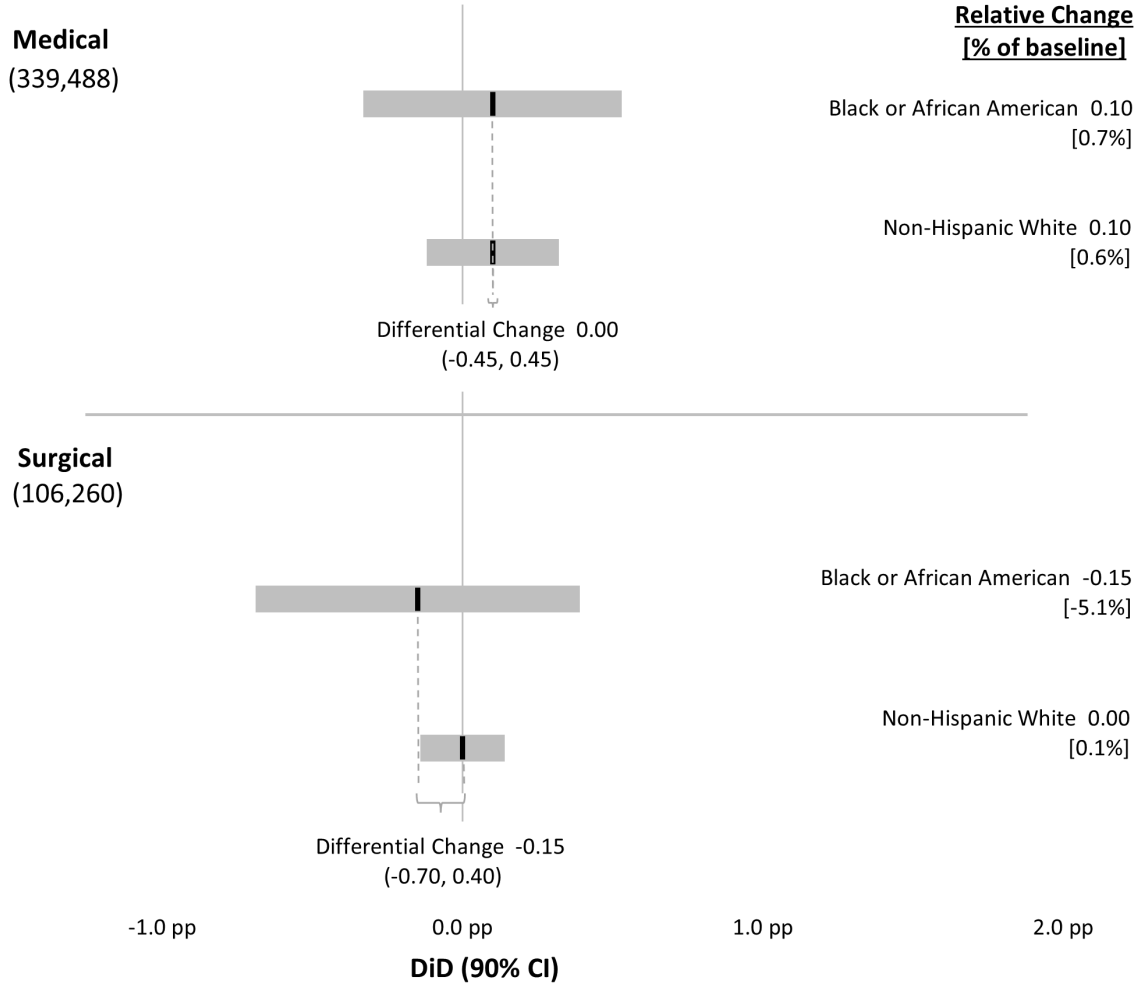
‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2019 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

In Model Years 1 and 2, we did not find any statistically significant relative changes or differential changes in the mortality rate during the 90-day PDP for Black or African American beneficiaries or Non-Hispanic White beneficiaries in either medical or surgical clinical episodes (Exhibit 42). For Black or African American beneficiaries in medical clinical episodes, the mortality rate increased by 0.10 pp (90% confidence interval: -0.32, 0.53; p=0.69), or 0.7%, relative to their comparison group. Similarly, for Non-Hispanic White beneficiaries in medical clinical episodes, the mortality rate increased by 0.10 pp (90% confidence interval: -0.11, 0.32; p=0.43), or 0.6%, relative to their comparison group. This resulted in a near-zero differential change in the mortality rate for medical clinical episodes. For Black or African American beneficiaries in surgical clinical episodes, the mortality rate declined by 0.15 pp (90% confidence interval: -0.68, 0.39; p=0.65), or 5.1%, relative to their comparison group, while for Non-Hispanic White beneficiaries, the mortality rate was nearly unchanged (estimate=0.00 pp; 90% confidence interval: -0.13, 0.14; p=0.95) relative to their comparison group. This resulted in a non-statistically significant differential decline of 0.15 pp (90% confidence interval: -0.70, 0.40; p=0.65) in the mortality rate for surgical clinical episodes.

While the analysis of Model Year 3 found a statistically significant differential increase in the mortality rate for Black or African American beneficiaries with medical clinical episodes compared to Non-Hispanic White beneficiaries, we did not observe differential impacts on the mortality rate in Model Years 1 and 2, mitigating concerns based on one single year of the model.

Exhibit 42: Change in the Mortality Rate During the 90-Day PDP for Black or African American Beneficiaries and Non-Hispanic White Beneficiaries, Model Years 1-2, October 2018 – December 2019



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean mortality rates. The grey bars indicate the 90% confidence interval of the DiD estimates. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Medicare Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, data sources, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

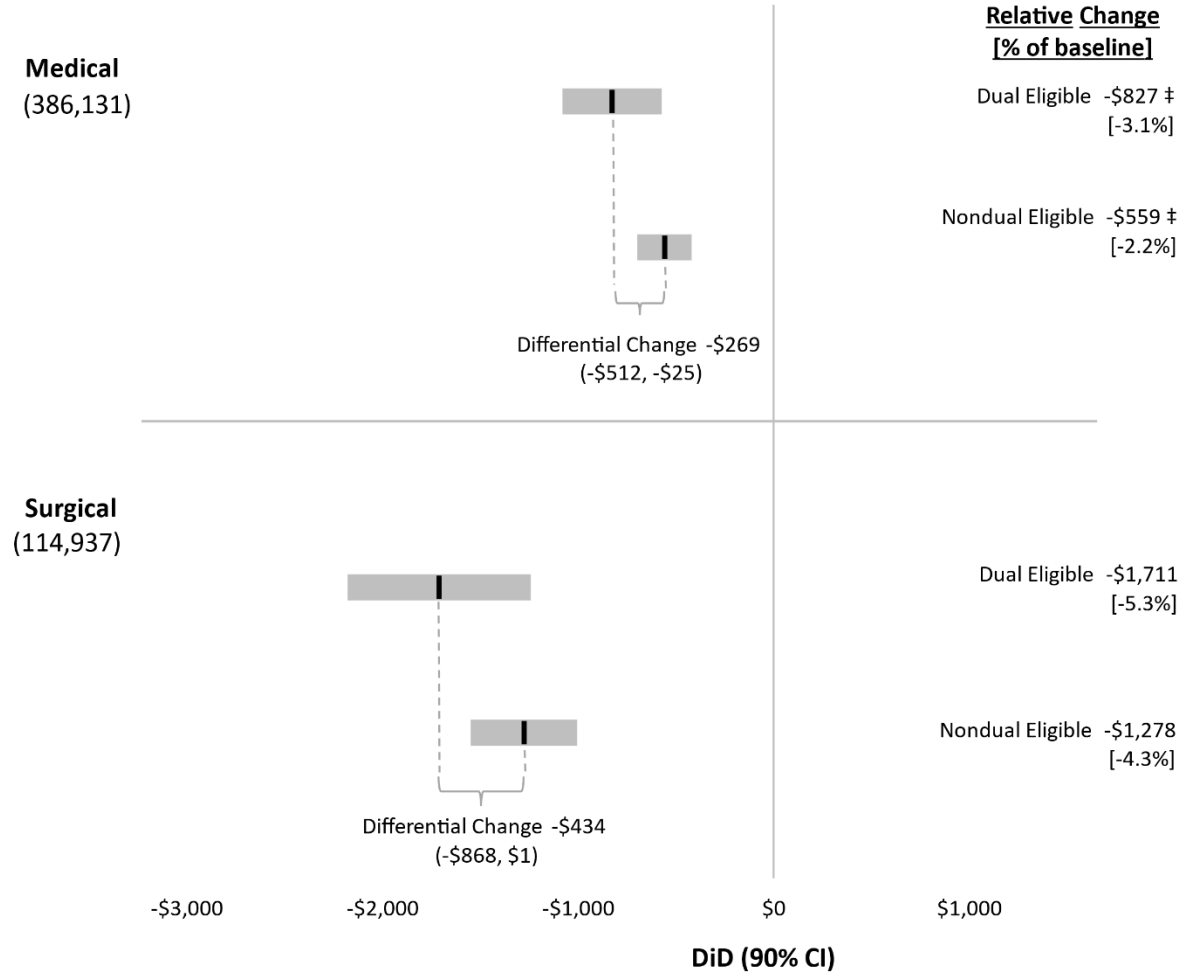
Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2019 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

Dual-Eligible Beneficiaries, Model Years 1 and 2

For medical clinical episodes, total episode payments decreased for BPCI Advanced dual-eligible beneficiaries and BPCI Advanced nondual-eligible beneficiaries relative to their comparison groups in Model Years 1 and 2 (Exhibit 43). The reduction in average episode payments was $827 (90% confidence interval: -$1,080, -$547; p<0.01), or 3.1% of the baseline mean, for dual-eligible beneficiaries and $559 (90% confidence interval: -$697, -$420; p<0.01), or 2.2%, for nondual-eligible beneficiaries. Total episode payments decreased by $269 (90% confidence interval: -$512, -$25; p=0.07) more for dual-eligible beneficiaries in medical clinical episodes compared to nondual-eligible beneficiaries.

For surgical clinical episodes, total episode payments declined by $1,711 (90% confidence interval: -$2,178, -$1,244; p<0.01), or 5.3%, for dual-eligible beneficiaries relative to their comparison group and by $1,278 (90% confidence interval: -$1,550, -1,005; p<0.01), or 4.3%, for nondual-eligible beneficiaries in Model Years 1 and 2. Total episode payments decreased by $434 (90% confidence interval: -$868, $1; p=0.10) more for dual-eligible beneficiaries in surgical clinical episodes compared to nondual-eligible beneficiaries, though the estimate was not statistically significant.

Exhibit 43: Change in Total Episode Payments Under BPCI Advanced for Dual-Eligible Beneficiaries and Nondual-Eligible Beneficiaries, Model Years 1-2, October 2018 – December 2019



Note: Total payments represent Parts A and B fee-for-service payments for the anchor stay or procedure and the 90-day post-discharge period. The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in dollars. Results are also presented as a percentage of the BPCI Advanced baseline mean. The grey bars indicate the 90% confidence interval of the DiD estimates. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

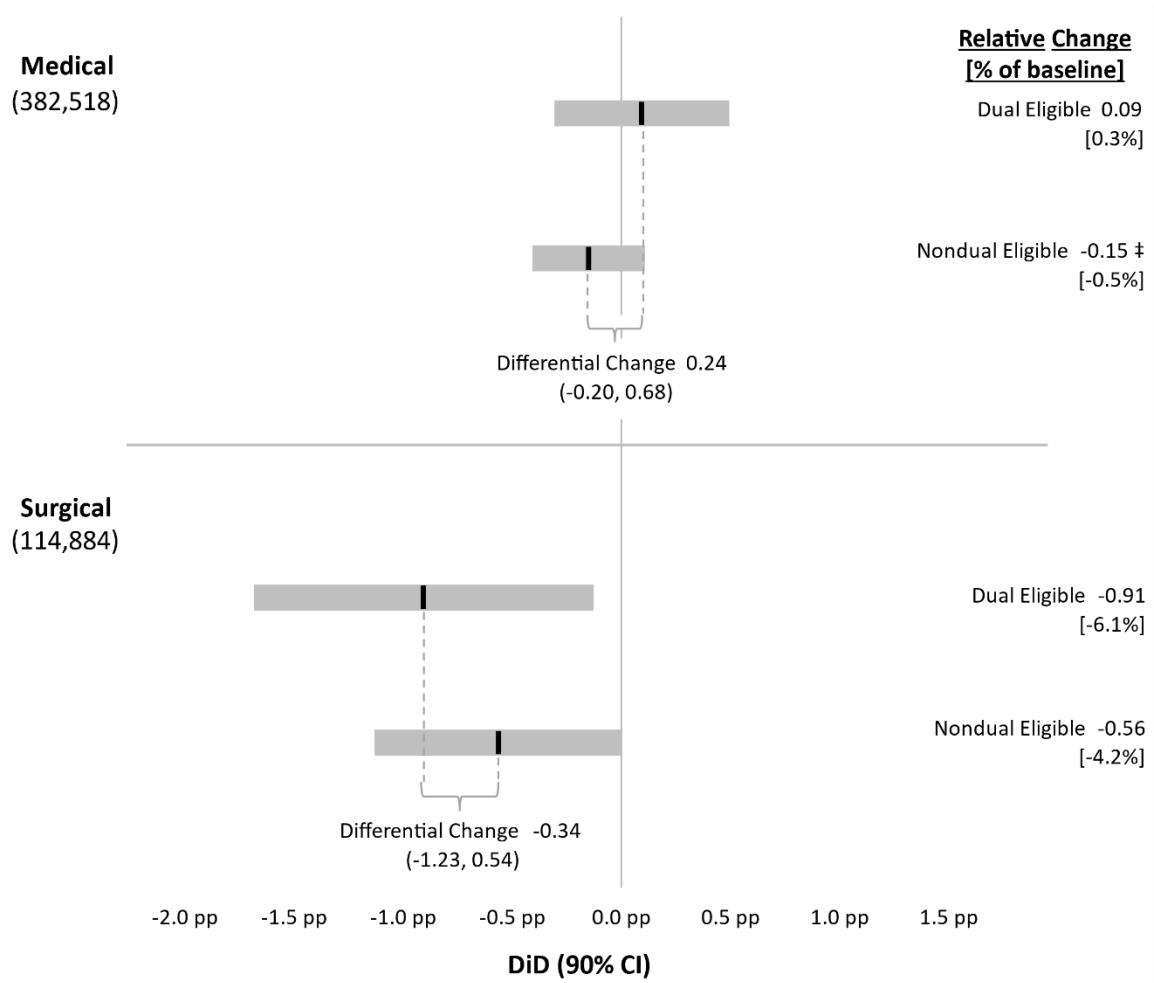
Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2019 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

For medical clinical episodes, although not statistically significant, the readmission rate increased by 0.09 pp (90% confidence interval: -0.31, 0.49; p=0.70), or 0.3% percent of the baseline mean, for dual-eligible beneficiaries relative to their comparison group in Model Years 1 and 2 (Exhibit 44). While also not statistically significant, the readmission rate declined by 0.15 pp (90% confidence interval: -0.41, 0.11; p=0.33), or 0.5%, for nondual-eligible beneficiaries in medical clinical episodes relative to their comparison group. The resulting differential change for dual-

eligible compared to nondual-eligible beneficiaries in medical clinical episodes was not statistically significant (estimate=0.24 pp; 90% confidence interval: -0.20, 0.68; p=0.36).

There was a statistically significant decline in the readmission rate for dual-eligible beneficiaries in surgical clinical episodes of 0.91 pp (90% confidence interval: -1.69, -0.13; p=0.06), or 6.1% in Model Years 1 and 2. The decline in the readmission rate for nondual-eligible beneficiaries was smaller in magnitude and was not statistically significant (estimate=-0.56 pp; 90% confidence interval: -1.13, 0.00; p=0.10). These declines resulted in a differential decrease of 0.34 pp (90% confidence interval: -1.23, 0.54; p=0.52) for dual-eligible beneficiaries compared to nondual-eligible beneficiaries in surgical clinical episodes, although this estimate is not statistically significant.

Exhibit 44: Change in the Unplanned Readmission Rate During the 90-day PDP Under BPCI Advanced for Dual-Eligible Beneficiaries and Nondual-Eligible Beneficiaries, Model Years 1-2, October 2018 – December 2019



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean readmission rates. The grey bars indicate the 90% confidence interval of the DiD estimates. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

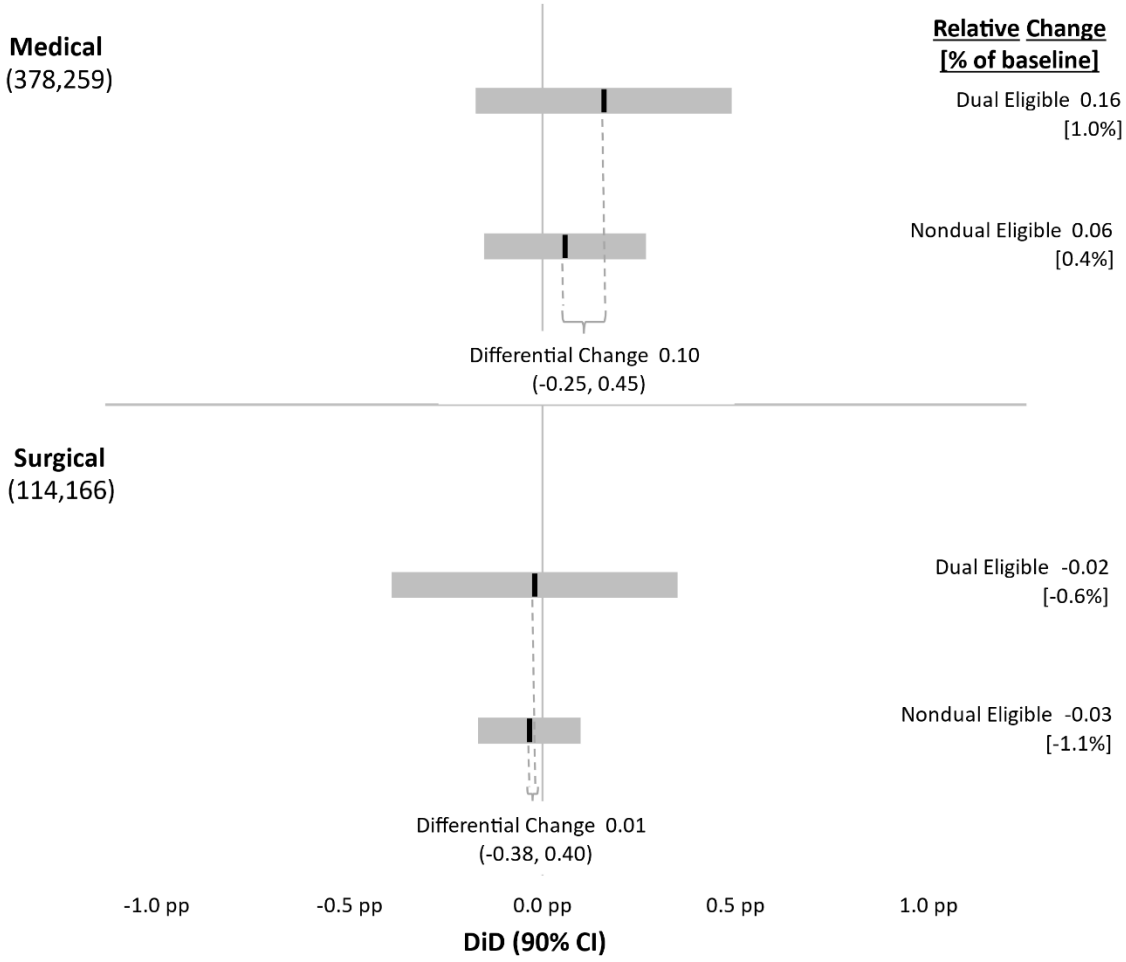
Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2019 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

In the analysis of the mortality rate during the 90-day PDP in Model Years 1 and 2, we do not find any statistically significant changes for dual- or nondual-eligible beneficiaries relative to their comparison groups (Exhibit 45). For medical clinical episodes, the mortality rate for dual-eligible beneficiaries increased by 0.16 pp (90% confidence interval: -0.17, 0.49; p=0.43), or 1.0%, relative to their comparison group. The mortality rate for nondual-eligible beneficiaries in medical clinical episodes increased by 0.06 pp (90% confidence interval: -0.15, 0.27; p=0.65), or 0.4%, relative to their comparison group.

For surgical clinical episodes, there was nearly no change in the mortality rate for dual-eligible beneficiaries (estimate=-0.02 pp; 90% confidence interval: -0.39, 0.35; p=0.92) or for nondual-eligible beneficiaries (estimate=-0.03 pp; 90% confidence interval: -0.17, 0.10; p=0.67) relative to their comparison groups in Model Years 1 and 2.

While the analysis of Model Year 3 found a statistically significant differential increase in the mortality rate for dual-eligible beneficiaries with medical clinical episodes compared to nondual-eligible beneficiaries, we did not observe differential impacts on the mortality rate in Model Years 1 and 2, mitigating concerns based on one single year of the model.

Exhibit 45: Change in the Mortality Rate During the 90-Day PDP for Dual-Eligible Beneficiaries and Nondual-Eligible Beneficiaries, Model Years 1-2, October 2018 – December 2019



Note: The estimates in this exhibit are the results of a difference-in-difference-in-differences (DiDiD) model. The DiDs (relative changes) and DiDiD (differential change) estimates represent the relative change in percentage points. Results are also presented as a percentage of the BPCI Advanced baseline mean mortality rates. The grey bars indicate the 90% confidence interval of the DiD estimates. See **Appendix C** for details of the DiD and DiDiD methodology, outcome definitions, and additional information on methods. See **Appendix K** for more detailed results. CI = confidence interval; DiD = difference-in-differences; PDP = post-discharge period; pp = percentage point(s).

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning October 1, 2018 and ending on or before December 31, 2019 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

3. Patient-reported functional status, care experiences, and overall satisfaction with care for populations that have been historically underserved

To evaluate whether quality of care varied for beneficiaries from historically underserved populations, we analyzed responses from the Model Year 4 beneficiary survey for historically

underserved populations with sufficient sample size to yield a minimum detectable difference of at least 10.0 pp. The historically underserved populations we analyzed include the following:

- Beneficiaries with hospital-initiated episodes who were:
  - Black or African American,
  - Hispanic,
  - Dually eligible for Medicare and Medicaid,
  - Lived in a ZIP code in the top 20% of the ADI,[41]
  - Lived in rural ZIP codes,[42]
- Beneficiaries with PGP-initiated episodes who:
  - Lived in rural ZIP codes.

For each historically underserved population, we pooled respondents across all 34 clinical episodes within each episode initiator type (hospital and PGP) and estimated differences between BPCI Advanced and comparison respondents (referred to as "differences"). We also discuss comparisons of the differences between each historically underserved population and their reference group (referred to as "differential impacts"). For example, we compare the difference between BPCI Advanced and comparison dual-eligible respondents to the difference between BPCI Advanced and comparison nondual-eligible respondents. Detailed results are presented in **Appendix L**.

[41] The ADI is a measure of neighborhood socioeconomic disadvantage. Neighborhoods with a higher ADI have lower neighborhood measures of income, education, employment, and housing quality. The ADI score, measured at the nine-digit zip code level, is defined in percentiles of the national population. Our subpopulation defined by ADI>=80 represents beneficiaries who reside in the top fifth most deprived neighborhoods in the country.

[42] Rurality was defined by the Federal Office of Rural Health Policy.

a. Key Findings

Patient-reported Outcomes Under BPCI Advanced for Underserved Populations

- Findings were neutral for patient-reported functional status for most of the historically underserved populations analyzed, and findings were generally less favorable or mixed for care experiences and satisfaction.
- Dual-eligible BPCI Advanced respondents with hospital-initiated episodes were slightly less likely to report favorable changes in functional status relative to dual-eligible comparison respondents. For nondual-eligible BPCI Advanced respondents with hospital episodes, there was no pattern of changes in functional status relative to comparison respondents. Therefore, BPCI Advanced had an unfavorable differential impact on functional status for dual-eligible respondents compared to the impact on nondual-eligible respondents.
- For respondents living in ZIP codes with a high Area Deprivation Index (ADI) and respondents living in rural ZIP codes (with either hospital or PGP episodes), BPCI Advanced had favorable differential impacts on functional status compared to the impacts on their reference populations. These favorable differential impacts were driven by unfavorable changes in functional status for the reference populations.
- BPCI Advanced respondents with hospital episodes who were (1) Hispanic, (2) lived in high-ADI ZIP codes, or (3) lived in rural ZIP codes were more likely to report unfavorable care experiences than their counterparts in the comparison group. Their reference populations (Non-Hispanic White, non-high ADI ZIP codes, or non-rural ZIP codes, respectively) reported mixed care experiences relative to their counterparts in the comparison group, resulting in larger unfavorable impacts on care experiences for the three underserved populations compared to their reference populations.
- Black or African American BPCI Advanced respondents and Hispanic BPCI Advanced respondents with hospital episodes were both less likely than comparison respondents to agree that medical staff took their preferences into account in deciding what services they should receive after leaving the hospital.
- Rural respondents with BPCI Advanced PGP episodes were more likely to report favorable care experiences and the highest levels of satisfaction with post-discharge care relative to comparison respondents.

b. How did patient-reported functional status change for BPCI Advanced respondents relative to comparison respondents from the historically underserved populations analyzed?

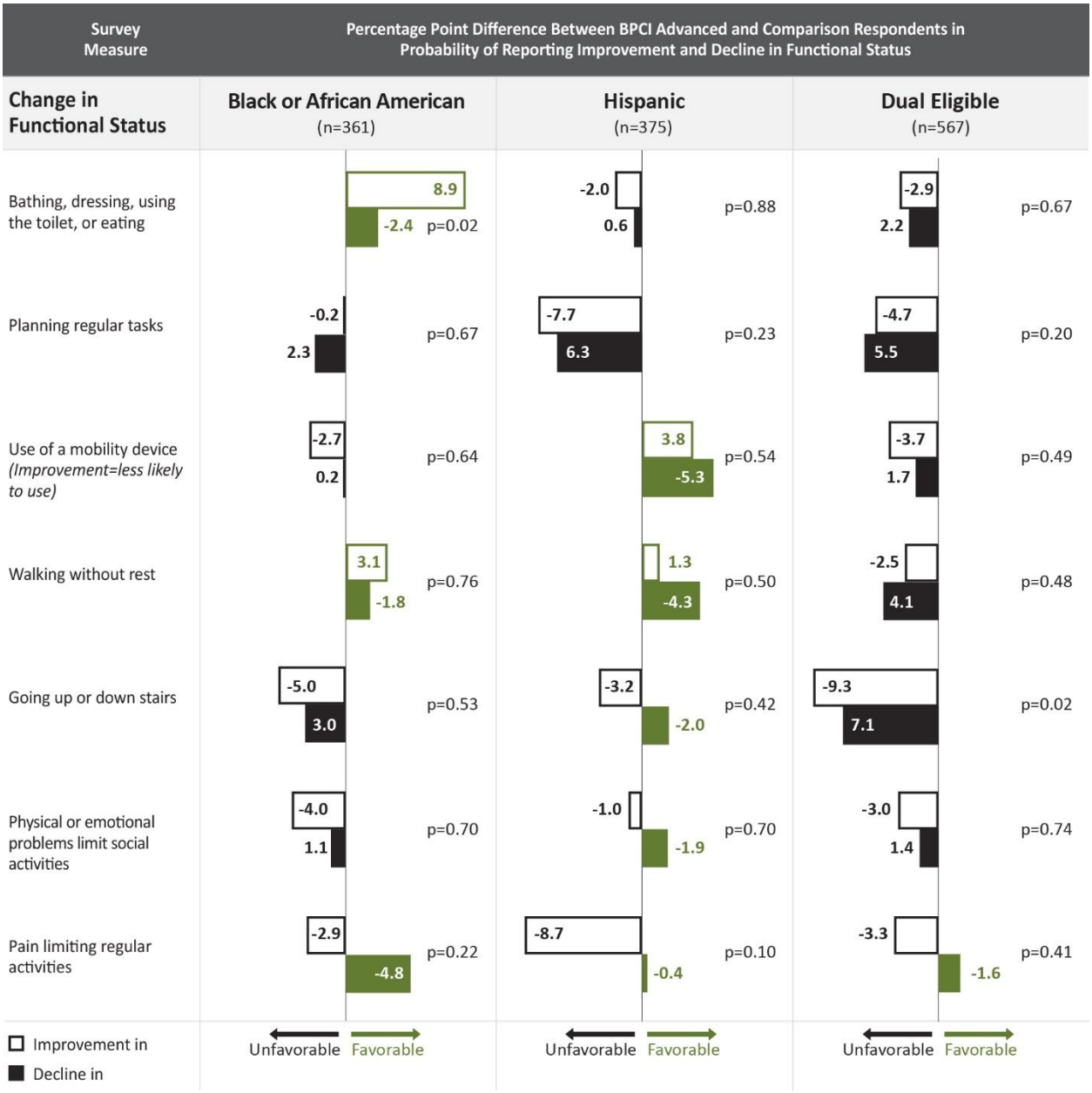
Dual-eligible BPCI Advanced respondents with hospital-initiated episodes were less likely to report favorable changes in functional status across all seven measures relative to dual-eligible comparison group respondents, and some results were large. Notably, dual-eligible BPCI Advanced respondents were 9.3 pp less likely to report improvement in the ability to go up or down stairs and 7.1 pp more likely to report decline relative to dual-eligible comparison respondents (p=0.02).

Results on patient-reported functional status for nondual-eligible BPCI Advanced respondents were mixed relative to nondual-eligible comparison respondents. This resulted in an unfavorable impact on patient-reported functional status for dual-eligible BPCI Advanced respondents compared to nondual-eligible comparison respondents, although only one measure was statistically significant: BPCI Advanced decreased the probability of dual-eligible respondents to report improvement in going up or down stairs by 7.9 pp more than the impact on nondual-eligible respondents (90% confidence interval: -14.4, -1.5 pp; p=0.08) (see **Appendix L**).

Among the other five historically underserved populations analyzed, there was generally no pattern of favorable or unfavorable changes in functional status relative to comparison respondents (Exhibit 46 and Exhibit 47). Within each population, BPCI Advanced respondents indicated some changes in functional status that were favorable relative to comparison respondents and some that were unfavorable, with few statistically significant differences.

BPCI Advanced had favorable differential impacts on patient-reported functional status for respondents with hospital-initiated episodes living in high-ADI ZIP codes and for respondents living in rural ZIP codes (with either hospital- or PGP-initiated episodes) compared to the impacts on their reference groups, although most differential impacts were not statistically significant. The BPCI Advanced Model had no differential impact on patient-reported functional status for Black or African American respondents and Hispanic respondents compared to the impact on non-Hispanic White respondents (see **Appendix L**).

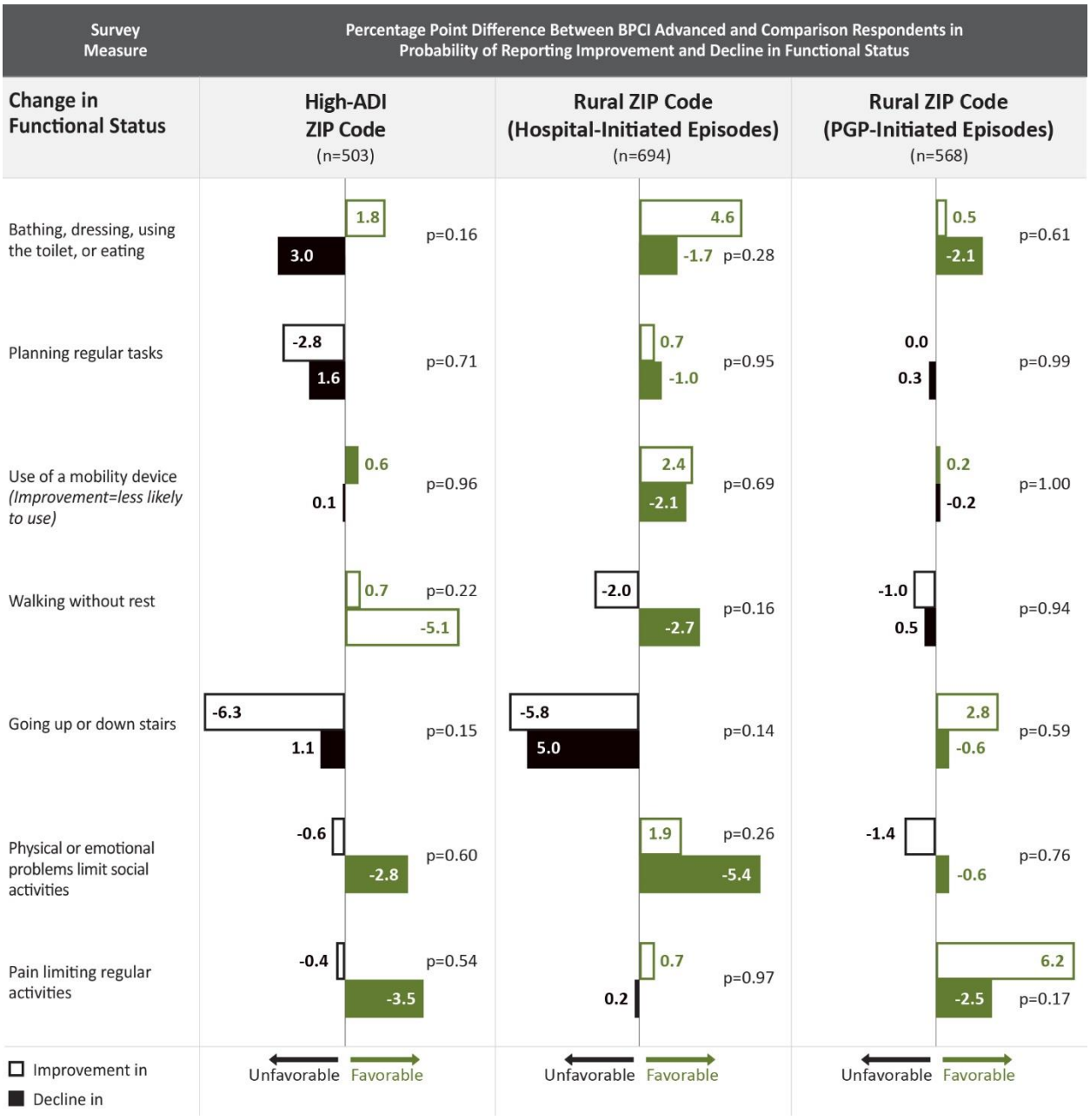
Exhibit 46: Differences in Patient-Reported Functional Status Between BPCI Advanced and Comparison Respondents Who are Black or African American, Hispanic, or Dual Eligible with Hospital-initiated Episodes, July 2021-August 2021



Note: The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. All responses were weighted for non-response and sampling design. Estimates were pooled across all 34 clinical episodes. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Master Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's race imputation algorithm based on beneficiaries' names and geography. Results are reported in percentage point terms. Values to the left, represented in black, indicate unfavorable results. Values to the right, represented in green, indicate favorable results. Thus, increases in decline are displayed on the left, and decreases in decline are displayed on the right. The p-values for functional status results indicate joint significance for differences in the proportion of respondents indicating one of three categories: improvement (or maintained highest function); stayed the same; or declined (or maintained lowest function). Sample sizes reflect the number of BPCI Advanced responses included in the analysis. Comparison sample sizes were roughly similar. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix L** for more detailed results.

***Source*:** The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Exhibit 47: Differences in Patient-Reported Functional Status Between BPCI Advanced and Comparison Respondents Residing in High-ADI and Rural ZIP Codes, July 2021 – August 2021



***Note*:** The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for trinary indicators. All responses were weighted for non-response and sampling design. Estimates were pooled across all 34 clinical episodes. The beneficiary ZIP code was used to determine high ADI and rural designations. Rurality was defined by the Federal Office of Rural Health Policy. High ADI indicated that a ZIP code was in the top 20% of the ADI. Results are reported in percentage point terms. Values to the left, represented in black, indicate unfavorable results. Values to the right, represented in green, indicate favorable results. Thus, increases in decline are displayed on the left, and decreases in decline are displayed on the right. The p-values for functional status results indicate joint significance for differences in the proportion of respondents indicating one of three

categories: improvement (or maintained highest function); stayed the same; or declined (or maintained lowest function). Sample sizes reflect the number of BPCI Advanced responses included in the analysis. Comparison sample sizes were roughly similar. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix L** for more detailed results. ADI = Area Deprivation Index; PGP = physician group practice.

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

c. *Did patient-reported care experiences and satisfaction differ for BPCI Advanced and comparison respondents from populations that have been historically underserved?*

BPCI Advanced respondents from historically underserved populations generally reported less favorable or mixed results for care experiences and satisfaction with care in Model Year 4. Respondents in three of the six historically underserved populations analyzed were more likely to report unfavorable care experiences across all eight care experience outcomes relative to their counterparts in the comparison group. These include respondents with hospital-initiated episodes who (1) were Hispanic, (2) lived in high-ADI ZIP codes, or (3) lived in rural ZIP codes (Exhibit 48 and Exhibit 49). Although differences were mostly not statistically significant, the changes were consistently large and unfavorable. For example, Hispanic BPCI Advanced respondents were 7.4 pp (90% confidence interval: -14.0, -0.8 pp; p=0.06) less likely than comparison respondents to agree that medical staff took their preferences into account in deciding what services they should receive after leaving the hospital.

BPCI Advanced had unfavorable differential impacts on care experiences for respondents from the three historically underserved populations discussed above—Hispanic, high-ADI, and rural (hospital-initiated episodes only)—compared to the impacts on their reference groups. Most differential impacts on care experiences were not statistically significant; however, BPCI Advanced decreased the probability of Hispanic respondents (relative to comparison respondents) to agree that they were able to manage health needs since returning home by 4.6 pp more than the impact on Non-Hispanic White BPCI Advanced respondents relative to their comparison group (90% confidence interval: -8.2, -0.9 pp; p=0.04). See **Appendix L** for detailed results for each historically underserved population and their reference group.

Results for Black or African American respondents were mixed relative to their counterparts in the comparison group, and most differences were small and not statistically significant. However, similar to Hispanic respondents, Black or African American BPCI Advanced respondents were 7.9 pp (90% confidence interval: -14.7, -1.2; p=0.05) less likely than their counterparts in the comparison group to agree that medical staff took their preferences into account when deciding what services they should receive after leaving the hospital. Differential impacts for Black or African American respondents compared to Non-Hispanic White respondents were mixed, and none were statistically significant.

Results for dual-eligible respondents with hospital-initiated episodes were mixed, relative to their counterparts in the comparison group, with unfavorable results for four care experience measures, a favorable result for one measure, and null results for three measures. No results were statistically significant. Differential impacts for dual-eligible respondents compared to nondual-eligible respondents were mixed, and none were statistically significant.

In contrast, although no results were statistically significant, rural respondents with PGP-initiated episodes were more likely to report favorable care experiences relative to their counterparts in the comparison group. Differential impacts on care experiences for rural respondents with PGP-initiated episodes compared to non-rural respondents with PGP-initiated episodes also were mixed, but one differential impact was statistically significant: BPCI Advanced increased the probability of rural respondents with PGP-initiated episodes to feel very or somewhat prepared to leave the hospital by 4.5 pp more than the impact on non-rural respondents with PGP-initiated episodes (90% confidence interval: 0.6, 8.4 pp; p=0.06).

For all six of the historically underserved populations analyzed, BPCI Advanced respondents were less likely than comparison respondents to indicate the highest levels of satisfaction with recovery, although the differences were not statistically significant. Similarly, for five of the six historically underserved populations analyzed, BPCI Advanced respondents were less likely than comparison respondents to report the highest levels of satisfaction with post-discharge care. Rural respondents with PGP-initiated episodes were the exception. In this population, BPCI Advanced respondents were 7.1 pp more likely than rural comparison respondents to report the highest levels of satisfaction with post-discharge care (p=0.02).

BPCI Advanced had unfavorable differential impacts on satisfaction with recovery or care for respondents in four of the six historically underserved populations analyzed—Black or African American respondents, Hispanic respondents, dual-eligible respondents, and respondents in high-ADI ZIP codes—compared to the impacts on their reference groups, although none of the differences were statistically significant.

The BPCI Advanced Model had slightly favorable differential impacts on satisfaction with recovery or care for rural respondents (with either hospital- or PGP-initiated episodes) compared to the impacts on their reference groups.

Exhibit 48: Differences in Care Experiences and Satisfaction Between BPCI Advanced and Comparison Respondents Who are Black or African American, Hispanic, or Dual Eligible with Hospital-initiated Episodes, July 2021 – August 2021

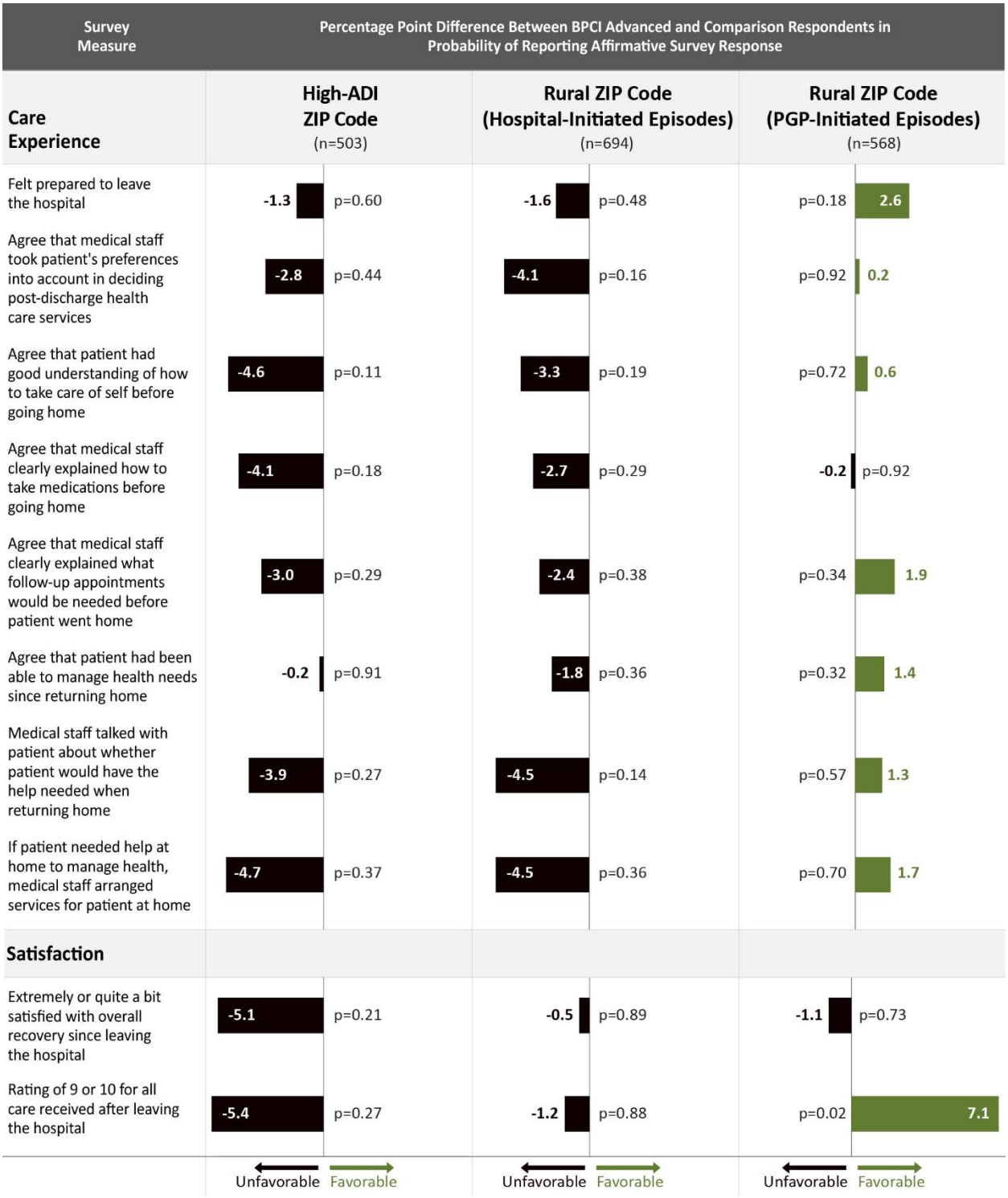
Survey Measure	Percentage Point Difference Between BPCI Advanced and Comparison Respondents in Probability of Reporting Affirmative Survey Response		
Care Experience	**Black or African American** (n=361)	**Hispanic** (n=375)	**Dual Eligible** (n=567)
Felt prepared to leave the hospital	-0.3 p=0.93	-3.1 p=0.28	p=0.35 1.9
Agree that medical staff took patient's preferences into account in deciding post-discharge health care services	-7.9 p=0.05	-7.4 p=0.06	-1.7 p=0.53
Agree that patient had good understanding of how to take care of self before going home	p=0.18 3.9	-2.7 p=0.29	-0.1 p=0.96
Agree that medical staff clearly explained how to take medications before going home	-0.5 p=0.88	-2.7 p=0.29	-1.8 p=0.40
Agree that medical staff clearly explained what follow-up appointments would be needed before patient went home	p=0.96 0.1	-0.2 p=0.94	0.0 p=1.00
Agree that patient had been able to manage health needs since returning home	-0.5 p=0.83	-3.1 p=0.12	p=0.93 0.2
Medical staff talked with patient about whether patient would have the help needed when returning home	-5.5 p=0.15	-4.6 p=0.13	-3.6 p=0.23
If patient needed help at home to manage health, medical staff arranged services for patient at home	p=0.99 0.1	-8.3 p=0.11	-3.5 p=0.37
Satisfaction			
Extremely or quite a bit satisfied with overall recovery since leaving the hospital	-4.7 p=0.32	-3.8 p=0.42	-1.1 p=0.75
Rating of 9 or 10 for all care received after leaving the hospital	-0.4 p=0.80	-2.8 p=0.67	-6.2 p=0.28
	Unfavorable Favorable	Unfavorable Favorable	Unfavorable Favorable

***Note*:** The estimates in this exhibit are the result of a cross-sectional, risk-adjusted multinomial logistic regression model for binary indicators. All responses were weighted for non-response and sampling design. Estimates were pooled across all 34 clinical episodes. The race and ethnicity data come from the Research Triangle Institute (RTI) race codes from the Master Beneficiary Summary File. The RTI race code is created based on beneficiaries' self-reporting to Medicare and the Social Security Administration, and RTI's

race imputation algorithm based on beneficiaries' names and geography. Results are reported in percentage point terms. Values to the left, represented in black, indicate unfavorable results. Values to the right, represented in green, indicate favorable results. The p-value for satisfaction with post-discharge care indicates joint significance for differences in the proportion of respondents indicating one of three categories: 9-10 rating; 7-8 rating; 0-6 rating. Sample sizes reflect the number of BPCI Advanced responses included in the analysis. Comparison sample sizes were roughly similar. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix L** for more detailed results.

Source: The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

Exhibit 49: Differences in Care Experiences and Satisfaction Between BPCI Advanced and Comparison Respondents Residing in Rural or High-ADI ZIP Codes, July 2021 – August 2021



***Note*:** The estimates in this exhibit are the result of a cross-sectional, multinomial logistic regression risk adjustment model for trinary indicators. All responses were weighted for non-response and sampling design. Estimates were pooled across all 34 clinical episodes. The beneficiary ZIP code was used to determine high ADI and rural designations. Rurality was defined by the Federal

Office of Rural Health Policy. High ADI indicated that a ZIP code was in the top 20% of the ADI. Results reported in percentage point terms. Values to the left, represented in black, indicate unfavorable results. Values to the right, represented in green, indicate favorable results. The p-value for satisfaction with post-discharge care indicates joint significance for differences in proportion of respondents indicating one of three categories: 9-10 rating; 7-8 rating; 0-6 rating. Sample sizes reflect the number of BPCI Advanced responses comprising analysis. Comparison sample sizes were roughly similar. See **Appendix C** for details of the risk-adjustment methodology, outcome definitions, and additional information on methods. See **Appendix L** for more detailed results. ADI = Area Deprivation Index; PGP = physician group practice.

***Source*:** The BPCI Advanced evaluation team's analysis of BPCI Advanced beneficiary survey responses collected from October 2021 through February 2022 for hospital discharges or outpatient procedures that occurred in July or August 2021.

III. Discussion and Conclusion

A. Discussion

The BPCI Advanced Model tests whether bundling Medicare payments for an episode of care can reduce Medicare expenditures while maintaining or improving quality of care. BPCI Advanced builds on the lessons learned from earlier bundled payment models, primarily the BPCI Initiative Model 2. Its refined payment approach is intended to expand provider participation as well as increase the likelihood that the Medicare program will achieve savings. In addition, performance on select quality metrics adjusts reconciliation payments by up to 10%, which qualifies BPCI Advanced as an Advanced APM and was intended to be an incentive to further boost participation. Physicians who achieve threshold levels of payments or patients through Advanced APMs may be eligible for the 5% APM incentive payment and may be excluded from the Merit-based Incentive Payment System (MIPS) reporting requirements and payment adjustment. However, as reported in the First BPCI Advanced Evaluation Report, the Advanced APM feature of the model was not a significant driver in decisions to participate in BPCI Advanced or in selecting clinical episodes among participants and episode initiators interviewed because most interviewees were uncertain whether clinicians would reach volume thresholds to qualify for Advanced APM incentive payments.[43] The quality adjustment is also intended to reinforce the quality aims of the model.

There was widespread participation in BPCI Advanced, with 1,295 hospital and physician group practice (PGP) episode initiators in Model Years 1 and 2 (2018 and 2019), which almost doubled to 2,041 episode initiators in Model Year 3 (2020).[44] At the start of Model Year 3, episode initiators selected to participate in 6.2 clinical episodes on average, with a total of 12,716 unique episode initiator-clinical episode combinations. Potential participants were provided preliminary target prices and data on their historical episode payments, which allowed them to evaluate their opportunities for achieving payment and quality goals prior to joining BPCI Advanced and selecting clinical episodes. Episode initiators voluntarily joined the model and elected to participate in one or more clinical episodes. Some episode initiators received guidance from conveners or third-party consultants to help them identify which clinical episodes presented the biggest financial opportunity. Other episode initiators saw the model as a learning opportunity and elected to participate in multiple clinical episodes regardless of the probability of financial gain in order to get experience with bundled payments and value-based care models. However, the COVID-19 public health emergency (PHE), which began in January 2020, presented a number of challenges to the broader health care system and to BPCI Advanced participants. CMS allowed participants to select participation agreement amendments to remove only episodes with a COVID-19 diagnosis or all episodes from reconciliation for Model Year 3. Even so, some participants chose to withdraw from the model given the challenges due to the PHE and other adjustments to the model that were implemented in Model Year 4 (2021), including changes to the target pricing methodology and clinical episode groupings, which were intended to bolster the model's ability to achieve Medicare savings. By the beginning of Model Year 4, there were 1,205 hospital and PGP episode initiators participating in the model. Although there were fewer episode initiators, due to the introduction of CESLGs in Model Year 4, the average number of clinical episodes selected by each episode

[43] The First Evaluation Report is available for download at https://innovation.cms.gov/innovation-models/bpci-advanced.

[44] In contrast, there were only 712 hospital and PGP episode initiators that participated in the BPCI Initiative Model 2.

initiator increased to 13.0, and the total unique number of episode initiator-clinical episode combinations increased to 15,684.

Overall, for the clinical episodes evaluated, the BPCI Advanced Model continued to achieve statistically significant reductions in average standardized episode payments in Model Year 3. The reduction in per-episode payments was over twice as large for surgical clinical episodes as it was for medical clinical episodes. These findings were similar in Model Years 1 and 2, as reported in the previous BPCI Advanced evaluation report.[45] For medical clinical episodes, hospital and PGP episode initiators reduced episode payments by similar amounts, but for surgical clinical episodes, the reduction in payments for PGP episode initiators was over twice as large as for hospital episode initiators.

Consistent with earlier analyses and other episode-based payment approaches, payment reductions in Model Year 3 were primarily due to lower payments for more intensive post-acute care (PAC) settings, particularly for skilled nursing facilities (SNF) and inpatient rehabilitation facilities (IRF).[46] This was true for both hospital and PGP episode initiators, although they differed in how they achieved these reductions. For medical clinical episodes, hospital and PGP episode initiators reduced the share of episodes first discharged to an institutional PAC setting, but hospital episode initiators had a larger reduction in the number of SNF days among episodes with at least one day of SNF use. For surgical clinical episodes, the reduction in the share of institutional PAC discharges was larger for PGPs than hospitals, but the reduction in the number of SNF days among episodes with at least one day of SNF use was similar for hospitals and PGPs. Unlike in previous analyses, however, there was less evidence that hospital episode initiators substituted lower institutional PAC use with increased HH use: there was no detectable change in HH payments for hospital medical or hospital surgical episodes in Model Year 3. This may be associated with the COVID-19 PHE, as some BPCI Advanced participants noted that patients were hesitant to allow HH staff in their homes after the onset of the PHE.

As in previous years, BPCI Advanced did not negatively affect quality of care in Model Year 3, and there was evidence of an improvement in quality as measured in claims data by unplanned readmission and mortality rates, despite declines in PAC use. The unplanned readmission rate and the mortality rate declined for PGP medical clinical episodes, and the readmission rate declined for episodes pooled across episode initiator type and across the clinical episodes evaluated. In Model Years 1 and 2, there was an improvement in the readmission rate for PGP surgical clinical episodes, and no impact on the readmission rate for PGP medical clinical episodes. The changes in impact estimates between Model Years 1 and 2 (2018-2019) and Model Year 3 (2020) may be due to the COVID-19 PHE, changes in participants and comparison groups, or fluctuations in the data. For example, there was a sharp decline in surgical volume in April 2020 in response to the

[45] The Third Evaluation Report is available for download at https://innovation.cms.gov/innovation-models/bpci-advanced.

[46] For findings from the evaluation of prior BPCI Advanced model years, see the Third Evaluation Report, available for download at https://innovation.cms.gov/innovation-models/bpci-advanced. For findings from the evaluation of the BPCI Initiative, see the CMS Bundled Payments for Care Improvement Initiative Models 2-4: Year 7 Evaluation & Monitoring Annual Report, available for download at https://innovation.cms.gov/innovation-models/bundled-payments. For findings from the evaluation of the Comprehensive Care for Joint Replacement Model, see the CMS Comprehensive Care for Joint Replacement Model: Performance Year 4 Evaluation Report, available for download at https://innovation.cms.gov/data-and-reports/2021/cjr-py4-annual-report.

pandemic when many hospitals paused elective surgical procedures. Several BPCI Advanced participants noted during site visits and interviews that the patients who did come to the hospital were of higher acuity after having postponed or delayed care during the pandemic. Interviewees also reported experiencing workforce redeployments, staff burnout, and general labor shortages, which impacted their BPCI Advanced activities, including strategies to reduce readmission rates. In addition, the estimated impacts for Model Years 1 and 2 and Model Year 3 may differ because a second cohort of participants was allowed to join BPCI Advanced in January 2020 (Model Year 3) and participants from the first cohort were able to reselect clinical episodes. Thus, some of the BPCI Advanced episode initiators and comparison providers differed between the years. Finally, the changes in impact estimates from Model Years 1 and 2 to Model Year 3 may be due to other fluctuations in data. We will continue to evaluate the impact of BPCI Advanced on the unplanned readmission rate in future reports.

In contrast to prior evaluation reports, there is some evidence from Model Year 4 that BPCI Advanced beneficiaries that responded to the survey reported slightly less favorable results for functional status, care experiences, and satisfaction with care relative to respondents in a matched comparison group. BPCI Advanced respondents with hospital-initiated episodes and those with PGP-initiated episodes were both slightly less likely to report improvements in functional status (or more likely to report declines) than comparison respondents in Model Year 4. In addition, BPCI Advanced respondents with hospital episodes were less likely to report favorable care experiences and the highest levels of satisfaction with care in Model Year 4. For BPCI Advanced respondents with PGP episodes, there was a mix of favorable and unfavorable results for care experiences and satisfaction with care across clinical episode service line groups (CESLGs). Differences in outcomes between BPCI Advanced and comparison respondents were small, approximately 1 to 2 percentage points (pp) on average. However, small differences in patient-reported outcomes reflect meaningfully large differences that affect a small proportion of individuals. The survey measures were constructed from questions with distinct response categories, such as, agree versus disagree, some help needed with task versus complete help needed with task. Average differences in functional status and care experiences roughly translate to one or two additional BPCI Advanced respondents out of 100 reporting an unfavorable outcome relative to comparison respondents.

This is the first time under BPCI Advanced or BPCI that there is evidence of unfavorable changes in functional status. Neither the BPCI Advanced Model Year 2 (2019) beneficiary survey results nor multiple years of survey results from the evaluation of BPCI Model 2 showed evidence of unfavorable functional status results among BPCI Advanced or BPCI respondents relative to comparison respondents. While there was no consistent pattern in care experiences or satisfaction with care in Model Year 2 of BPCI Advanced, there was some evidence that BPCI respondents had unfavorable care experiences and satisfaction with care relative to comparison respondents under Model 2 of the BPCI Initiative. There were changes made to the BPCI Advanced Model in Model Year 4, including the requirement to participate in broad CESLGs rather than individual clinical episodes, which could have contributed to the unfavorable results presented in this report. We will continue to field surveys in Model Year 5 (2022), and the results, to be reported in the next evaluation report, will provide additional evidence on the impact of the model on functional status, care experiences, and satisfaction with care.

Although BPCI Advanced expanded participation and achieved lower episode payments across multiple clinical episodes, it was only partly successful in reducing spending by the Medicare

program. During Model Year 3, the BPCI Advanced Model generally resulted in estimated net losses after considering reconciliation payments for both hospital and PGP medical episodes and estimated net savings for hospital and PGP surgical episodes, with an overall estimated net loss to the Medicare program of $113.7 million, or 0.8% of Medicare program payments under the counterfactual (or what payments would have been if the BPCI Advanced Model had not occurred). Compared to Model Years 1 and 2, Medicare losses were slightly larger in Model Year 3. Although the declines in average episode payments were larger in Model Year 3, episode volume was lower, resulting in a smaller decline in non-standardized payments in Model Year 3 than in Model Years 1 and 2. In addition, reconciliation payments per episode to participants were larger in Model Year 3 compared to Model Years 1 and 2, and despite lower episode volume, total reconciliation payments were larger in Model Year 3.

With few exceptions, the evidence suggests that target prices were too high for medical clinical episodes for both hospital and PGP episode initiators. For surgical clinical episodes, the evidence generally suggests target prices were reasonably accurate. These are similar to the findings from Model Years 1 and 2. The target price methodology remained unchanged in Model Years 1 through 3, but in MY4, CMS changed the methodology with the intention of improving the model's ability to achieve Medicare savings. Based on preliminary data, per-episode reconciliation payments were lower in Model Year 4 than in prior model years. If the evaluation continues to estimate reductions in Medicare fee-for-service (FFS) payments for BPCI Advanced participants, the estimated financial outcome for Medicare may be more favorable in Model Year 4 than in Model Years 1 and 2 and Model Year 3. The Fifth Evaluation Report will report estimates of Medicare program savings in Model Year 4.

We evaluated the impact of BPCI Advanced on outcomes for beneficiaries from populations that have been historically underserved. In Model Years 1 and 2 and in Model Year 3, there were larger relative declines in episode payments for the two historically underserved populations evaluated, Black or African American beneficiaries (compared to Non-Hispanic White beneficiaries) and beneficiaries dually eligible for Medicare and Medicaid (compared to nondual-eligible beneficiaries), in both medical and surgical clinical episodes. There were no changes in the readmission rate for underserved populations relative to their comparison groups in Model Year 3, but in Model Years 1 and 2, the readmission rate declined for dual-eligible beneficiaries with surgical episodes relative to the comparison group. There were no relative changes in the mortality rate for Black or African American beneficiaries or for dual-eligible beneficiaries relative to their comparison groups in either Model Years 1 and 2 or Model Year 3. However, there were differential increases in the mortality rate for Black or African American beneficiaries with medical episodes compared to Non-Hispanic White beneficiaries and for dual-eligible beneficiaries with medical episodes compared to nondual-eligible beneficiaries in Model Year 3. These differential increases are largely a result of statistically significant declines in the mortality rate for the reference populations, with no statistically significant change in the mortality rate among BPCI Advanced beneficiaries in the underserved population. These results suggest that Black or African American beneficiaries and dual-eligible beneficiaries were less likely than their reference populations to benefit from the BPCI Advanced Model during Model Year 3.

In addition, we analyzed responses to the survey for beneficiaries from historically underserved populations during Model Year 4. Findings were neutral for patient-reported functional status for all historically underserved populations analyzed except one. Dual-eligible BPCI Advanced

respondents with hospital-initiated episodes were less likely to report favorable changes in functional status relative to comparison group respondents in Model Year 4. The average difference across outcomes was approximately 4 pp, indicating that roughly four additional dual-eligible BPCI Advanced respondents out of 100 reported unfavorable outcomes, relative to the comparison group. For nondual-eligible BPCI Advanced respondents, there was no pattern of changes in functional status relative to their comparison group. Therefore, BPCI Advanced had an unfavorable differential impact on functional status for dual-eligible respondents compared to nondual-eligible respondents, an average differential of roughly 3 pp.

For two historically underserved populations, respondents living in high-ADI ZIP codes and respondents living in rural ZIP codes (with either hospital- or PGP-initiated episodes), BPCI Advanced had favorable differential impacts on patient-reported functional status compared to the impacts on their reference groups. These favorable differential impacts were driven by unfavorable changes in functional status for the reference populations and were roughly 2 to 3 pp in magnitude, on average.

Findings were generally less favorable or mixed for care experiences and satisfaction with care. BPCI Advanced respondents with hospital episodes in some populations (Hispanic beneficiaries, those who lived in rural ZIP codes, and those who lived in high-ADI ZIP codes) were less likely to report favorable care experiences in Model Year 4 relative to comparison respondents. For example, Black or African American BPCI Advanced respondents and Hispanic BPCI Advanced respondents with hospital-initiated episodes were 7 to 8 pp less likely than comparison respondents to agree that medical staff took their preferences into account in deciding what services they should receive after leaving the hospital. Rural beneficiaries with BPCI Advanced PGP-initiated episodes were the exception: in Model Year 4, they were more likely to report favorable care experiences and the highest levels of satisfaction with post-discharge care relative to comparison respondents.

B. Limitations

We estimated the impact of the model using a difference-in-differences (DiD) design, which is dependent on a comparison group that represents what would have happened to episodes treated by BPCI Advanced episode initiators if the model had never existed. Thus, we selected a matched comparison group that was similar to BPCI Advanced providers on key factors expected to influence payment, utilization, and quality outcomes. To identify matched providers that were balanced with BPCI Advanced providers across various characteristics, a subset of BPCI Advanced episode initiators had to be excluded from our impact estimates. Across the clinical episodes evaluated, 69.2% to 82.7% of BPCI Advanced hospital episode initiators were included in the analysis, and 79.4% to 100.0% of PGP episode initiators were included. To estimate net savings to Medicare, we extrapolated these analyses to all episode initiators that initiated episodes in the clinical episodes we evaluated. Sensitivity testing that included all episode initiators showed that our findings were robust to alternative samples.

The share of BPCI Advanced episode initiators in the evaluation sample presented above illustrates the challenges of evaluating the BPCI Advanced Model. As described in prior evaluation reports, the primary difficulty is finding a group of comparison providers that are similar to BPCI Advanced hospital and PGP episode initiators. Because the BPCI Advanced Model has broad voluntary participation, there are fewer non-participating hospitals and PGPs that are unassociated

with the model and are also similar to participants on key characteristics. Constructing comparison groups for PGP episode initiators introduced additional challenges. PGP episode initiators were able to form new tax identification numbers (TINs) specifically to participate in BPCI Advanced, with new clinician members and ownership status. As a result, there was no baseline claims data to use for matching purposes for some PGP episode initiators. To preserve as many PGP episode initiators as we could in our sample, we linked new PGP episode initiators to baseline data based on their clinician members, allowing us to retain 85.0% of BPCI Advanced episodes initiated in the clinical episodes evaluated.

We assess the impact of BPCI Advanced for a subset of clinical episodes due to limited sample size or difficulty identifying a suitable matched comparison sample of providers. We evaluated as many clinical episodes as possible, resulting in 17 clinical episodes for hospital episode initiators, which account for 91.1% of all hospital episodes, and 17 clinical episodes for PGP episode initiators, which account for 94.0% of all PGP episodes. In future reports, we will continue to estimate the impact of the model on as many CESLGs as possible.

As noted above, the DiD design is dependent on a comparison group that represents what would have happened in the absence of the BPCI Advanced Model. A fundamental assumption for this to hold is that outcomes for BPCI Advanced and comparison episodes had parallel trends during the baseline period (often referred to as the parallel trends assumption). We test this assumption by conducting a statistical test to determine whether the quarterly baseline trends for BPCI Advanced and the comparison group were the same. Trends that are not parallel during the baseline suggest that the estimated impacts may not be entirely due to the BPCI Advanced Model but rather to differences between BPCI Advanced episode initiators and the comparison group that existed prior to the model. That is, non-parallel trends during the baseline period may provide evidence that the impact estimates are biased; there may be an impact due to the model, but it may be smaller or larger than the impact estimated. See **Appendix C** for additional details of the parallel trends testing methodology.

We constructed comparison groups by matching BPCI Advanced episode initiators to non-participating providers at the clinical-episode level, separately for hospitals and PGPs, and we performed the impact analysis by pooling episodes across the clinical episodes evaluated. We found evidence of non-parallel trends in only 7% of the clinical-episode level outcomes tested, at the 10% level of statistical significance, supporting the validity of our comparison group selection approach. However, when we tested for parallel trends during the baseline period for each pooled grouping, we found evidence that the trends were not parallel for some groupings and for some outcomes, including total allowed episode payments, readmission payments, and the unplanned readmission rate during the 90-day PDP, among others. In all cases, the parallel trends coefficients were negative, indicating the outcomes were declining at a faster rate during the baseline for BPCI Advanced than for the comparison group. This implies that the estimated reductions in these measures may overstate the true model impact. See **Appendix G** for the results of the parallel trends tests.

While some of our findings may be an overestimate of the BPCI Advanced Model impact, the overall conclusions are reliable. The estimates are robust to sensitivity tests and the main drivers of the reductions in total allowed payments generally do not have parallel trends failures. SNF, IRF, and HH payments and the share of episodes first discharged into an institutional PAC setting often

passed the parallel trends test for groupings by episode initiator type (e.g., hospital medical clinical episodes, surgical PGP clinical episodes), the analyses of which provide important information on how BPCI Advanced episode initiators achieved reductions in total episode payments.

In addition, we performed a series of sensitivity analyses of the total episode payments outcome. First, we calculated the impact estimates for total episode payments after subtracting the potential gap between BPCI Advanced and the comparison group in the intervention period caused by the divergence in the baseline trends, referred to as the potential bias to the impact estimate (Exhibit 50).[47] We estimated the potential bias to be -$255 for episodes pooled across the clinical episodes evaluated, which means the impact estimate reported may be inflated by $255. Subtracting the potential bias from the impact estimate resulted in a per-episode reduction in payments of $773, or 2.9% of the baseline mean. This calculation did not change the inference of the confidence intervals—the upper and lower bounds of the confidence intervals less the potential bias were still less than $0, indicating that BPCI Advanced resulted in a decline in total episode payments. For medical clinical episodes, the potential bias was -$273, resulting in an estimated impact less the bias of $523 or 2.0%. The potential bias also did not change the inference of the confidence intervals, indicating that BPCI Advanced resulted in a decline in episode payments for medical clinical episodes. For surgical clinical episodes, the potential bias was -$171, resulting in an estimated impact less the bias of $1,629 or 5.2%. The potential bias also did not change the inference of the confidence intervals, indicating that BPCI Advanced resulted in a decline in episode payments.

Exhibit 50: Comparison of the Potential Bias and the Impact of BPCI Advanced on Total Payments, Hospital and PGP Episode Initiators, Model Year 3 (January 1, 2020 – December 31, 2020)

Clinical Episode	BPCI Advanced DiD	Potential Bias	DiD Less Potential Bias	BPCI Advanced DiD CI	
				90% LCI	90% UCI
All Clinical Episodes	-$1,028 ‡	-$255	-$773	-$1,205	-$852
Medical	-$796 ‡	-$273	-$523	-$974	-$619
Hospital	-$756 ‡	-$200	-$556	-$950	-$563
PGP	-$667	-$244	-$423	-$1,025	-$310
Surgical	-$1,800	-$171	-$1,629	-$2,149	-$1,451
Hospital	-$933	-$133	-$800	-$1,534	-$331
PGP	-$2,147	-$147	-$2,000	-$2,552	-$1,741

***Note*:** Total payments represent Part A and B fee-for-service payments for the anchor stay or procedure and the 90-day post-discharge period. This payment outcome is standardized to remove the effect of geographic and other payment adjustments. The potential bias was calculated as the product of the parallel trends coefficient times the number of quarters since the baseline period (12 quarters). The DiD less potential bias was calculated as the DiD estimate minus the potential bias. The BPCI Advanced DiD estimates and the 90% confidence intervals are the results of a DiD model on the matched BPCI Advanced and selected comparison group sample. See **Appendix C** for additional details related to methodology. The DiD estimates represent the relative change in dollars. CI = confidence interval; DiD = difference-in-differences; LCI = lower confidence interval; PGP = physician group practice; UCI = upper confidence interval.

[47] The potential bias represents the amount by which BPCI Advanced payments would have continued to decline if they had remained on the same baseline trend. We calculated the potential bias as the coefficient from the parallel trends test for the outcome (which represents the average quarterly decline in the baseline period) times the number of quarters since the baseline period (12 quarters). See **Appendix G** for the parallel trends coefficients.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

We also compared the results of our pooled estimates to the volume-weighted average of the clinical episode-level impact estimates. We found that the impact estimates were robust to the weighted-average results, but the weighted-average results tended to be smaller by 2% to 18%, depending on the grouping (Exhibit 51). Next, we calculated impact estimates for medical and surgical clinical episode groupings without the use of a matched comparison group. This alternative sample included all episodes initiated by BPCI Advanced hospitals and PGPs within the clinical episodes evaluated (whether the episode initiators were matched to comparison providers), and all episodes initiated at non-participating hospitals and PGPs (referred to as the "national comparison sample"). For medical clinical episodes, the national comparison sample resulted in a decline in total allowed payments that was 9.2% smaller in magnitude than the reported impact estimates. For surgical clinical episodes, the national comparison sample resulted in a decline that was 28.3% larger in magnitude than the reported impact estimates. (Both the medical and surgical clinical episodes failed the parallel trends test when using the national comparison sample.) Overall, the results of these sensitivity analyses indicate that while the true impact of BPCI Advanced on total episode payments may be smaller than the estimates reported, we can be reasonably confident that impact of BPCI Advanced on total episode payments during Model Year 3 was meaningfully large.

Exhibit 51: Comparison of the Impact of BPCI Advanced on Total Payments, Hospital and PGP Episode Initiators, Model Year 3 (January 1, 2020 – December 31, 2020)

Clinical Episode	Weighted Average	National Comparison Sample	BPCI Advanced DiD	90% LCI	90% UCI
All Clinical Episodes	-$872	-	-$1,028 ‡	-$1,205	-$852
Medical	-$685	-$723 ‡	-$796 ‡	-$974	-$619
Hospital	-$750	-	-$756 ‡	-$950	-$563
PGP	-$550	-	-$667	-$1,025	-$310
Surgical	-$1,673	-$2,310 ‡	-$1,800	-$2,149	-$1,451
Hospital	-$859	-	-$933	-$1,534	-$331
PGP	-$2,097	-	-$2,147	-$2,552	-$1,741

***Note*:** Total payments represent Part A and B fee-for-service payments for the anchor stay or procedure and the 90-day post-discharge period. This payment outcome is standardized to remove the effect of geographic and other payment adjustments. The weighted average indicates the volume weighted average of the clinical episode-level estimates, where the volume was calculated using the BPCI Advanced Model Year 3 volume. The national comparison sample is the result of a DiD model without matching, including episodes from BPCI Advanced episode initiators that initiated episodes in both the baseline period and Model Year 3 and episodes initiated at non-participating hospitals and PGPs. The BPCI Advanced DiD estimates and the 90% confidence intervals are the results of a DiD model on the matched BPCI Advanced and selected comparison group sample. The DiD estimates represent the relative change in dollars. See **Appendix C** for additional details related to methodology. DiD = difference-in-differences; LCI = lower confidence interval; PGP = physician group practice; UCI = upper confidence interval.

‡ We rejected the null hypothesis that BPCI Advanced and matched comparison providers had parallel trends for this outcome (with 90% confidence). See **Appendix G** for parallel trends test results.

Source: The BPCI Advanced evaluation team's analysis of Medicare claims and enrollment data for episodes with anchor stays or procedures beginning April 1, 2013 and ending on or before December 31, 2017 (baseline period) and episodes with anchor stays or

procedures beginning January 1, 2020 and ending on or before December 31, 2020 (intervention period) for BPCI Advanced episode initiators and matched comparison providers.

Our analysis of the beneficiary survey identified differences in responses between BPCI Advanced and comparison respondents. Because the survey data were only collected during the intervention period, we cannot determine whether these differences were pre-existing during the baseline period, or if they were caused by the BPCI Advanced Model. Furthermore, given the smaller sample sizes for the beneficiary survey, particularly for the analyses of the historically underserved populations, we may not be able to draw conclusions about statistical significance for results that are small in magnitude. For analysis of individual CESLGs and historically underserved populations, we are sufficiently powered to estimate minimum detectable differences between BPCI Advanced and the comparison group in the range of 5 to 10 pp, but consistent differences of lesser magnitude may be meaningful and important.

In addition, due to the nature of primary data collection, the beneficiary survey may not be representative of all BPCI Advanced beneficiaries. First, the beneficiary survey was collected from episodes spanning two months (beneficiaries with episodes initiated in July and August 2021). Thus, the BPCI Advanced beneficiaries in the survey sample are not guaranteed to be representative of all the beneficiaries in BPCI Advanced during Model Year 4, which spans the full calendar year (January through December 2021). Second, response rates for the beneficiary survey were approximately 27% for BPCI Advanced beneficiaries attributed to hospitals and 29% for BPCI Advanced beneficiaries attributed to PGPs and were generally lower among beneficiaries from historically underserved populations. While we applied non-response and sampling weights to all observations to make our respondents reflect the overall BPCI Advanced population on key beneficiary-, hospital-, and neighborhood-level characteristics, we cannot guarantee that our results are generalizable to the remaining BPCI Advanced beneficiaries that did not respond to the survey.

Our conclusion that BPCI Advanced resulted in net losses to Medicare during Model Year 3 is based on several assumptions. First, it is based on the DiD estimates, which may be biased because BPCI Advanced and the comparison group were not on parallel trends during the baseline period for some of the pooled estimates. However, the conclusion that BPCI Advanced resulted in losses is not affected. This potential bias from non-parallel trends would result in our estimated reduction in payments being overestimated, or larger than the true model's reduction in payments. If this were the case, this would result in even larger losses to Medicare than those presented in this report. Second, we extrapolate our DiD estimates to episode initiators not included in our sample due to limitations identifying suitable matched comparison providers. Third, the reconciliation payments that we used do not account for several model adjustments that are applied at the episode initiator and convener level, aggregated across the clinical episodes in which they participate (i.e., the stop-loss/stop-gain provision, the composite quality score adjustment, and the post-episode spending penalty amount). Finally, we estimated net savings to Medicare for the 17 hospital clinical episodes and 17 PGP clinical episodes evaluated. These clinical episodes represent 92.1% of all BPCI Advanced episodes.

We have begun to assess health equity in the BPCI Advanced Model for the first time in this evaluation report. For the analysis of historically underserved populations, we utilized the comparison groups selected for the main analysis. We compared outcomes for subgroups of beneficiaries from historically underserved populations relative to their counterparts in the

comparison group, and we compared the relative change for each historically underserved population to a reference population. This methodology relies on the assumption that our treatment and comparison groups are balanced on key characteristics for beneficiaries from each underserved population and each reference population. However, since we constructed the comparison groups such that the treatment and comparison groups would be balanced across the full sample of beneficiaries, we cannot guarantee that the subgroups are also balanced. In addition, as in the analysis of the full sample of episodes, the methodology used to analyze historically underserved populations relies on the parallel trends assumption, that is, that outcomes for BPCI Advanced and comparison episodes within each subpopulation had parallel trends during the baseline period. This implies that in instances where BPCI Advanced participants were moving in a favorable direction during the baseline, estimated coefficients may overstate any favorable changes under the model.

We identify beneficiary race or ethnicity using the Research Triangle Institute (RTI) race code, which has known limitations. Race and ethnicity are treated as if they are mutually exclusive, and therefore, do not allow for overlap in categories. The RTI race code also has low validity for certain groups when compared to self-reported data; however, many studies find high validity for the Black or African American race code.[48] We used the RTI race code in this analysis because alternative data sources covering both the baseline and the intervention period were not available. In 2021, CMS made available new data developed by RAND to improve the measurement of race and ethnicity. However, these data are not available for the BPCI Advanced evaluation baseline period, April 2013 through December 2017. We will continue to explore improved measures of race and ethnicity that can be used for evaluation purposes.

Beginning January 1, 2020, CMS allowed Medicare coverage of total hip arthroplasty (THA) to be performed in the outpatient setting. During Model Year 3 (2020), the BPCI Advanced Model only included THAs performed in the inpatient setting, which may have provided incentives for BPCI Advanced participants to perform more THAs in the inpatient setting than they would have if the BPCI Advanced Model had not occurred. If this were the case, it would have led to an overestimate of payment reductions under the model. However, due to the COVID-19 PHE, providers (both in BPCI Advanced and those not in the model) may have been more likely to shift THAs to the outpatient setting than they would have otherwise. This may have overcome any potential incentives to shift toward the inpatient setting.

The analyses presented in this report cover 2020 and 2021, two years that were heavily affected by the COVID-19 PHE. We include controls for COVID-19 in the regressions that estimate the results presented in this report, and we verified that BPCI Advanced and the comparison group had similar rates of episodes with a COVID-19 diagnoses and similar rates of COVID-19 in the county. However, the COVID-19 PHE had widespread effects on the entire health system that may not be adequately captured in data. In site visits and interviews, participants reported that patients may have delayed care or wanted to avoid engaging with PAC services, and they noted difficulties with staffing, burnout, and labor shortages. On the other hand, they reported increased use of telehealth

[48] For example, see Jarrín, O. F., Nyandege, A. N., Grafova, I. B.; Dong, X., Lin, H. (2020). Validity of Race and Ethnicity Codes in Medicare Administrative Data Compared With Gold-standard Self-reported Race Collected During Routine Home Health Care Visits. *Medical Care*, 58(1), e1–e8.

and remote patient monitoring. Thus, we cannot determine for certain whether some of the results presented are due to the BPCI Advanced Model or the COVID-19 PHE.

C. Conclusion

In Model Year 3 (2020), the BPCI Advanced Model successfully reduced total episode payments and improved unplanned readmission and mortality rates for PGP medical clinical episodes. Findings for patient-reported functional status, care experiences, and satisfaction with care from Model Year 4 (2021) were generally slightly less favorable or neutral for BPCI Advanced relative to comparison respondents. While there were no changes in mortality rates for the historically underserved populations analyzed relative to their comparison groups, there is evidence that mortality rates declined for some reference populations in Model Year 3. These results suggest that Black or African American beneficiaries and dual-eligible beneficiaries were less likely than their reference populations to benefit from the BPCI Advanced Model during Model Year 3. Findings for patient-reported functional status from Model Year 4 were mixed or neutral for the historically underserved populations analyzed, with the exception that dual-eligible BPCI Advanced respondents with hospital-initiated episodes reported less favorable changes than comparison respondents. Findings for patient-reported care experiences and satisfaction with care from Model Year 4 were generally less favorable or neutral for the historically underserved populations analyzed. Future reports will continue to evaluate payment, utilization, and quality outcomes under BPCI Advanced for all BPCI Advanced beneficiaries, as well as for beneficiaries from historically underserved populations.

In Model Year 3, the model accrued net Medicare losses for medical clinical episodes and net Medicare savings for surgical clinical episodes, resulting in small (less than 1%) net Medicare losses overall. Changes to the target pricing methodology and clinical episode groupings that CMS implemented in Model Year 4 are intended to bolster the model's ability to achieve Medicare savings. CMS implemented additional changes to the target pricing methodology in Model Year 6 (2023), which are intended to further improve the pricing methodology and keep providers and suppliers engaged in value-based care. Achieving savings is important because the Secretary of Health and Human Services has the authority to expand models that reduce federal spending while maintaining or improving quality for beneficiaries. Future evaluation reports will assess how these changes impacted participation in the model and Medicare program savings.