The Hilltop Institute
UMBC

report

Maryland Primary Care Program Performance Evaluation (2019–2022)

August 20, 2024

Suggested Citation: Goetschius, L., Gill, C., Fakeye, O., Han, F., & Henderson, M. (2024, August 20). *Maryland Primary Care Program performance evaluation (2019–2022)*. Baltimore, MD: The Hilltop Institute, UMBC.

Acknowledgements

Hilltop would like to thank Chad Perman, Jason Dubois, Emily Gruber, Kayleigh Hope, William Henderson, and Christa Speicher for their helpful feedback.

Additionally, Hilltop developed this evaluation through a contract from the Maryland Primary Care Program (MDPCP) with funding from the Centers for Disease Control and Prevention's National Initiative to Address COVID-19 Health Disparities Among Populations at High-Risk and Underserved, including Racial and Ethnic Minority Populations and Rural Communities.

Hilltop currently maintains an interagency agreement with MDPCP and has created and provided tools and reports for measuring and stratifying patient risk. These instruments have been available for use by MDPCP providers since October 2019. Hilltop was awarded the contract to perform this evaluation through a competitive procurement process. The evaluation team and work were separate from Hilltop's work on the MDPCP risk tools. The Maryland Department of Health did not find any conflict of interest with this study.

Table of Contents

List of Tables and Figures

Tables

Figures

Executive Summary

The Maryland Primary Care Program (MDPCP) began in January 2019 as a key component of the state of Maryland's Total Cost of Care (TCOC) Model, an agreement between the Centers for Medicare & Medicaid Services (CMS) and the State to limit statewide growth in Medicare per capita expenditures while advancing higher quality of care. As one of three primary interventions of the TCOC—which in turn builds on the Maryland All-Payer Model—MDPCP is a voluntary program that provides funding and support for the delivery of advanced primary care throughout the state. It allows primary care providers to play a greater role in the prevention and management of chronic disease, as well as in the prevention of unnecessary hospital utilization, with the ultimate goal of improving quality of care while reducing Medicare expenditure growth trends.

This report presents an evaluation of the estimated causal impact of implementation of MDPCP on utilization and expenditure for Medicare beneficiaries and is the State's commissioned independent evaluation of the program to fulfill the 2024 Joint Chairmen's Report (JCR) requirement. JCRs are legislatively mandated reports on topics of relevance to the State, and this JCR requested that the evaluation should outline cost savings from MDPCP reducing unnecessary utilization or hospitalization for patients participating in MDPCP over the increased expenditures from provider incentives. Using a combination of propensity score matching and difference-in-differences analyses, Hilltop identified the effect of MDPCP by examining changes in outcomes for individuals attributed to MDPCP-participating primary care practices, net of concurrent changes in outcomes for comparable individuals in Maryland over the first four years of the program implementation period (2019–2022). Given that MDPCP is a component of a suite of TCOC policy initiatives, these results can be interpreted as occurring over and above the other components of the TCOC Model. In addition to estimating overall impacts on total expenditure, inpatient utilization, emergency department (ED) utilization, and avoidable hospital events, we examined effects for various subgroups of beneficiaries, stratifying by race/ethnicity, dual eligibility for Medicare and Medicaid, area characteristics, and practice characteristics. Last, we evaluated whether the COVID-19 pandemic modified the effect of MDPCP on these utilization or spending outcomes.

The Hilltop team found that, on average, MDPCP led to a reduction of 4.33% in total spending (as measured by total Part A and Part B Medicare payments) and reduction of 7.18% in the probability of any inpatient utilization per beneficiary per calendar year quarter. Annualized, the impact on total spending translates to a reduction of $424.68 per beneficiary per calendar year when accounting for the fact that not all MDPCP-attributed beneficiaries were continuously attributed for all four quarters of a year. The team documented smaller reductions in the probability of any ED utilization and the incidence of avoidable hospital events: 1.70% and 1.96%, respectively, with the latter effect statistically insignificant.

Table ES.1, below, displays the results.

Table ES.1. Summary of Estimated Impacts of MDPCP on Expenditure and Utilization Outcomes for Medicare Beneficiaries in Maryland, 2019–2022

Outcome	Quarterly Effect Estimate *% Change**	95% Confidence Interval *% Change**	
		Lower	Upper
Total Medicare Parts A & B expenditure	**-$119.60** ***-4.33%***	**-$151.66** ***-5.49%***	**-$87.54** ***-3.17%***
Any inpatient utilization	**-0.0030** ***-7.18%***	**-0.0038** ***-9.10%***	**-0.0021** ***-5.03%***
Any ED utilization	**-0.0013** ***-1.70%***	**-0.0025** ***-3.27%***	**-0.0002** ***-0.03%***
Any avoidable hospital event	-0.0003 *-1.96%*	-0.0008 *-5.22%*	0.0003 *1.96%*

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter of participation, after accounting for changes among matched comparators and controlling for individual-level fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level. *Defined as the percentage change in the average value of the outcome from the calendar quarter before a beneficiary's participation in MDPCP began.

Hilltop conducted subgroup analyses to assess the extent to which these outcomes varied across distinct patient and practice characteristics. We documented larger spending reductions for Black patients relative to White patients, for dually eligible individuals relative to non-duals, and for individuals who had ever resided in an area with high levels of area deprivation relative to individuals who had not resided in an area with high levels of area deprivation. While these estimates are not statistically differentiable across groups, they suggest that MDPCP is impacting traditionally underserved groups and, in so doing, advancing health equity in Maryland. We did find evidence that the MDPCP impact estimates on total spending and inpatient utilization were moderated during the COVID-19 pandemic, meaning that MDPCP led to a smaller reduction in expenditure and utilization during the pandemic relative to before the pandemic.

Finally, Hilltop compared the aggregate Medicare expenditure savings estimated for the MDPCP participants to the program's incurred operational costs for payments to participating providers from 2019 to 2022. The analysis indicates that the total savings in Medicare expenditure were sufficient to fully cover program costs and may have led to overall net cost savings.

Table ES.2, below, presents the results.

Table ES.2. Estimated Medicare Expenditure Savings and Operational Costs during the First Four Years of Implementation of the MDPCP, 2019–2022

Outcome	Aggregate Effect	95% Confidence Interval	
		Lower	Upper
Expenditure	-$672.4 million	-$852.7 million	-$492.2 million
Program Cost	$510.5 million	$510.5 million	$510.5 million
Net Program Cost	**-$161.9 million**	**-$342.2 million**	**$18.3 million**

Notes: Net program cost is calculated as the difference between the total aggregate savings on Medicare Parts A & B expenditure among participating beneficiaries, and the total of the program's payments (net of recoupments) to participating providers in primary care practices and care transformation organizations for per beneficiary per month care delivery fees and incentives over the four-year period. The aggregate program estimate is calculated as the overall impact estimate multiplied by the total number of person-quarters in the treatment period (N = 5,622,377). A positive value indicates that program operational costs exceeded estimated savings, while a negative value indicates that estimated savings on Medicare expenditure were greater than the program's payments to participating providers.

This evaluation has limitations that should be considered when interpreting the results. First, any evaluation based on observational data can potentially reflect the impact of confounding factors. This evaluation employed matching techniques and a difference-in-differences regression methodology to mitigate potential confounding. Additionally, we conducted several specification checks to support the validity of our comparisons and believe that our methodology mitigates the risk that unobserved confounders introduce bias into our impact estimates.

Second, the COVID-19 pandemic occurred six quarters into the study period. While the difference-in-differences technique should account for shocks that were common to both the treatment and comparison groups, we acknowledge that MDPCP practices may have responded to the pandemic in systematic ways that practices for the comparison group did not. We assessed the extent to which the MDPCP impact estimates changed during the COVID-19 pandemic, relative to the period prior to the pandemic, and found that the MDPCP (negative) impacts on spending and utilization both moderated during COVID-19: the effects were still negative but grew smaller in magnitude relative to what they had been prior to the pandemic. While we are unable to discern the mechanisms driving this result, this could have been caused by MDPCP practices conducting outreach to maintain their beneficiaries' link to the health care system during this global event.

Finally, since MDPCP is a voluntary program for practices, we cannot assume that the positive effects of MDPCP would be similar if the program was rolled-out to the approximately 300 practices that have not yet chosen to participate. Hilltop documented heterogeneous effects by joining cohort, with the earliest joiners in the first quarter of 2019 experiencing larger effects than individuals whose exposure to MDPCP began in the first quarter of 2020 or 2021. Thus, while we found that the first four years of MDPCP were approximately revenue-neutral and may have led to net cost savings, we cannot conclude that a full-state roll-out would yield similar results.

The main results from this evaluation differ from a recent federally funded analysis of MDPCP, situated within a broader evaluation of the TCOC (Peterson et al., 2024a). We believe that there are three reasons that prevent direct comparison of these results. First, the current evaluation conducts an analysis spanning all participating practices and beneficiaries from quarter 1 of 2019 to quarter 4 of 2022, rather than only for practices that joined the program in 2019. Second, the current evaluation uses CMS's attribution methodology for assigning Medicare beneficiaries to MDPCP-participating and non-MDPCP providers, instead of recreating the attribution methodology in the data set used for evaluation. Third and most importantly, this evaluation identifies the impact of MDPCP using within-individual changes in outcomes over time, rather than regression-adjusted within-practice changes in average outcomes over time. This difference—that is, building impact estimates from the individual upward rather than from the practice downward—centers the analysis on the beneficiary-level response to MDPCP. Both approaches are methodologically valid but seek answers to slightly different questions.

Maryland Primary Care Program Performance Evaluation, 2019–2022

Introduction

The Maryland Primary Care Program (MDPCP) began in January 2019 as a key element of the state's Total Cost of Care (TCOC) Model, an agreement between CMS and the state of Maryland. MDPCP is a voluntary program that provides funding and support for the delivery of advanced primary care throughout the state. It allows primary care providers to play an increased role in the prevention and management of chronic disease, as well as in the prevention of unnecessary hospital utilization, with the ultimate goal of improving quality of care while reducing Medicare TCOC growth.

This report documents the causal impact of the introduction of MDPCP on utilization and expenditure and is the State's commissioned evaluation of the program to fulfill the 2024 Joint Chairmen's Report (JCR) requirement of an independent evaluation of MDPCP. JCRs are legislatively mandated reports on topics of relevance to the State, and this JCR requested that the evaluation should outline cost savings from MDPCP reducing unnecessary utilization or hospitalization for patients participating in MDPCP over the increased expenditures from provider incentives (Chairmen of the Senate Budget and Taxation Committee and House Appropriations Committee, 2024). Using a combination of propensity score matching and difference-in-differences analyses, the Hilltop team identified the effect of MDPCP by examining changes in outcomes for Traditional Medicare-insured (that is, fee-for-service) beneficiaries attributed to MDPCP-participating practices, net of changes in outcomes for comparable beneficiaries in Maryland over the same time period. In addition to estimating overall impacts on total expenditure, inpatient utilization, emergency department (ED) utilization, and avoidable hospital (AH) events, we examined effects for various subgroups stratified by race, dual eligibility for Medicare and Medicaid, area characteristics, and practice characteristics. Additionally, we conducted an analysis to discern the extent to which the impact estimates of MDPCP changed during the COVID-19 pandemic.

Section 1 of this report discusses structural details of MDPCP; Section 2 documents what is known about the impact of MDPCP and other similarly enhanced primary care programs; Section 3 presents the data used in this evaluation; Section 4 describes the analytic methodology; Section 5 presents results; Section 6 discusses limitations; and Section 7 discusses the results (and, in particular, situates the findings presented in this study with those from other recent evaluations of the program). An Appendix includes additional results and methodological and technical detail.

Section 1. Institutional Details of the Maryland Primary Care Program

Overview

MDPCP is a programmatic element of Maryland's TCOC Model, which is designed to operate from January 1, 2019, through December 31, 2026 (Centers for Medicare & Medicaid Services,

2024a). MDPCP was created in partnership between the Center for Medicare and Medicaid Innovation (CMMI) and the Maryland Department of Health (MDH) to support the TCOC's goal of transforming the delivery of health care across the state through investment in a robust and enhanced primary care system. It is one of four aligned programs constituting the TCOC Model; the other three are the Hospital Payment Program, the Care Redesign Program, and Population Health Improvement Outcomes-Based Credits.

MDPCP is intended to create a network of strong and effective advanced primary care practices throughout the state. The program involves a transformation of the primary care delivery system through targeted and data-driven care management of complex patients and chronic disease, the integration of behavioral health, screening and referrals for unmet social needs, and data-driven quality improvement (Neall et al., 2019).

The main aims of MDPCP are to reduce avoidable utilization of acute health care services, reduce aggregate health services expenditure, and improve quality outcomes for all residents of the state (Schrader et al., 2021). The program seeks to build a strong, effective, and sustainable primary care system characterized by continuous, relationship-based care that is responsive to medical concerns, behavioral health needs, and social determinants of health (Neall et al., 2019). These population health goals are intended to be achieved by leveraging the elements of advanced primary care, including strategic financial investments of population-based payments in primary care infrastructure, data tools for monitoring performance, and the dissemination and implementation of quality improvement processes. The envisioned statewide health care transformation will result in better prevention and management of disease conditions, yielding reductions in AH and ED utilization that ultimately curb the rising trend of high and unsustainable health care expenditure (Neall et al., 2019).

MDPCP is a multi-payer program and began operations with fee-for-service (FFS) Medicare in January 2019. A commercial payer, CareFirst, joined in 2020 as an aligned payer adopting a value-based payment model with similar goals and payment structure to MDPCP (Neall et al., 2019). This evaluation; however, focuses on the impact of the introduction of MDPCP on outcomes for FFS Medicare beneficiaries.

Participation

To be eligible to participate in MDPCP, a primary care practice must be located within the state of Maryland, enrolled in the Medicare program with good standing, and maintain a minimum of 125 attributed FFS Medicare beneficiaries during each program year (PY). Attribution is based on beneficiaries' utilization over a 24-month lookback period. Providers within the practice are eligible if they provide primary care services and have a primary care specialty (Center for Medicare & Medicaid Innovation, 2019). Federally qualified health centers (FQHCs) were initially excluded from participation but became eligible in PY 2021 (Center for Medicare & Medicaid Innovation, 2023).

Eligible practices join the program during annual open enrollment periods for the first five implementation years (Neall et al., 2019). Practices join at the standard (Track 1) or advanced (Track 2) level, with specific care delivery, care quality, and administrative requirements for each Track. Track 1 is a temporary level for practices working to meet the requirements of advanced primary care. Track 1 practices must proceed to Track 2 no later than the end of their third year of participation. Practices in Track 2 have met advanced primary care requirements and are willing to accept a modified payment structure while enrolled. Track 2 practices receive greater financial support and a partial prepayment for regular evaluation and management services. All participating practices have the option to partner with a care transformation organization (CTO) to assist with staffing, care coordination, and technical assistance needs relevant to meeting program requirements.

Requirements

In addition to providing comprehensive primary care services, practices participating in MDPCP are required to deliver and improve on five key functions of advanced primary care: access and continuity; comprehensiveness and coordination; beneficiary and caregiver experience; planned care for health outcomes; and care management. Participating practices are also required to track and report on specific electronic Clinical Quality Measures (eCQMs) of clinical performance (Neall et al., 2019).

Payments & Support

In addition to their regularly billed Medicare FFS payments, primary care practices participating in MDPCP are eligible to receive the following four forms of enhanced and prospective Medicare payment to support transformation and implementation of advanced primary care.

- Care Management Fees (CMF): For implementing changes and services, participating practices in both Track 1 and Track 2 receive a prospective, non-visit-based, risk-adjusted per beneficiary per month (PBPM) payment made prospectively on a quarterly basis. Payment tiers are based on disease burden as assessed using the Hierarchical Conditions Category (HCC) risk adjustment model. Practices partnering with a CTO may share 30% or 50% of the CMF depending on the CTO's level of support to the practice.
- Performance-Based Incentive Payment (PBIP): The PBIP is an annual prospective payment designed to incentivize and reward practice-level performance on clinical quality, beneficiary experience of care, and reductions in acute utilization. Practices and CTOs receive separate PBIPs. Although practices receive the full payment in advance, they retain only the portion earned based on achievement levels relative to defined thresholds on two component indices:
    - *Quality Component (50%)*: This component includes patient experience of care (15%) and four clinical outcomes—hemoglobin A1c control in patients with diabetes (8.75%), control of high blood pressure (8.75%), screening and management of body mass index (8.75%), and screening and management of depression (8.75%).

  - *Utilization Component (50%)*: This component assesses ED utilization (16.7%) and acute hospital utilization (33.3%).

- Comprehensive Primary Care Payment (CPCP): A feature of the MDPCP's hybrid payment methodology, the CPCP is a partially capitated payment made to advanced (Track 2) practices and FQHCs based on primary care service use by attributed beneficiaries. A portion of the participating practice's expected FFS reimbursement for select primary care services (SPCS) delivered to attributed beneficiaries is paid prospectively each quarter, while the remainder is distributed as these SPCS claims are billed and paid, using a reduced FFS (or prospective payment system for FQHCs) fee schedule that has been proportionately adjusted to deduct the CPCP disbursed in advance. Reimbursed amounts for SPCS claims for beneficiaries attributed to Track 2 practices reflect this reduced fee schedule in Medicare Part B claims. Prior to the start of each year, Track 2 practices select the percentage of their expected SPCS claims they would like to receive as CPCP (ranging from 25%-65% in the first year or 40%-65% in subsequent years), and CMS calculates the CPCP as 110% of the standard reimbursement rate for this portion of SPCS claims. The surplus is intended to compensate the providers for additional care coordination and delivery requirements of advanced primary care. Participating practices are permitted to retain the entire CPCP payment, net of any recoupments, as an incentive to drive timely, efficient, and innovative care processes (The Lewin Group, 2024).
- Health Equity Advancement Resource and Transformation (HEART) Payment: This provides additional support to participating practices that serve socioeconomically disadvantaged populations toward advancing health equity. Starting in PY 2022, these practices receive an additional $110 PBPM for attributed beneficiaries in the highest quartile of HCC risk and residing in locations that score in the highest quintile of the area deprivation index (ADI). HEART payments can be used by practices to procure chronic disease care kits, provide housing support or food assistance, or engage community health workers (Maryland Department of Health, 2023a). CTOs also receive a share of the HEART payments.

During the first four years of MDPCP implementation, over $510.5 million in net payments was distributed to participating practices and partnering CTOs as a combination of these payment types (Table 1). An additional population-based practice track (Track 3) was introduced in January 2023 after the project period of this evaluation. In place of CMFs, PBIPs, and CPCPs, Track 3 practices and partner CTOs receive a total primary care payment (TPCP) that consists of a portion of FFS billings paid prospectively (termed population-based payments), and a flat visit fee paid at the time of service for specified primary care services. The TPCP is adjusted for the practice's performance on measures of quality, utilization, and total cost of care (Centers for Medicare & Medicaid Services, n.d.-a).

Table 1. Annual Net MDPCP Payments to Participating Primary Care Practices and CTOs

Payment Type	Year 1 (2019)	Year 2 (2020)	Year 3 (2021)	Year 4 (2022)	Years 1-4 (2019-2022)
CMF payments	$53,003,874	$102,611,859	$132,155,845	$110,293,405	$398,064,984
PBIP payments	$5,717,430	$7,444,697	$15,581,051	$10,002,708	$38,745,886
CPCP payments	$701,367	$4,673,400	$14,445,164	$26,192,851	$46,012,782
HEART payments	N/A	N/A	N/A	$27,692,401	$27,692,401
Annual Total	**$59,422,672**	**$114,729,956**	**$162,182,060**	**$174,181,365**	**$510,516,053**

Notes: N/A – not applicable. The spending totals in this table reflect net payments (which differ from gross payments) to participating practices and CTOs across both Track 1 and Track 2 of MDPCP. Net payments are calculated as gross payments less retrospective recoupments/debits for overpayment due to ineligible beneficiaries, excessive beneficiary out-of-practice utilization for office visits, provider termination or withdrawal, or redundant/unearned payments for the same services. As such, gross payments tend to exceed net payments.
Source: MDPCP net payment data received from MDPCP Program Management Office (PMO).

Participating practices receive support and technical assistance from CMMI and the MDH Program Management Office (PMO). Practices are provided with a team of dedicated coaches, on-demand data services of the state-designated health information exchange Chesapeake Regional Information System for our Patients (CRISP), population health analytic services from The Hilltop Institute at UMBC, and contractors helping with additional patient-related needs (Neall et al., 2019; Schrader et al., 2021).

Program Reach & Uptake

During its first three years of implementation, MDPCP recorded an annual increase in the counts of participating practices, providers, and beneficiaries (Table 2). By the third year of implementation (2021), MDPCP participation had increased to 525 primary care practices across Maryland, with about 2,150 primary care providers serving a diverse population of about 377,000 attributed beneficiaries. In addition to being in all 24 Maryland counties (including Baltimore City), MDPCP practices are also located in primary care health professional shortage areas (HPSAs), mental health HPSAs, medically underserved areas, and rural areas.

Table 2. Annual Enrollment of Primary Care Practices, Providers, and Beneficiaries in MDPCP

Characteristic	Implementation Year 1 (PY 2019)	Implementation Year 2 (PY 2020)	Implementation Year 3 (PY 2021)	Implementation Year 4 (PY 2022)
Number of participating primary care practices*	380	476	525	508
Number of participating primary care providers*	1,500	2,000	2,150	2,150
Number of attributed FFS Medicare beneficiaries	214,640	325,770	376,785	365,211

Notes: *Approximate numbers. For PY 2022, the number of participating practices provided is from the calendar year quarter with the highest enrollment. PY – program year.
Sources: Schrader et al. (2021); Schrader et al. (2022).

COVID-19 Impact on MDPCP Operations

The integration of public health initiatives with primary care within the MDPCP framework is designed to foster rapid response to public health emergencies (Schrader et al., 2021). The effectiveness of this programmatic feature was demonstrated in the coordinated approach by participating practices to provide continued telehealth access for COVID-19 and non-COVID-19 care, a steady supply of personal protective equipment, point-of-care viral or polymerase chain reaction (PCR) testing, therapeutics such as oral antivirals, and vaccination to their patients during the COVID-19 pandemic (Schrader et al., 2022). At the onset of the COVID-19 outbreak in 2020, the program undertook several initiatives to support practices and CTOs in addressing pandemic-related concerns and effectively continue to deliver advanced primary care to beneficiaries. These efforts included a webinar series to keep participating practices updated on timely information on epidemiological developments and safe practices; provision of daily clinical data to practices on hospital admission and ED visits to anticipate avoidable complications; supporting the provision of telehealth services catering to vulnerable patients and those needing accommodations to be seen in person; and the option to enroll in ImmuNet for distribution of COVID-19 vaccines as available (Schrader et al., 2021).

Participating practices reported that the continued support provided by MDPCP, as well as the financial flexibility afforded by the advanced payments, allowed their providers to retain staff and maintain open access to their services despite a decline in in-person visits (Schrader et al., 2022). The effectiveness of these strategies was demonstrated in a lower likelihood of COVID-19 incidence, hospitalizations, and mortality among patients of MDPCP practices relative to counterparts in non-MDPCP practices (Gruber et al., 2023).

Section 2. Review of MDPCP-Related Literature

Several analyses of (or related to) MDPCP have been conducted in recent years. Below is an overview of the relevant results.

MDPCP Annual Reports

The MDH PMO publishes annual reports describing the background of MDPCP, along with a status of implementation that includes a description of participating practices, summary of payments made to participating providers to support the delivery of advanced primary care, practice performance metrics, alignment of program activities with the Statewide Integrated Health Improvement Strategy (SIHIS), and future directions and recommendations to CMS.[1]

The January 2023 MDPCP Summary Report includes data on MDPCP performance across the first four program years (2019, 2020, 2021, and January-September of 2022) (Maryland Department of Health, 2023b). The report compares overall outcomes for Medicare beneficiaries attributed to MDPCP practices to a sample of non-participating comparators that represents "a subset of the statewide non-participating [Medicare beneficiary] population, demographically matched to participants by age band, sex, dual eligibility for Medicaid, and county of residence." The report examines trends in inpatient use, ED utilization, AH events for diagnoses similar to Prevention Quality Indicators (PQIs), and Medicare PBPM expenditure risk-adjusted for HCC score. Inpatient utilization among MDPCP beneficiaries fell from 15.2% from 2019 to 2022, relative to a reduction of 12.3% among the comparison group of non-participants. For the entire statewide FFS Medicare population (which includes MDPCP participants and non-participants), inpatient utilization decreased by 12.2%.

Between 2019 and 2022, utilization of ED services dropped by 17.9% among MDPCP participants and by 17.3% among the non-participating population. The PQI-like AH event rate fell 29.2% for the MDPCP population and 29.2% for the equivalent non-participating population. Finally, spending rose by 11.7% for the MDPCP population and 16.0% for the equivalent non-participation population, for a relative decrease of 4.3 percentage points.

Annual Legislatively Mandated Reports

Beginning in 2020, the Maryland Health Services Cost Review Commission (HSCRC) has conducted an annual evaluation of MDPCP and released a JCR to the Maryland state legislature. For the latest available report, an assessment of the program's performance during PY 2023, "as directed by the committees, HSCRC analyzed the impact of MDPCP on both the Medicare total cost of care and inpatient utilization" (Maryland Health Services Cost Review Commission, 2023, p. 14).

[1] Annual reports are linked under the "General Information" section on the MDPCP main page: https://health.maryland.gov/mdpcp/Pages/home.aspx

The study's methodology attributes Medicare beneficiaries to MDPCP and non-MDPCP primary care practices and uses a difference-in-differences analytic methodology to assess changes in outcomes for the MDPCP attributed population relative to changes in outcomes for a comparable non-participating Medicare population. This is similar to the approach taken in this analysis: as documented below, the methodology for this evaluation uses an in-state comparison group (i.e., within Maryland) and difference-in-differences methodology.

HSCRC's analysis of the first four years of the program (2019-2022) found that, while MDPCP has made progress integrating advanced primary care into the health care system and has shown reductions in inpatient health care utilization, the program was associated with a net increase in Medicare expenditure during PY 2022 relative to 2018 (prior to the start of the program) when factoring in the operational costs of payments to the primary care practices participating in the program (Maryland Health Services Cost Review Commission, 2023). Specifically, the 2023 report estimated that the MDPCP population experienced a 2.66% reduction in inpatient utilization and $114.1 million in aggregate Medicare TCOC reductions between 2018 and 2022 relative to the comparison population. The authors noted that the program's operational costs for PY 2022, which totaled $198.6 million, exceeded the estimated savings, resulting in a net deficit of $84.5 million.

The HSCRC's annual JCRs document variation in MDPCP's estimated impact from year to year. For example, there were larger reductions in Medicare expenditure and inpatient hospital utilization during PY 2020 among MDPCP-participating practices compared to non-participating practices. The reports do not provide 95% confidence intervals for assessing the statistical significance of effect estimates. Hence, it is not possible to evaluate the statistical precision of these program estimates. Furthermore, the evaluation methodology does not analyze the effect of MDPCP on non-hospital Medicare expenditure, nor ED utilization or incidence of AH events.

Evaluation of the TCOC Model

In April 2024, contractors on behalf of CMMI released an evaluation report of the TCOC Model (Peterson et al., 2024a). Using a cross-state methodology comparing Maryland to approximately one quarter of the rest of the country from 2011-2022, the evaluation finds that the TCOC model reduced total Medicare spending by 2.1%, with 1.0% of this reduction in spending attributable to the TCOC Model beginning in 2019. This reduction is concentrated in hospital spending, with a 1.4 percentage point reduction from 2019-2022 relative to the 2017-2018 period, with a smaller but statistically insignificant reduction in non-hospital spending.

Additionally, the evaluation found that the TCOC reduced all-cause acute care hospital admissions by 16.2% relative to baseline, with a 5.6 percentage point differential reduction starting with the introduction of the TCOC in 2019. The model also resulted in a 3.1 percentage point reduction in outpatient ED visits and observation stays over the 2019-2022 period relative to 2017-2018. Table 3 presents selected key findings.

Table 3. Selected Key Findings from the TCOC Progress Report

Outcome	2017-2018 Estimated Impact for MDAPM	2019-2022 Impact from MDAPM and TCOC	Difference Due to TCOC
Total Medicare Part A & B FFS spending	-1.1%*	-2.1%**	-1.0 pp**
Hospital spending (inpatient and outpatient)	-4.7%**	-6.1%**	1.4 pp **
Non-hospital spending	3.6% **	3.1% **	-0.5 pp
All-cause acute care hospital admissions	-10.6% **	-16.2% **	-5.6 pp **
Outpatient ED visits and observation stays	-2.8% **	-5.9% **	-3.1 pp**

Notes: * p<0.10; ** p<0.05. MDAPM – Maryland All-Payer Model.
Source: Exhibit ES.3. in the TCOC Progress Report, Peterson et al., 2024

It is important to note that the TCOC Model is a *suite* of policy initiatives that incorporated aspects of the prior statewide all-payer policies (e.g., hospital global budgets) and new initiatives (e.g., MDPCP). As of 2022, MDPCP was the second largest component of the TCOC, with an investment of $195 million. Hospital global budgets were the largest investment, at over $1 billion (Peterson et al., 2024a, p. 20). Given that hospital global budgets are a continuation of prior policy in Maryland, while MDPCP was a new policy initiative effective as of 2019, it is possible that the introduction of MDPCP drove a portion of this observed reduction in spending and inpatient utilization from 2019-2022.

MDPCP-Specific Analysis

As noted above, in April 2024, contractors on behalf of CMMI released an evaluation report of the TCOC Model, which included a sub-analysis on MDPCP. This analysis used a practice-level approach, restricted to practices that joined MDPCP in 2019, and used a within-Maryland comparison group to estimate the impact of MDPCP on spending and utilization. The authors found that MDPCP did not reduce outpatient emergency department use, readmissions, or total spending during 2019-2022 above and beyond the other TCOC Model components. However, they did find some evidence for MDPCP having "possibly reduced admissions from 2019 to 2022 by 2.5% (P = 0.10)" (Peterson et al., 2024a, p. 118). Additionally, the study found that MDPCP led to "modest improvements on timely follow-up after exacerbation of a chronic condition" from 2019-2022 (1.9%, P = 0.04). The study concluded that, for beneficiaries attributed to 2019 practices, the program led to approximately $6 million of annual savings (not statistically significant); but, given annual program costs of $96 million, led to a net loss of approximately $90 million annually.

As with the HSCRC evaluations, the methodology employed by this study is broadly consistent with that used in the current evaluation: the authors created a matched treatment-comparison group using practices within Maryland, and then employed a difference-in-differences methodology to estimate the effect of MDPCP as the change in outcomes for treated units net of the change in outcomes for comparison units. However, the study's methodology deviates from the methodology used in this evaluation in several respects.

First, the evaluation focused only on practices that joined MDPCP in 2019. While this was the largest cohort of "joiners"—in our analytic sample, beneficiaries first attributed to MDPCP practices in 2019Q1 composed 38.3% of all beneficiaries attributed to participating practices from 2019 to 2022—this strategy, by definition, does not investigate effects for practices and beneficiaries that joined the program in subsequent years.

Second, the evaluation created its own attribution methodology using proprietary practice rosters purchased from a third-party vendor due to the limited information available for primary care practices in Maryland not participating in MDPCP (Peterson et al., 2024b). The authors note that their approach to attribution resulted in a 62% overlap in 2019 with the results of CMS's official attribution of MDPCP beneficiaries to participating practices for payment purposes. This degree of overlap is, however, lower than the approximately 75% overlap reported for the later years—with the increase likely due to changes in the payment attribution made beginning in 2020 (Peterson et al., 2024b, Exhibit E.6). This is in contrast to the current evaluation, which uses the full list of MDPCP-attributed beneficiaries from CMS for matching.

Third, the authors used the practice as the basis for matching. We acknowledge that MDPCP is a practice-level intervention, which argues for practice-level matching; however, for this evaluation, we did not have access to practice-level information for comparison practices, and as such, were unable to include practice-level characteristics in our estimation.[2] Moreover, the voluntary nature of MDPCP and the widespread uptake among primary care practices in Maryland as of 2019Q1 imply that it may be challenging to identify an adequate number of comparator practices through a matching algorithm (Peterson et al., 2024a, p. 87). The current evaluation, instead, implements a matching algorithm at the individual level, matching MDPCP-participating Medicare beneficiaries during their first quarter of attribution with comparable individuals who were at one point in the study period attributed to a non-participating primary care provider and who never participated in MDPCP.

Finally, the methodology used in the TCOC evaluation employs practice-level fixed effects in the regression analyses, whereas the methodology in the current evaluation uses individual-level fixed effects. That is, the TCOC evaluation controls for unobserved, time-invariant practice-level characteristics (such as whether a practice is located in a given area) and identifies the effect of MDPCP as the regression-adjusted within-practice change in average outcomes for individuals attributed to participating practices, net of the change for individuals attributed to non-participating practices. Crucially, this specification estimates a "combined" effect of both group-level and individual-level outcomes. If, for example, MDPCP practices retain high-cost beneficiaries at a higher rate than comparison practices, this functional form could conflate this

[2] Additionally, the landscape of primary care provision in Maryland potentially argues against a practice-level matching strategy. MDPCP is a voluntary program in which practices opt to join; as such, practices may join based on their expected costs and benefits of doing so. Practices that are better poised to meet program requirements, for example, may be more likely to join than practices that, for example, lack electronic health systems or other attributes. Thus, it may be challenging to located non-participating practices that resemble participant practices.

retention—which may be reflected in higher practice-level average per-beneficiary costs—with an individual-level MDPCP effect.

In contrast, the methodology used in this evaluation uses variation at the *individual* level to identify the effect of MDPCP as the *within-person* change in outcomes for individuals ever attributed to MDPCP practices, net of the within-person change for individuals never attributed to MDPCP practices. Unlike a strategy that uses repeated cross-sections of individuals over time and estimates program effects as differences in averages across individuals over time, the empirical strategy used in this evaluation *follows* individuals over time. The only individuals that contribute to impact estimates in this evaluation are those with both a pre-treatment period and at least one post-treatment period; thus, while a practice-level estimate may reflect both individual effects and changing practice composition, the empirical strategy in this evaluation focuses on estimating individual-level effects.

Other Studies

Comprehensive Primary Care Plus Evaluation

The Comprehensive Primary Care Plus (CPC+) was a national advanced primary care medical home demonstration model that operated from January 2017 through December 2021. Similar to MDPCP, CPC+ aimed to improve "quality, access, and efficiency of primary care" (Centers for Medicare & Medicaid Services, n.d.-b). Participating practices were drawn from 18 regions throughout the country (although none in Maryland), and the model shared several key operational details with MDPCP: practices joined one of two tracks, with Track 2 practices being required to "provide more enhanced care delivery approaches to better support patients with complex needs and received additional financial support to help them" (O'Malley et al., 2023, p. 2). Additionally, like MDPCP, CPC+ practices received financial support, data feedback, and individualized and group learning supports in order to facilitate the delivery of advanced primary care.

An independent evaluation contracted by CMS found that CPC+ practices reduced annual ED visits by approximately 1-3 percentage points over the five program years, with reductions in acute hospitalizations emerging later in the program. Overall, despite significant reductions in acute medical hospitalization expenditures of approximately 3-4% across tracks, "CPC+ did not reduce total Medicare expenditures without enhanced payments and it increased expenditures with enhanced payments" (O'Malley et al., 2023, p. 144). The evaluation notes that practices participating in both CPC+ and Medicare Shared Savings Program (SSP) experienced more favorable effects than those not participating in SSP.

MDPCP and COVID-19

The COVID-19 public health emergency occurred one year into MDPCP. In response, the program "continued to expand comprehensive, advanced primary care across the state, while addressing the COVID-19 pandemic through broad implementation of telehealth, testing, therapeutics, and vaccination" (Maryland Health Services Cost Review Commission, 2021). A 2023 study examined

the extent to which attribution to an MDPCP-participating practice was associated with improved COVID-19-related outcomes, relative to a matched comparison group. Using Medicare claims from January 1, 2020 – October 31, 2021, and COVID-19 vaccination data from January 1, 2020 – March 31, 2022, the authors found that MDPCP-participating practices performed more favorably on most outcomes measured relative to comparison practices.

Specifically, Medicare beneficiaries attributed to MDPCP practices had higher vaccination rates (84.47% vs. 77.93%), were more likely to receive monoclonal antibody therapy if diagnosed with COVID-19 (8.45% vs 6.11%) and received more telehealth services than Medicare beneficiaries who were attributed to non-participating practices (62.95% vs. 54.53%). Additionally, the Medicare beneficiaries attributed to participating practices compared to those attributed to non-participating practices had lower rates of COVID-19 diagnosis (6.55% vs. 7.09%), COVID-19 inpatient admission (1.81% vs. 2.06%), and COVID-19-related death (0.56% vs. 0.77%) (Gruber et al., 2023). This study expands upon an earlier, similar study, which found that, based on Medicare claims from 2020, MDPCP beneficiaries had a lower incidence of COVID-19 diagnosis, COVID-19-related admissions, and COVID-19-related mortality (Perman et al., 2021).

While these studies indicate that MDPCP-participating practices performed well during COVID-19, the findings imply interpretational nuances for this evaluation of MDPCP. Specifically, to the extent that MDPCP practices responded to COVID-19 *differentially* relative to their non-MDPCP counterparts, individuals attributed to MDPCP practices may have experienced spending and utilization changes during this period that may be driven by MDPCP's response to COVID-19.

Section 3. Data

This section discusses data sources, outcomes, and covariates.

Data Sources

Hilltop used three primary data sources for this evaluation: primary care practice beneficiary attribution files provided by CMS, Medicare FFS claims from the Claims and Claim-Line Feed (CCLF), and CCLF beneficiary demographics files.

Attribution Files

Hilltop used MDPCP attribution files from 2019Q1–2022Q4 to delineate the "treatment" pool, which is further refined in the matching process. These files are generated by CMS quarterly and are a list of all individuals attributed to MDPCP-participating practices. The attribution methodology uses historical Medicare claims for primary care services to "determine the Participant Practice to which beneficiaries will be attributed" (Center for Medicare & Medicaid Innovation, 2019). Notably, this uses a two-year lookback period ending four months prior to the start of the quarter; thus, the attribution files for 2019 Q1 will be based entirely on pre-MDPCP

claims data.[3] Beneficiaries are attributed to providers for whom they are expected to receive the plurality of care in the next quarter.

In addition to MDPCP attribution files, we also used primary care attribution files for practices that are not currently participating in MDPCP. As with the MDPCP attribution files, these are generated quarterly by CMS. This forms the pool of individuals out of which we created the comparison group for the study. Crucially, this first filter on the comparison group should enhance the validity of the comparison group, since all individuals in the comparison group are attributed to a primary care practice at least once in the study period. Thus, our comparison between the treatment and comparison groups should implicitly control for a "primary care effect" by excluding Medicare beneficiaries who are not connected with a primary care provider.

CCLF Claims

For both covariates and outcomes, we primarily used Medicare FFS claims data in the CCLF extract provided for MDPCP monitoring and evaluation. These files consist of final-action claims covering Part A and Part B utilization and excludes claims related to Substance Abuse and Mental Health Services Administration's (SAMHSA) federal regulation 42 CFR Part 2.[4] While these files are created by CMS on a monthly basis and constitute a 36-month rolling window of claims, Hilltop aggregated available claims data into master claims data files ranging from April 2016–December 2022. Specifically, Hilltop developed a methodology to aggregate the CCLF claims into a "master" claims database spanning the project period for this analysis. This process ingests 55 CCLF data extracts and aggregates them into one claim database, correcting for any updated Medicare Beneficiary Identifiers (MBI) variables in the process. To accomplish this, Hilltop processed 14 terabytes of CCLF data, and the resulting master claims files are 1.7 terabytes.

The data volumes are significant. For example, this aggregation results in 77,299,053 Part A claims and 42,237,560 Part B claims from January 1, 2018–December 31, 2022. For additional details on the aggregation process, see the Appendix. Additionally, the Part B claims for SPCS delivered to attributed beneficiaries in MDPCP Track 2 practices reflect the reduced fee schedule applied due to the structure of the CPCP, in which a practice-selected portion of expected SPCS billings is paid to the practices prospectively at the start of each quarter. SPCS claims billed during the quarter are paid at a lower Medicare FFS reimbursement rate to account for the prospective CPCP.

CCLF Beneficiary Demographics Files

In addition to Parts A and B claims, each CCLF data file contains a beneficiary demographics file containing demographic information and summary information on eligibility, spending, utilization, and condition history for previous calendar years. Hilltop used information on ZIP

[3] Based on communications and internal documentation from the MDPCP Program Management Office.

[4] See https://www.ecfr.gov/current/title-42/chapter-I/subchapter-A/part-2

code, county, race, sex, age, reason for original Medicare entitlement, eligibility status, dual eligibility status, and selected conditions.

Outcomes and Covariates

We created both outcomes and covariates using CCLF claims files and the beneficiary demographics files. We created the outcomes for inpatient admissions, ED visits, AH events, and total Medicare spending (Parts A and B). We operationalized the utilization outcomes as 0/1 indicators for whether individual *i* incurred any utilization in quarter *q* of year *y*.

Hilltop opted for this binarized outcome structure for four reasons. First, we sought to avoid the undue influence of outliers in our impact estimates. While most individuals have no ED or inpatient utilization in a given quarter, a small number of individuals incur a significant amount of utilization and it is possible for these individuals to have a significant effect on the impact estimates. Second, using a 0/1 outcome structure, impact estimates can be interpreted as approximate probabilities: for example, an effect size of -0.005 translates to approximately 0.5 percentage points reduction in the probability of incurring that particular outcome. Third, this formulation is operationally expedient: as noted above, the underlying data constitutes 1.7 terabytes, and recreating visits from claims is computationally intensive. Finally, given the person-quarter granularity of the analytic data set, this structure will capture both the intensive and extensive margins for individuals with only one utilization event in a given time period.

We used 2022 Prevention Quality Indicator (PQI) measure definitions from the Agency for Healthcare Research and Quality (AHRQ) to operationalize the AH events outcome (Agency for Healthcare Research and Quality, 2022). Consistent with the definition of "avoidable hospital event" used in the MDPCP risk prediction tools, we identified AH events occurring in both the inpatient and ED settings. We winsorized spending at the 99th percentile (that is, replaced all values exceeding the 99th percentile with the 99th percentile) and replaced any negative values with $0.

We used fifteen covariates for matching. First, we matched exactly on the year-quarter of attribution, dual eligibility status during the quarter of attribution, first three digits of residential ZIP code (ZIP3) during the quarter of attribution, and race group (Black/White/other). We refined our matching algorithm using propensity score matching based on an additional eleven covariates: the average quarterly spending (winsorized to the 99th percentile) over the prior calendar year prior to attribution/participation; average quarterly inpatient utilization over the prior calendar year; average quarterly ED utilization over the prior calendar year; average quarterly AH event incidence over the prior calendar year; age; sex; and a flag for whether the original entitlement for Medicare is old age. Additionally, for the baseline spending and utilization variables, we included the change in spending (winsorized at the 99th and 1st percentiles) and utilization from the prior year.

In addition to the use of covariates in the matching process, we used six covariates to determine group membership in our subgroup analyses, including 4 variables not directly used in the matching algorithm: ADI (ZIP code-level), urban/rural status (county-level), CTO status, and

whether the practice was owned by a health system. The subgroup analysis section of the methodology provides more detail.

Finally, we leveraged the detailed, monthly eligibility information in the beneficiary demographics file as the basis for the analytic data set. We operationalized eligibility using the "ELIG_YYYY_MM" flags, which roll up information on Maryland residency, health maintenance organization (HMO) status (to distinguish FFS from Medicare Advantage enrollment), Parts A and B coverage, and date of death into one indicator for MDPCP eligibility.

Additional details on the construction of outcomes and covariates are presented in the Appendix.

Section 4. Methodology

This section discusses Hilltop's methodological approach to evaluating the causal impact of MDPCP on expenditure and utilization.

Overview

The goal of program evaluation is to estimate the causal effect of a program or initiative. This is typically conceptualized through a "potential outcomes" framework, and the causal impact of the program is defined as the difference between the observed outcomes for the treated group after implementation of the intervention, and what the outcomes for the treated group would have been had the group not been treated (Rubin, 2005).

Since it is not possible to observe outcomes for the treated group in both intervention and non-intervention states simultaneously—that is, the outcome with treatment and the outcome had the treated group not been treated—researchers rely on comparison groups as approximations of the counterfactual, or what would have happened to the treatment group in the absence of treatment. Treatment effects are then estimated by comparing outcomes for the treated group with those of the comparison group.

Random assignment of treatment yields unbiased, "gold standard" treatment effects. However, random assignment is often difficult (and potentially unethical) to implement in large-scale program rollouts. In these cases, it is incumbent on program evaluators to utilize an appropriate comparison group that represents the outcomes that the treatment group would have had in the absence of treatment. Failure to do so introduces selection bias in the estimate of the treatment's effect; that is, the treatment group differs from the comparison group because of both the treatment and other reasons. Thus, differences in outcomes between these groups are due to both the treatment and other reasons. In this case, the treatment effect contains bias.

This study uses two techniques to mitigate the extent to which selection bias may affect our results. First, as noted above, we used propensity score matching to construct a treatment-comparison group in which each individual in the treated group has a counterpart in the comparison group who was equally likely to have been in MDPCP but was not, ultimately, in

MDPCP. Then, using the matched data set, we used a difference-in-differences methodology to estimate the impact of MDPCP on beneficiary-level expenditure and utilization outcomes.

Analytic Populations

For this evaluation, we restricted analyses to Maryland residents for both treatment and comparison groups. This was intentional. Maryland is a unique state in terms of health policy, with both the All-Payer Model (2014-2018) and TCOC Model (2019-2026), and as of the time of writing, no other state had directly comparable policies. Given Maryland's unique position in the landscape of Medicare financing, a within-state research design is optimal for the evaluation of MDPCP. Using an out-of-state comparison group, researchers would not be able to tease apart the effect of MDPCP from the effect of the TCOC writ large.

Additionally, as noted above, Hilltop's treated group consisted of Medicare beneficiaries who were ever attributed to an MDPCP-participating provider, and our comparison group was drawn from the pool of Medicare beneficiaries who were ever attributed to a non-participating primary care provider over the study period (2019Q1–2022Q4) and who were never attributed to an MDPCP-participating provider over the study period. Thus, this methodology implicitly controls for a "primary care effect" and does not compare individuals attributed to an MDPCP-participating practice and individuals who are unconnected with the health system.

Finally, we imposed additional criteria on the beneficiary-quarter analytic data set. Specifically, we included only beneficiaries who were eligible for Medicare FFS Parts A and B, lived in Maryland, were not enrolled in Medicare Advantage, and were alive for at least one month in the given quarter. This final set of restrictions was intended to create a comparison of similar groups over time. Failing to impose this restriction could, for example, make it impossible to distinguish for a given individual whether they had genuinely zero utilization in a particular quarter, or whether that zero utilization was due to (for example) moving out of the state or dying in the previous quarter.

Matching

The individuals who were attributed to an MDPCP practice at any point during the four-year study period (2019-2022) constitute the basis for the treatment group for the analyses (ever-treated). Individuals who were at some point during the study period (2019-2022) attributed to a non-MDPCP primary care practice but who were never attributed to an MDPCP-participating provider constitute the basis for the comparison group (never-treated). To increase the likelihood that the comparison pool serves as an accurate counterfactual for the ever-treated population, we matched individuals from the never-treated pool to individuals from the ever-treated pool as of the latter's first quarter of attribution using pre-treatment, baseline, and time-invariant characteristics.

Hilltop's study employed a two-stage matching algorithm that was developed to meet the requirements of this evaluation (summarized in Table 4). In the first stage, we matched ever-treated individuals to never-treated individuals with exactly the same value or level of specified

variables. We required that matches be exact by year-quarter: that is, for an individual who joins an MDPCP practice in (for example) 2019Q1, we located a match from an individual in the pool of never-joiners who met the analytic data set eligibility requirements as of 2019Q1. Additionally, we enforced exact matching on race group, dual eligibility status as of the first quarter of attribution, and ZIP3 (i.e., the first three digits of each beneficiary's ZIP code) as of the first quarter of attribution.[5] Thus, for each individual in the treatment group, there is a corresponding individual in the comparison group of the same race group, with the same dual eligibility status as of the first quarter of attribution, and in the same geographic area as of the first quarter of attribution. This is intended to facilitate the subgroup analysis, in which we estimated the differential impact of MDPCP by race, dual status, locational characteristics, and practice characteristics. We sought to develop a methodology in which treated individuals with certain characteristics required for the subgroup analysis (for example, race) would have comparison-group analogues with those same characteristics. This was intended to mitigate the role for any potential composition bias in the subgroup analyses.

Table 4. Criteria Used in Two-Stage Matching Method

Criteria for Stage 1: Exact Match	Criteria for Stage 2: Propensity Score Match
Race group (Black, White, other*)	Age (continuous)
Medicare-Medicaid dual eligibility status (full or partial dual, not)	Sex (male, female)
Geographic area (ZIP3)	Medicare eligibility due to old age (yes, no)
Year-quarter	Average quarterly Medicare expenditure in prior calendar year
	Change in average quarterly Medicare expenditure
	Average quarterly inpatient utilization in prior calendar year
	Change in average baseline quarterly inpatient utilization
	Average quarterly ED utilization in prior calendar year
	Change in average baseline quarterly ED utilization
	Average quarterly AH event incidence in prior calendar year
	Change in average baseline quarterly AH event incidence

Notes: *The "Other" race group includes beneficiaries who are identified as Asian, North American Native, Hispanic/Latinx, Native Hawaiian, Other Pacific Islander, or any other racial-ethnic group (e.g., Middle Eastern, North African) in the beneficiary demographics file. These groups were combined into a single "Other" group because of the relatively small size of these groups compared to Black and White beneficiaries.
Baseline spending includes all Part A and B Medicare spending with the exception of Part B durable medical equipment.

There are 1,248 unique combinations from the exact matching of these four factors (three race groups; two dual eligibility statuses; 13 ZIP3 groups; and 16 year-quarter combinations from 2019Q1-2022Q4). We partitioned the cells by population in order to refine the matching process. For small cells, we conducted random 1:1 matches with replacement to link each treated individual with an individual from the comparison pool. For large cells, we conducted a second-stage match by estimating within-cell logistic regressions modeling an individual's

[5] While ZIP codes are not geographic polygons, the structure of ZIP codes implies that individuals in ZIP codes with the same first three digits should be relatively geographical proximal. The first three digits of ZIP codes indicate the "Sectional Center Facility," a centralized sorting and distribution hub that distributed mail to post offices according to the fourth and fifth digits of a given ZIP code. For additional information, see Stevens (2006) and U.S. Postal Service Office of Inspector General (2013).

likelihood of attribution to an MDPCP-participation provider on baseline characteristics. We defined "large" cells as those with an events-per-variable rate of at least 15. Given that the propensity score models include eleven covariates, this translates to cells with at least 165 MDPCP-attributed individuals. Small cells are defined as those with fewer than 165 MDPCP-attributed individuals (Peduzzi et al., 1996).

For each of the large cells, we estimated logistic propensity scores to model the probability that individual *i* is attributed to an MDPCP-participating provider as a function of baseline covariates. Given that we conducted exact matching based on race group, dual eligibility status from the prior quarter, ZIP3, and year-quarter, we use a parsimonious set of covariates for these within-cell propensity models to maximize the probability of model convergence. In particular, we included the average quarterly spending (Parts A and B, winsorized to the 99th percentile) over the prior calendar year (that is, year t-1); average quarterly inpatient utilization over the prior calendar year; average quarterly ED utilization over the prior calendar year; average quarterly AH event incidence over the prior calendar year; age; sex; and a flag for whether the original entitlement for Medicare is old age. Additionally, for the baseline spending and utilization variables, we included the change in spending (winsorized at the 99th and 1st percentiles) and utilization from the prior year. Specifically, we estimated:

$$
\begin{aligned}
In\,MDPCP_{i,t} = {} & \alpha + \beta_1 Avg.Quarterly\ Medicare\ Spending_{i,t-1} \\
& + \beta_2 Avg.Quarterly\ Inpatient\ Stays_{i,t-1} + \beta_3 Avg.Quarterly\ ED\ Visits_{i,t-1} \\
& + \beta_4 Avg.Quarterly\ AH\ Events_{i,t-1} \\
& + \beta_5 \Delta Avg.Quarterly\ Medicare\ Spending_{i,t-2,t-1} \\
& + \beta_6 \Delta Avg.Quarterly\ Inpatient\ Stays_{i,t-2,t-1} \\
& + \beta_7 \Delta Avg.Quarterly\ ED\ Visits_{i,t-2,t-1} \\
& + \beta_8 \Delta Avg.Quarterly\ AH\ Events_{i,t-2,t-1} + \beta_9 Age_{i,t} + \beta_{10} Sex\ (Female)_i \\
& + \beta_{11} Medicare\ Eligibility\ due\ to\ Old\ Age_i + \epsilon_{i,t}
\end{aligned}
$$

This methodology estimates separate propensity score regressions within each of the large-cell exact match groups (that is, each combination of race group, ZIP3, dual eligibility status, and year-quarter), and allows for differing dynamics of MDPCP participation—both in terms of which practices participate in MDPCP and which individuals are attributed to MDPCP-participating practices—for each exact match group. Within each exact-matching group, we conducted 1:1 matching with replacement based on coarsened propensity score. Where there are multiple matches for a given ever-treated individual based on coarsened propensity score, we randomly selected a matched never-treated individual.[6]

The final analytical sample consists of 15,278,197 person-quarters comprising 547,390 ever-treated individuals. Each ever-treated individual has a matched counterpart in the never-treated population. Since we matched with replacement, individuals can appear multiple times from the

[6] Many matching algorithms are possible. We selected this matching algorithm to meet the needs of the subgroup analysis and balance rigor with methodological transparency.

never-treated population; thus, the 547,390 matches constitute 222,824 unique individuals from the never-treated population.

Table 5 presents the distribution of first treatment cohort for the "ever" joiners. Hilltop found that 38.3% of all treated individuals first joined MDPCP in 2019Q1; 23.6% first joined in 2020Q1; and 9.1% first joined in 2021Q1. Together, these three primary cohorts compose 71.0% of the unique beneficiary count from the ever-treated population.

Table 5. Distribution of Treated Individuals by First Joining Quarter

Joining Cohort	Total Unique Beneficiaries	% of Total Attributed Beneficiaries from 2019-2022
2019Q1	209,879	38.3%
2019Q2	2,742	0.5%
2019Q3	3,125	0.6%
2019Q4	3,786	0.7%
2020Q1	129,099	23.6%
2020Q2	30,544	5.6%
2020Q3	16,334	3.0%
2020Q4	15,322	2.8%
2021Q1	49,658	9.1%
2021Q2	14,783	2.7%
2021Q3	10,344	1.9%
2021Q4	14,282	2.6%
2022Q1	12,773	2.3%
2022Q2	79	0.0%
2022Q3	22,257	4.1%
2022Q4	12,383	2.3%
Total	**547,390**	**100.0%**

Notes: This table displays the distribution of individuals attributed to MDPCP-participating practices by their first quarter of exposure to MDPCP.

Table 6 shows the distribution of beneficiary quarters in the analytic data set following the start of MDPCP by joining cohort. Hilltop found that the 2019Q1 joiners constitute 51.5% of the analytic data set, and the top three joining cohorts (2019Q1, 2020Q1, and 2021Q1) together compose 82.0% of treated experience. While the 2019Q1 joiners are the largest joining cohort, they also have the longest treatment duration in the analytic data set: from 2019Q1–2022Q4. As such, this cohort has an outsized effect on the overall MDPCP impact estimates.

Table 6. Distribution of Treated Individuals across Study Period

Joining Cohort	Total Beneficiary Quarters	% of Total Attributed Beneficiary Quarters
2019Q1	2,893,880	51.5%
2019Q2	36,338	0.6%
2019Q3	38,220	0.7%
2019Q4	43,109	0.8%
2020Q1	1,362,573	24.2%
2020Q2	300,008	5.3%
2020Q3	145,585	2.6%
2020Q4	123,508	2.2%
2021Q1	352,695	6.3%
2021Q2	96,314	1.7%
2021Q3	58,477	1.0%
2021Q4	67,108	1.2%
2022Q1	47,818	0.9%
2022Q2	228	0.0%
2022Q3	44,133	0.8%
2022Q4	12,383	0.2%
Total	**5,622,377**	**100.0%**

Notes: This table displays the distribution of post-treatment experience for individuals attributed to MDPCP-participating practices by their first quarter of exposure to MDPCP.

Balance

Hilltop assessed the success of the matching process using standardized differences along baseline covariates. Small standardized differences indicate that the matched comparison population statistically resembles the treated population and increases the likelihood that the comparison group serves as a valid counterfactual for the treated group.

Pre-Match

First, we examined the balance of the treatment and comparison pool prior to the matching process. For this analysis, we used the pool of treated individuals as of their first quarter of attribution, as well as the entire pool of potential comparison individuals (across all quarters). See Table 7 for the results.

Table 7. Balance Tests for Treatment Pool and Comparison Pool Prior to Match

Baseline Characteristic	Treatment Pool Mean	Comparison Pool Mean	Standardized Difference
Average baseline quarterly Medicare Parts A & B expenditure*	$3,499.1	$3,568.5	-0.0091
Average baseline quarterly inpatient utilization	0.1627	0.1651	-0.0047
Average baseline quarterly ED utilization	0.2645	0.2579	0.0113
Average baseline quarterly AH event incidence	0.056	0.0537	0.0083
Change in average baseline quarterly Medicare Parts A and B expenditure†	$909.9	$689.3	0.0271
Change in average baseline quarterly inpatient utilization	0.0353	0.0235	0.0198
Change in average baseline quarterly ED utilization	0.0334	0.0179	0.0235
Change in average baseline quarterly AH event incidence	0.0114	0.0067	0.0146
Original Medicare Entitlement due to Age	82.08%	78.91%	0.0802
Sex (female)	58.97%	56.57%	0.0486
Age in years, mean	72.1	72.5	-0.0443
Race (Black)	22.4%	24.7%	-0.0556

Notes: *Spending is winsorized at the 99th percentile, with negative values re-coded to 0.
†Change in spending in winsorized at the 99th and 1st percentiles.

Hilltop found that, overall, the comparison pool and treatment pool were reasonably similar by baseline utilization: baseline spending and utilization do not meaningfully differ. We interpreted this as support for our first filter of requiring that all comparison individuals were at one point attributed to a non-participating primary care practice over this period. However, there were significant differences in other characteristics. Spending and utilization were rising significantly for the pool of treated individuals relative to the comparison pool, which speaks to the need to match based on these factors. Additionally, there were demographic differences: the treatment pool was significantly more likely to have original Medicare entitlement due to old age than the comparison pool; was more likely to be female; was younger; and was less likely to be Black.

Post-Match

As noted above, the matching process identified a counterpart for every MDPCP-attributed individual at the first quarter of attribution, matching exactly on year-quarter, race group, dual eligibility status, and ZIP3—and using propensity score matching on sex, age, original Medicare eligibility reason, baseline spending and utilization, and change in baseline spending and utilization. Table 8, below, presents results of the balance tests for the matched treatment and comparison samples.

Table 8. Balance Tests for Matched Treatment and Comparison Group

Baseline Characteristic	Treatment Mean	Matched Comparison Mean	Standardized Difference
Average baseline quarterly Medicare Parts A & B expenditure *	$3,384.3	$3,435.4	-0.0071
Average baseline quarterly inpatient utilization	0.1574	0.1594	-0.0041
Average baseline quarterly ED utilization	0.2568	0.2591	-0.0041
Average baseline quarterly AH event incidence	0.0532	0.0538	-0.0026
Change in average baseline quarterly Medicare Parts A & B expenditure[2]	$905.2	$855.0	0.0066
Change in average baseline quarterly inpatient utilization	0.0339	0.031	0.0051
Change in average baseline quarterly ED utilization	0.0325	0.0285	0.0063
Change in average baseline quarterly AH event incidence	0.0108	0.0093	0.0047
Original Medicare Entitlement due to Age	82.51%	82.73%	-0.0058
Sex (female)	58.92%	59.38%	-0.0092
Age in years, mean	72.1	72.4	-0.0273
Race (Black)	22.07%	22.07%	0.0000

Notes: *Spending is winsorized at the 99th percentile, with negative values re-coded to 0.
†Change in spending in winsorized at the 99th and 1st percentiles.

These results indicate strong balance between the treatment and comparison samples. Hilltop noted that individuals in the treatment group were, on average, 0.3 years younger than individuals in the comparison group. Internal analysis indicates that this appears to be driven by the smaller joining cohorts, for whom a proportionally higher number is matched using only exact matching without the propensity score match refinement.

Appendix Table A.2 presents the balance tests separately for the primary joining cohorts (2019Q1, 2020Q1, and 2021Q1).

Difference-In-Differences Methodology

Using this matched treatment and comparison groups, we conducted a difference-in-differences analysis. This is a widely used technique that estimates the effect of a program by estimating changes in outcomes over time between the treated and comparison groups. The treatment effect is identified as the change in outcomes for the treatment group, net of any changes that also occurred in the comparison group over the same period. The remaining difference is attributed to the treatment effect of the program.

Hilltop estimated the impact of MDPCP on spending and utilization using a difference-in-differences approach designed for data with multiple time periods and staggered treatment onsets (Callaway & Sant'Anna, 2021). This method is conceptually similar to a traditional difference-in-differences analysis in that it quantifies the impact of an intervention by comparing

treatment and comparison groups over time before and after starting the intervention. However, it differs in that it isolates "clean" comparisons between individuals in the treatment and comparison groups over time and excludes "forbidden" comparisons that occur when members of the treatment group start receiving the treatment at different times and that can result in biased, inaccurate causal estimates (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021; Roth et al., 2023). Further, the difference-in-differences estimator we used relies on the assumption of post-treatment parallel trends rather than the stronger assumption of parallel trends for all groups and time periods (Callaway & Sant'Anna, 2021). The post-treatment parallel trends assumption is generally more realistic to satisfy, especially in situations with staggered treatment onsets. It only assumes that, had MDPCP not occurred, the average change in outcomes for the treatment and comparison groups would have been similar during the treated year-quarters (Marcus & Sant'Anna, 2021; Roth et al., 2023).

This methodology estimates separate 2x2 difference-in-differences regressions for each starting cohort (g) (e.g., 2019Q1, 2019Q2, 2019Q3) and each post-treatment year-quarter (t) resulting in group-time average treatment effect, or ATT(g,t). The final estimates are then aggregated across the constituent sub-estimates using weighted averages (Callaway & Sant'Anna, 2021; Rios-Avila et al., 2021). While many estimates are possible, we focused on three sets of outcomes: the overall average treatment on the treated (ATT); the overall ATT for the three largest joining cohorts (2019Q1, 2020Q1, and 2021Q1); and the event time estimates for the three largest cohorts. It is important to note that our methodology leverages the longitudinal structure of the data and includes individual-level fixed effects that control for unobserved, time-invariant factors (such as, for example, individual-level general health status). This specification uses within-individual changes in outcomes to estimate the effects of MDPCP; that is, for each underlying 2x2 difference-in-difference regression, the estimate of the ATT(g,t) is a weighted average of individual-level changes over time, net the corresponding changes in the individuals in the comparison group. Additionally, while it is in principle possible to use both the never-treated and not-yet-treated groups as comparators for the treated population, this evaluation uses the never-treated groups because, due to the matching, this population should serve as a reliable counterfactual for the treatment group.[7]

Crucially, this methodology isolates the impact of MDPCP by focusing on the individual level; our results are not intended to capture group-level effects (as if, for example, MDPCP practices retain high-cost patients at a higher rate than non-MDPCP practices). While the methodology does not require individuals to remain consistently enrolled in MDPCP, results are only identified for a given post-MDPCP period based on individuals that are present in the analytic data set; moreover, only individuals that are present in the pre-MDPCP period and at least one post-MDPCP period will contribute to the overall impact estimates.

[7] Additionally, as noted in Appendix Table A.2, the composition of MDPCP joiners changes over time. The 2019Q1 joining cohort tends to be older, more likely White, and with more individuals having "old age" as the basis for original Medicare eligibility; the later cohorts, however, tend to be younger, more likely Black, with fewer individuals having "old age" as the basis for original Medicare eligibility. This relative dissimilarity argues against using the not-yet-treated as a counterfactual for treated individuals.

Finally, to account for the fact that our matching method results in a degree of duplication among the comparison beneficiaries—and thus correlation in errors—we clustered standard errors at the individual level. All analyses were conducted in Stata 18.

Methodology for Subgroup Analyses

Hilltop estimated six sets of subgroup analyses: by race group, dual status in the first quarter of attribution, ADI level, rural/urban status, CTO status, and practice ownership. To facilitate this analysis, we attributed each characteristic to each pair of treatment-comparison individuals as time-invariant; thus, for example, we used the pool of MDPCP-attributed individuals that were ever attributed to a practice that was matched with a CTO, and we used the corresponding matched comparison individuals as comparators for these treated individuals. We performed all subgroup analyses in an analogous manner. See the Appendix for more details regarding the methodology used for the subgroup analyses.

Methodology for Analysis of the Impact of COVID-19

Hilltop's study tested whether the COVID-19 pandemic altered the effect of MDPCP on utilization or spending. To do this, we estimated quarterly impacts of MDPCP on the four outcomes—total Medicare expenditure, inpatient utilization, ED utilization, and AH event incidence—by calendar time. This necessarily conflates experience in the program across joining cohorts (for example, the 2020Q1 "effect" will be the first quarter in the program for 2020Q1 joiners, and the 5th quarter in the program for the 2019Q1 joiners) but allowed us to statistically test whether the MDPCP effect significantly differed during the COVID-19 period relative to the pre-COVID-19 period. We then aggregated the individual quarter-level estimates to two periods: pre-COVID-19, defined as 2019Q1–2020Q1, and during COVID-19, defined as 2020Q2–2021Q4. This analysis excludes individuals that joined MDPCP during 2022Q1 or later. Confidence intervals were created by the bootstrapping method to compare the quarter-level MDPCP effects during and prior to COVID periods.

Pre-Treatment ATT(g,t) Effects

In order to evaluate whether the parallel trend assumption holds in the matched analytic data set, Hilltop performed a visual inspection and statistical test of the pre-treatment ATT(g,t) estimates as a placebo check. These placebo checks indicate whether the ever-treated and never-treated comparison groups have different outcomes prior to their involvement in MDPCP. For the visual inspection, we plotted the ATT(g,t) point estimates and 95% confidence intervals for each year-quarter for each joining cohort (i.e., 16 individual plots per outcome) and evaluated whether there was evidence of differential pre-trends at the cohort-level. Additionally, we used the statistical test of pre-treatment ATT(g,t) effects that is built into the CSDID Stata module to test for two potential effects: 1) aggregated pre-treatment effects across all pre-treatment windows (aggregated chi-square) and 2) pre-treatment effects specifically between the last pre-treatment period and the treatment period because that most closely approximates the post-treatment parallel trends assumption (post-treatment chi-square). Both of these chi-square tests have the null hypothesis that the pre-treatment ATT(g,t) effects = 0 and statistically

significant results would be evidence against the parallel trends assumption in the analytic data set.

Table 9 presents the results. There was no evidence consistent with the violation of the parallel trends assumption.

Table 9. Pre-Treatment ATT(g,t) Effects by Outcome

Outcome	Aggregated Chi-Square, p value	Post-Treatment Chi-square, p value
Total Medicare Parts A & B expenditure	$X^2(48) = 2.6900$, $p > 0.9999$	$X^2(16) = 3.4440$, $p = 0.9996$
Any inpatient utilization	$X^2(48) = 3.1844$, $p > 0.9999$	$X^2(16) = 3.1844$, $p = 0.9997$
Any ED utilization	$X^2(48) = 0.0295$, $p > 0.9999$	$X^2(16) = 5.3230$, $p = 0.9939$
Any AH event	$X^2(48) = 0.3957$, $p > 0.9999$	$X^2(16) = 0.4511$, $p > 0.9999$

Notes: ATT – average treatment effect on the treated. G-group. T-time.

Pre-Trends

We also assessed the extent to which, using these matched treatment and comparison groups, the outcomes trended differentially for ever-treated and never-treated comparison units prior to treatment. That is, for individuals and their matched comparison group counterparts who are first attributed to MDPCP in quarter *q* of year *y*, we assessed the extent to which total spending, inpatient admissions, ED utilization, and AH event rates trended differentially in the four quarters prior to (but not including) the first quarter of attribution. Differential pre-trends are a threat to the validity of the analysis: they suggest that any post-attribution differences may have occurred regardless of MDPCP.

We statistically tested for the presence of parallel pre-trends using the following specification for all periods prior to the first quarter of attribution for each ever-treated individual and matched comparison individual:

$$Outcome_{i,q} = \gamma_1 * t_q + \gamma_2 * t_q * Ever\ MDPCP_i + \alpha_i$$

For each outcome, we regressed the outcome on a time trend, and on the time trend interacted with a 0/1 indicator for whether the individual *i* will be in the MDPCP attribution sample in the future. We included individual-level fixed effects (α_i), as in our baseline specification; as such, the indicator for $Ever\ MDPCP_i$ is not identified in this pre-trend analysis. A statistically significant coefficient on the interaction term – that is, γ_2 – indicates that, controlling for time-invariant individual-level factors, the outcomes were trending differentially for the treatment and comparison group prior to the start of MDPCP, and thus calls the validity of the regressions into question. For this exercise, the standard errors were clustered at the individual level to account for duplicates in the matched comparison group.

Table 10 presents the results.

Table 10. Pre-Trend Test by Outcome

Outcome	Differential Time Trend (γ_2)	*p* value	N
Total Medicare Parts A & B expenditure	$12.51	0.112	4,233,501
Any inpatient utilization	0.0001	0.642	4,233,501
Any ED utilization	-0.0003	0.329	4,233,501
Any AH event	-0.0002	0.105	4,233,501

Notes: All regressions include individual-level fixed effects and standard errors are clustered at the individual level. N – number of observations.

While this does not confirm that the counterfactual trend in outcomes for the treated group would have been parallel to the actual trend in outcomes, these results indicate that pre-treatment outcome trends in the comparison group did not systematically differ from those in the treatment group across all cohorts and supports the use of the difference-in-differences methodology.

Data Granularity

We operationalized our matching model at the individual level, and our difference-in-differences model at the beneficiary-quarter level. That is, in our difference-in-differences specification, each observation is the expenditure, utilization, and treatment status for individual *i* in quarter *q* of year *y*. The following sub-sections elaborate on both operational decisions.

Individual vs. Practice-Level

Since MDPCP is operationalized at the practice-level—that is, practices decide whether or not to opt into MDPCP—the ideal comparison group would be made up of individuals who resemble treated individuals, who are attributed to practices that resemble MDPCP practices, and who live in areas that are impacted by similar laws, policies, and other external factors. In practice, however, it is difficult to identify comparison practices that resemble MDPCP practices. As noted above, MDPCP had a wide reach upon its introduction and is voluntary; as such, the non-participating practices within Maryland may be relatively inapt comparators to participating practices. In principle, it could be possible to locate apt comparator practices from other states, but doing so would risk conflating the impact of Maryland's unique system writ large, with that of MDPCP specifically. Simply stated, the quarterly attribution files for non-participating providers—upon which we rely to construct our comparison group—do not provide information about non-participating practices.

Taken together, this implies that it is conceptually challenging—or potentially impossible—to locate apt comparator practices. Moreover, data limitations prevent us from attempting to do so. Thus, we focused on identifying program estimates at the individual level rather than the practice level by matching similar individuals who did and did not participate in MDPCP, and then estimating the within-person change in outcomes over time. While this should, in principle, yield accurate program impacts, it does not allow us to identify practice-level mechanisms through

which MDPCP effects operate. Additionally, it effectively rules out changing practice composition as a measurable impact of MDPCP. To the extent that this empirical strategy allows us to estimate accurate impact estimates, we believe that the trade-off is justified.

Quarterly vs. Yearly

Hilltop opted for a person-quarter granularity rather than person-year granularity for three reasons. First, MDPCP attribution occurs quarterly, and individuals can—and do—enter or leave MDPCP practices throughout the course of the year. Using a person-quarter architecture allowed us to define the treatment period for individuals as of the first quarter of attribution; using a person-year architecture, by contrast, could potentially bias detectable MDPCP effects downward since it would be necessary to impute pre-treatment experience to the "treatment" period for individuals joining later in a calendar year (for example, in 2019Q4). Second, the person-quarter granularity allowed us to more precisely estimate the impact of COVID-19. A nationwide emergency was declared on March 13, 2020;[8] thus, while this overlaps with the final 19 days of 2020Q1, we believe that the first quarter represents a sufficiently complete snapshot of pre-COVID spending and utilization, and that assessing differential program effects beginning in 2020Q2 provides a natural means of assessing the differential impact of COVID-19. Finally, this person-quarter architecture allowed us to test for immediate program impacts.

However, we recognize that the use of a person-quarter architecture may make it more difficult to detect treatment effects; while underlying variance in the outcome variables would likely be lower than that of a person-year architecture, the theoretical treatment effect should also be lower than that of an annual granularity. Thus, it is possible that we are unable to detect small treatment effects in a person-quarter architecture that we would be able to detect in a person-year architecture. Moreover, person-quarter estimates may be more subject to intra-year variability in spending and utilization if, for example, treatment and comparison individuals systematically—and differentially—alter their spending and utilization over the course of the year. While we are unaware of any evidence to suggest this, we replicated our person-quarter analysis at the person-year level and included results in the Appendix.

Interpretation of Results

These results should be interpreted as an "intent to treat" estimate. Individuals were not required to remain in MDPCP for the duration of the study window; they were only required to be attributed to an MDPCP-participating provider at least once (and treatment was assumed to be an absorbing state). Thus, while some individuals in the treatment group may remain in MDPCP for the entirety of the study period, this is not required; at the extreme, it is possible for an individual to be attributed to an MDPCP-participating provider once and then exit MDPCP (but remain in the analytic data set).

[8] See CMS press release for more detail: https://www.cms.gov/newsroom/press-releases/cms-takes-action-nationwide-aggressively-respond-coronavirus-national-emergency

It is also important to note that this analysis uses the time of individual-level attribution to signal the start of "treatment" and not the time period when a *practice* joins MDPCP. While this will often overlap, in our analysis, it is possible for individuals to, for example, enter a practice in 2021Q1 that had joined MDPCP in 2019Q1. We believe that our use of both sources of variation (individual-level and practice-level) in the timing of the start of MDPCP mitigates, to a certain extent, potential non-random selection by practices into MDPCP.

Section 5. Results

This section describes the study participants and reports the impact of MDPCP on Medicare utilization and spending. In order to provide a comprehensive picture of the results for each outcome, we summarize the impact of MDPCP in three ways: 1) overall, which aggregates impact estimates for all cohorts across all timepoints; 2) overall joining cohort, which presents aggregated overall impact estimates by cohort for the three largest cohorts (i.e., 2019Q1, 2020Q1, 2021Q1); and 3) by treatment time and joining cohort, which shows the treatment effect at the year-quarter-level separately for the three largest cohorts.

In all results tables, bolded values in the "Effect" and "95% Confidence Interval (CI)" columns indicate that the 95% CI did not include zero, meaning that the effect was statistically significant at the $p<0.05$ level.

Matched Sample

The final analytical sample consists of 547,390 ever-treated individuals and their matched counterparts in the comparison population. Since we match with replacement, individuals can appear multiple times from the comparison population; thus, the 547,390 matches constitute 222,824 unique individuals from the comparison population. Table 11 below summarizes the demographic characteristics of the final analytic data set.

Table 11. Demographic Characteristics of the Matched Study Participants in the Analytic Data Set

	Treatment Group N=547,390	Matched Comparison Group N=547,390	Total Sample N=1,094,780
Mean age in years (SD)	72.14 (10.40)	72.43 (10.93)	72.28 (10.67)
Race and Ethnicity			
Asian*	16,339 (3.0%)	20,081 (3.7%)	36,420 (3.3%)
Black	120,795 (22.1%)	120,795 (22.1%)	241,590 (22.1%)
Hispanic/Latinx*	13,064 (2.4%)	11,375 (2.1%)	24,439 (2.2%)
North American Native*	238 (0.0%)	239 (0.0%)	477 (0.0%)
Other*†	4,297 (0.8%)	4,326 (0.8%)	8,623 (0.8%)
Unknown*	12,974 (2.4%)	10,891 (2.0%)	23,865 (2.2%)
White	379,683 (69.4%)	379,683 (69.4%)	759,366 (69.4%)
Sex			
Female	322,539 (58.9%)	325,014 (59.4%)	647,553 (59.1%)
Male	224,851 (41.1%)	222,376 (40.6%)	447,227 (40.9%)

	Treatment Group N=547,390	Matched Comparison Group N=547,390	Total Sample N=1,094,780
Original Medicare Entitlement Due to Age			
Yes	451,668 (82.5%)	452,861 (82.7%)	904,529 (82.6%)
No	95,722 (17.5%)	94,529 (17.3%)	190,251 (17.4%)
Medicaid-Medicaid Dual Eligible			
Full or partial dual	81,819 (14.9%)	81,819 (14.9%)	163,638 (14.9%)
Non-dual	465,571 (85.1%)	465,571 (85.1%)	931,142 (85.1%)
County of Residence			
Urban	466,792 (85.3%)	462,213 (84.4%)	929,005 (84.9%)
Rural	80,589 (14.7%)	85,159 (15.6%)	165,748 (15.1%)
Area Deprivation Index (ADI)			
In highest 20%	93,603 (17.2%)	99,736 (18.4%)	193,339 (17.8%)
Not in highest 20%	449,812 (82.8%)	442,537 (81.6%)	892,349 (82.2%)

Notes: *Included in the "Other" race group for matching and subgroup analyses.
†"Other" references any other racial or ethnic group (e.g., Middle Eastern, North African)
N- number of observations. SD- standard deviation.

Table 12 shows the average values of study outcomes for the quarter immediately prior to the start of MDPCP for the treatment and matched comparison populations.

Table 12. Baseline Value of Outcome Variables, by Treatment and Comparison Status

Baseline Characteristic	Treatment Group Mean	Matched Comparison Group Mean
Quarterly total Medicare Parts A & B expenditure	$2,764.3	$2,761.9
Any quarterly inpatient utilization	0.0418	0.0420
Any quarterly ED utilization	0.0765	0.0758
Any quarterly AH event	0.0153	0.0154

Note: This table shows the average value of study outcomes in the quarter immediately prior to the start of MDPCP treatment exposure for the MDPCP-attributed population and matched comparators.

It is important to note several factors in this table. First, since this is a comprehensive analysis of MDPCP, the first quarter of attribution ranges from 2019Q1–2022Q4; as such, the pre-MDPCP values in this table span 2018Q4–2022Q3. Second, since each of the three utilization variables (inpatient, ED, and AH events) is a binary variable, their means can be interpreted as percentages: that is, on average, 4.18% of individuals attributed to MDPCP practices had at least one inpatient service in the quarter immediately prior to the start of MDPCP, while 4.20% of matched comparison individuals had at least one inpatient service in the quarter immediately prior to the start of MDPCP for their matched treatment unit. Similarly, 7.65% of MDPCP-attributed individuals had at least one ED service in the quarter immediately prior to the start of MDPCP, compared to 7.58% of comparison individuals. Finally, 1.53% of MDPCP-attributed individuals had at least one AH event in the quarter immediately prior to the start of MDPCP, compared to 1.54% of comparison individuals.

Third, the relative frequencies indicate that it may be challenging to detect effects for AH events. This is a relatively rare event, with a quarterly incidence rate of approximately 1.5%; to the extent that underlying risk is steady over time, then our difference-in-differences methodology—which uses individual fixed effects to control for underlying, time-invariant factors at the individual level—may also account for a meaningful portion of the variation in this measure.

Finally, this mirrors the balance tests shown above: treatment and matched comparison individuals are well-balanced by average pre-treatment outcomes. While the difference-in-differences method does not require this condition, it supports our matching strategy.

Overall Results

Quarterly Effect Estimates

Table 13 presents overall results for each of the four outcomes. The effect size as a percentage of baseline is calculated as the effect size relative to the average value of each outcome in the quarter immediately prior to the first quarter of attribution for the treated population.

Table 13. Overall MDPCP Impact Estimates

Outcome	Quarterly Effect Estimate *% Change**	95% Confidence Interval *% Change**	
		Lower	Upper
Total Medicare Part A & B expenditure	**-$119.60** **-4.33%**	**-$151.66** **-5.49%**	**-$87.54** **-3.17%**
Any inpatient utilization	**-0.0030** **-7.18%**	**-0.0038** **-9.10%**	**-0.0021** **-5.03%**
Any ED utilization	**-0.0013** **-1.70%**	**-0.0025** **-3.27%**	**-0.0002** **-0.03%**
Any AH event	-0.0003 -1.96%	-0.0008 -5.22%	0.0003 1.96%

Notes: The adjusted effect estimates presented are average changes in the measure level (spending) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter of participation, after accounting for changes among matched comparators and adjusting for individual fixed effects. *Defined as the percentage change in average value of the outcome from the calendar quarter before a beneficiary's participation in MDPCP began. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

Hilltop found that, across all treated cohorts, the introduction of MDPCP led to significant declines in expenditure, inpatient utilization, and ED utilization. Relative to matched individuals and controlling for time-invariant, individual-level factors, individuals attributed to MDPCP experienced an average spending reduction of $119.60 per quarter, or 4.33% of baseline expenditure. This effect is likely driven by a statistically significant 7.18% reduction in inpatient utilization. Hilltop also found evidence for a smaller effect on ED utilization, reducing by an average of 1.70% per quarter from baseline, and statistically significant at the 5% level. Finally, Hilltop estimated a negative impact of MDPCP on AH events but was unable to statistically differentiate this effect from zero. This relative statistical imprecision may be due to the rarity of

AH events: as noted in Table 13, above, only approximately 1.5% of treatment and comparison group individuals experienced AH events in the quarter immediately prior to participation in MDPCP.

Results by Primary Cohort

The Hilltop team disaggregated the results into the three largest joining cohorts in order to better understand effect heterogeneity over time: 2019Q1 joiners (38.3% of all attributed beneficiaries from 2019 to 2022), 2020Q1 joiners (23.6%), and 2021Q1 joiners (9.1%). Collectively, these cohorts constitute 71% of the ever-treated sample and 82% of the treated experience. See Table 14 for the results by primary cohort.

Table 14. Results by Primary Cohort

Outcome	Quarterly Effect Estimate *% Change**	95% Confidence Interval *%Change**	
		Lower	Upper
2019 Q1 Joiners (38.3% of ever-treated)			
Total Medicare Parts A & B expenditure	**-$123.15** ***-4.51%***	**-$178.28** ***-6.52%***	**-$47.82** ***-1.75%***
Any inpatient utilization	**-0.0033** ***-7.22%***	**-0.0053** ***-11.59%***	**-0.0013** ***-2.84%***
Any ED utilization	-0.0005 *-0.61%*	-0.0035 *-4.26%*	0.0025 *3.04%*
Any AH event	0.0001 *-0.59%*	-0.0012 *-7.03%*	0.0014 *8.20%*
2020 Q1 Joiners (23.6% of ever-treated)			
Total Medicare Parts A & B expenditure	-$29.10 *-1.00%*	-$114.40 *-3.93%*	$55.99 *1.92%*
Any inpatient utilization	-0.0009 *-2.02%*	-0.0031 *-6.95%*	0.0014 *3.14%*
Any ED utilization	-0.0018 *-2.13%*	-0.0047 *-5.55%*	0.0012 *1.42%*
Any AH event	-0.0006 *-3.34%*	-0.0021 *-11.68%*	0.0009 *5.01%*
2021 Q1 Joiners (9.1% of ever-treated)			
Total Medicare Parts A & B expenditure	-$32.27 *-1.06%*	-$164.47 *-5.41%*	$99.50 *3.27%*
Any inpatient utilization	-0.0015 *-4.01%*	-0.0047 *-12.57%*	0.0017 *4.55%*
Any ED utilization	-0.0012 *-1.85%*	-0.0051 *-7.85%*	0.0026 *4.00%*
Any AH event	0.0011 *9.85%*	-0.0003 *-2.69%*	0.0026 *23.29%*

Table 14 Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. *Defined as the percentage change in average value of the outcome from the calendar quarter before a beneficiary's participation in MDPCP began. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level. Q1- first quarter of the calendar year.

Hilltop documented heterogeneous effects by joining cohort. Specifically, the 2019Q1 joiners experience significant declines in expenditures and inpatient utilization, but these effects are smaller in magnitude and lack statistical precision for later joiners. This may, in part, be due to the nature of the MDPCP uptake: participation is voluntary, and the practices that anticipated the greatest benefit in terms of patient impact may have been the first to enroll.

Results by Time in MDPCP by Primary Cohort

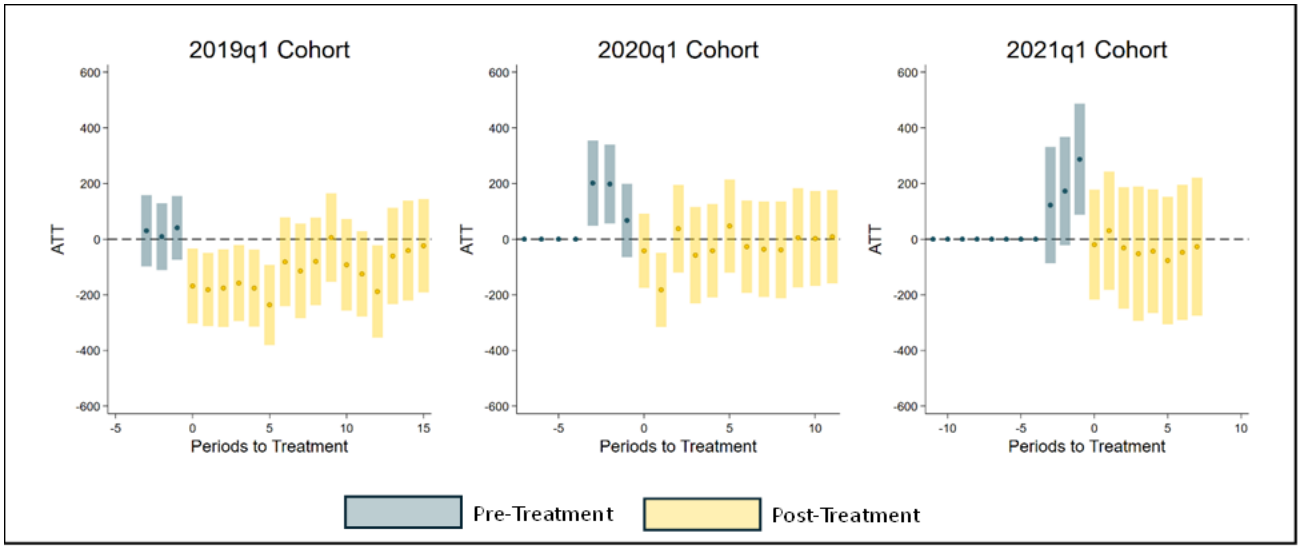
Finally, we present results by the number of quarters since an individual was first attributed to MDPCP. Aggregated event-time analyses necessarily conflate calendar time with event time: that is, for a 2019Q1 joiner, the first quarter in the program will be 2019Q1; for a 2020Q1 joiner, the first quarter in the program will be 2020Q1. In the absence of significant exogenous shocks, these first quarter estimates can be aggregated to estimate an overall "first quarter" effect. Given the magnitude of the COVID-19 shock, however, we do not aggregate event time across all cohorts; instead, we present event-time estimate for the three primary joining cohorts (2019Q1, 2020Q1, and 2021Q1).

Total Medicare Parts A & B Expenditure

Figure 1, below, presents the event-time estimates of the impact of MDPCP on total spending for the three primary cohorts. In these graphs, each vertical bar represents the 95% CI for the difference in total expenditure between the MDPCP and comparison groups in that joining cohort for a single year-quarter and the dots on the bars represents the difference-in-differences point estimate. Blue bars indicate quarters prior to the joining cohort's initial participation in MDPCP, while yellow bars indicate quarters after the cohort members were first attributed to an MDPCP practice. The yellow bars can be used to understand the impact of MCPCP on each cohort over time. For the 2019Q1 joiners, we documented a consistent and statistically significant effect starting in the first quarter that individuals were attributed to MDPCP-participating practices. This effect is consistent until their fifth quarter from initial attribution, which coincides with 2020Q2—the first quarter in which the effect of COVID-19 should be apparent. MDPCP effects for subsequent quarters, while negative, are not individually statistically significant. For the 2020Q1 joining cohort, the effect is statistically significant and negative during the second quarter following initial attribution (that is, 2020Q2) and negative for all other periods, but not statistically significantly so. Finally, for the 2021Q1 joiners, the individual quarter-level effects, while negative, are not statistically significant. It is worth noting that spending was trending downward for the 2020Q1 joiners in the quarters immediately prior to joining, while spending was trending upward for the 2021Q1 joiners during the same pre-treatment period. Ideally, while pre-treatment outcomes do not trend differentially prior to

treatment, the differing directions of these later-cohort pre-trends suggest that systematic confounders do not play a role.

Figure 1. Event-Time Results for Total Medicare Parts A + B Expenditure, by Primary Cohort

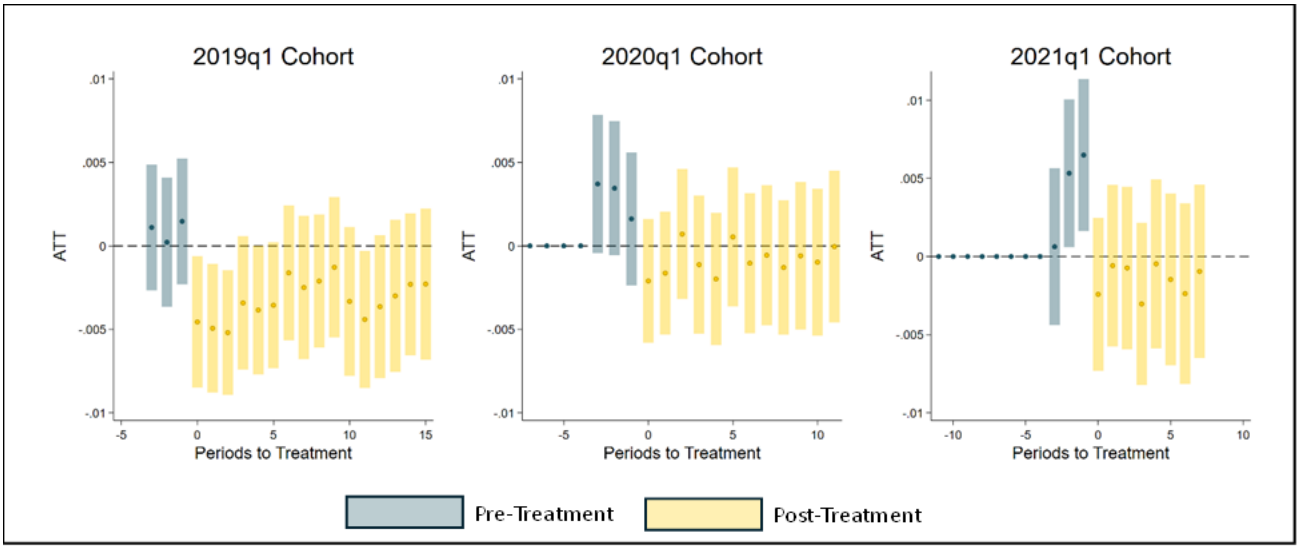


Notes: The bar for each period (year-quarter) depicts the adjusted difference-in-differences (DiD) effect estimate evaluated as the average change in total Medicare Parts A & B expenditure relative to the quarter prior to the program start among the cohort of MDPCP-attributed Medicare beneficiaries minus the concurrent change among their matched comparators, after adjusting for individual fixed effects. DiD effect estimates for quarters prior to the joining cohort's initial participation in MDPCP are in blue bars, while yellow bars are for quarters after the cohort members were first attributed to an MDPCP practice. The dots in each bar mark the level/magnitude of the DiD effect estimates, while the length of each bar spans the 95% confidence interval of the DiD effect estimate. Q1- first quarter of the calendar year.

Inpatient Admissions

Figure 2, below, presents the event-time estimates of the impact of MDPCP on inpatient utilization for the three primary cohorts. As above, each vertical bar in the graphs represents the 95% CI for the difference between the MDPCP and comparison groups in that joining cohort for a single year-quarter and the dots on the bars represents the difference-in-differences point estimate. Blue bars indicate quarters prior to the joining cohort's initial participation in MDPCP, while yellow bars indicate quarters after the cohort members were attributed to an MDPCP practice. The yellow bars can be used to understand the impact of MCPCP on each cohort over time. As with the results for spending, Hilltop documented a consistent and statistically significant negative effect for 2019Q1 joiners starting in the first quarter that individuals were attributed to MDPCP-participating practices. This effect dissipates over time through the 9th quarter since their initial attribution (i.e., 2021Q1). For the 2020Q1 and 2021Q1 joiners, the individual quarter-level effects, while largely negative, are not statistically significant. Similar to the total Medicare spending results, inpatient admissions appear to be trending downward for the 2020Q1 joiners in the quarters immediately prior to joining and trending upward for the 2021Q1 joiners during the same pre-treatment period. However, the weight of the evidence indicates that the treatment effect was not unduly influenced by pre-treatment trends.

Figure 2. Event-Time Results for Any Inpatient Utilization, by Primary Cohort

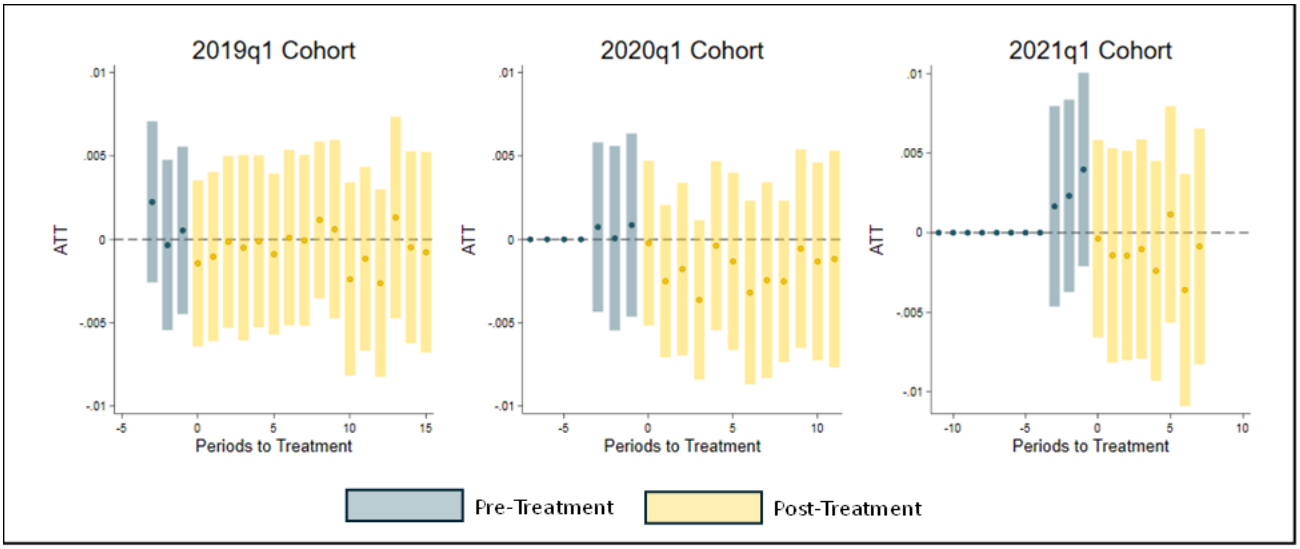


Notes: The bar for each period (year-quarter) depicts the adjusted difference-in-differences (DiD) effect estimate evaluated as the average change in the occurrence of any inpatient service utilization relative to the quarter prior to the program start among the cohort of MDPCP-attributed Medicare beneficiaries minus the concurrent change among their matched comparators, after adjusting for individual fixed effects. DiD effect estimates for quarters prior to the joining cohort's initial participation in MDPCP are in blue bars, while yellow bars are for quarters after the cohort members were first attributed to an MDPCP practice. The dots in each bar mark the level/magnitude of the DiD effect estimates, while the length of each bar spans the 95% confidence interval of the DiD effect estimate. Q1-first quarter of the calendar year.

Emergency Department Utilization

Figure 3, below, presents the event-time estimates of the impact of MDPCP on ED utilization for the three primary cohorts. Each vertical bar in the graphs represents the 95% CI for the difference between the MDPCP and comparison groups in that joining cohort for a single year-quarter and the dots on the bars represents the difference-in-differences point estimate. Blue bars indicate quarters prior to the joining cohort's initial participation in MDPCP, while yellow bars indicate quarters after the cohort members were attributed to an MDPCP practice. The yellow bars can be used to understand the impact of MCPCP on each cohort over time. Hilltop did not find evidence for statistically significant negative quarter-level event-time impacts for any primary cohort, although the pattern of event-time results indicates that 2020Q1 joiners may have experienced a more consistent reduction in ED utilization relative to either 2019Q1 joiners or 2021Q1 joiners.

Figure 3. Event-Time Results for Any ED Utilization, by Primary Cohort

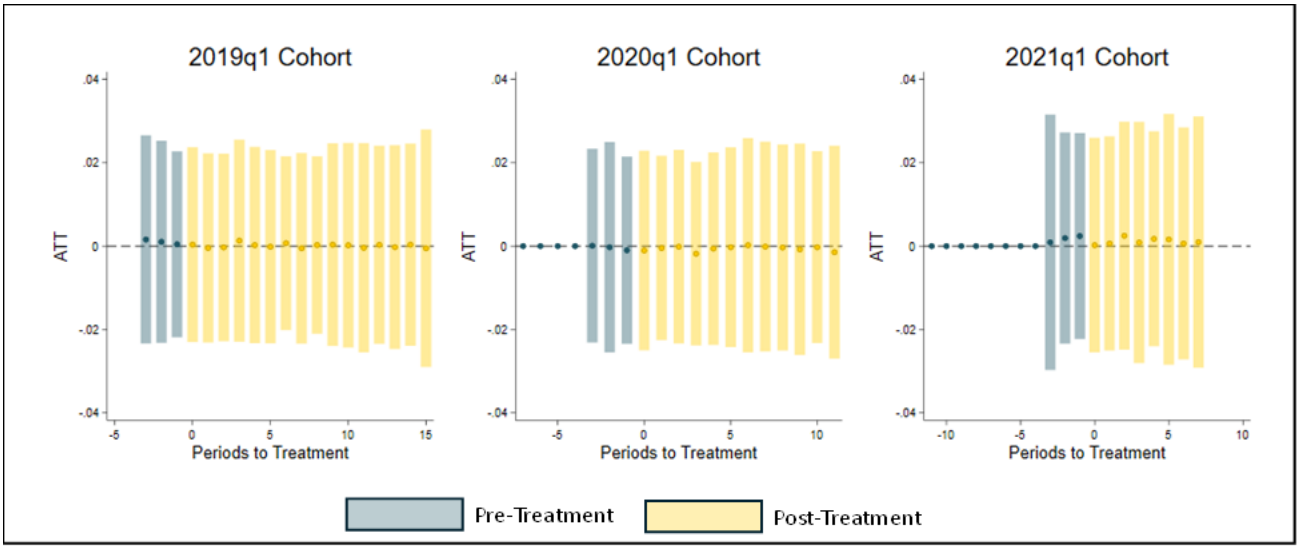


Notes: The bar for each period (year-quarter) depicts the adjusted difference-in-differences (DiD) effect estimate evaluated as the average change in the occurrence of any emergency department service utilization relative to the quarter prior to the program start among the cohort of MDPCP-attributed Medicare beneficiaries minus the concurrent change among their matched comparators, after adjusting for individual fixed effects. DiD effect estimates for quarters prior to the joining cohort's initial participation in MDPCP are in blue bars, while yellow bars are for quarters after the cohort members were first attributed to an MDPCP practice. The dots in each bar mark the level/magnitude of the DiD effect estimates, while the length of each bar spans the 95% confidence interval of the DiD effect estimate. Q1- first quarter of the calendar year.

Avoidable Hospital (AH) Events

Figure 4, below, presents the event-time estimates of the impact of MDPCP on AH events for the three primary cohorts. Each vertical bar in the graphs represents the 95% CI for the difference between the MDPCP and comparison groups in that joining cohort for a single year-quarter and the dots on the bars represents the difference-in-differences point estimate. Blue bars indicate quarters prior to the joining cohort's initial participation in MDPCP, while yellow bars indicate quarters after the cohort members were attributed to an MDPCP practice. The yellow bars can be used to understand the impact of MCPCP on each cohort over time. Hilltop did not find evidence for statistically significant negative quarter-level event-time impacts for any of the primary cohorts. This pattern of results is consistent with the overall, quarter-level results presented in Table 13.

Figure 4. Event-Time Results for Avoidable Hospital Event Incidence, by Primary Cohort



Notes: The bar for each period (year-quarter) depicts the adjusted difference-in-differences (DiD) effect estimate evaluated as the average change in the occurrence of any AH event relative to the quarter prior to the program start among the cohort of MDPCP-attributed Medicare beneficiaries minus the concurrent change among their matched comparators, after adjusting for individual fixed effects. DiD effect estimates for quarters prior to the joining cohort's initial participation in MDPCP are in blue bars, while yellow bars are for quarters after the cohort members were first attributed to an MDPCP practice. The dots in each bar mark the level/magnitude of the DiD effect estimates, while the length of each bar spans the 95% confidence interval of the DiD effect estimate. Q1- first quarter of the calendar year.

Alternative Outcome Specifications & Robustness Checks

For the full cohort difference-in-differences models, we conducted several sensitivity analyses with alternative specifications to assess the robustness of our results. These included:

- Year-level versions of the outcome variables (i.e., total Medicare spending, any inpatient hospitalization, any ED visit, any AH event)[9]
- Operationalization of the inpatient and ED utilization outcomes as continuous variables defined as the number of inpatient hospitalizations (identified using claim counts) or ED visits per quarter, respectively
- Disaggregating the AH event outcome by place of service: AH events in the inpatient setting and AH events in the ED

The pattern of results from the sensitivity analyses mirrored those seen using the original specifications, with the exception of the results quantifying the impact of MDPCP on ED visits. The effect of MDPCP was not statistically significant when using the year-level or the continuous

[9] The difference-in-differences approach used in this evaluation requires all individuals treated at the same time (e.g., 2020) to have the same pre-treatment period (e.g., 2019). Therefore, in this sensitivity check, we only included beneficiaries who had a full year (four quarters) of pre-treatment data (73% of beneficiaries).

versions of the outcome specification. See the Appendix for more detail on these sensitivity analyses, as well as their results.

Hilltop conducted several additional robustness checks and alternative specifications intended to test the sensitivity of the observed results to analytic choices throughout the modeling process. In particular, we used different definitions for, and combinations of, propensity score matching variables; imposed additional restrictions for pre-MDPCP membership in the analytic data set; required that comparison group individuals be attributed to a non-participating primary care practice at the time of match; used as an exact matching variable dual status in the quarter prior to MDPCP attribution, instead of as of the first quarter of attribution; used four quarters of pre-treatment experience as the baseline, instead of the quarter immediately prior to treatment; and used only individuals matched via propensity score, and not the "small group" exact matches. Importantly, none of these iterations substantively altered the pattern of baseline results and are therefore not included in this report.

Results by Subgroup

Hilltop ran six sets of subgroup regressions: four at the individual level and two at the practice level. Our individual-level subgroups are race group (Black, White, other), dual status at baseline, urban/rural location (modal county), and ADI level (for individuals who have ever, and never, lived in a ZIP code in at least the top 20% of disadvantage statewide). Our practice-level subgroups are whether the practice has ever been affiliated with a CTO and whether the practice was owned by a health system as of 2023.

Each subgroup regression was run on a different population. For a given characteristic (for example, practice CTO status), we included treated individuals with that characteristic and their corresponding matched controls. Certain outcomes are not available at baseline, and certain subgroup regressions may have small sample sizes. As such, the results should be interpreted with cautionary attention to these limitations.

Race Group

The overall pattern of results presented in Table 13—that MDPCP leads to a reduction in quarterly total Medicare Parts A & B expenditure and a reduction in inpatient utilization—does not vary significantly across racial groups. For total spending, the point estimates are negative and statistically significant for Black and White beneficiaries. It is important to note that while the point estimate of the impact of MDPCP for Black beneficiaries is larger than for that of White beneficiaries—an average quarterly reduction of $170.84 for the former compared to an average quarterly reduction of $115.29 for the latter—the 95% CIs around these estimates overlap, meaning they are not statistically distinguishable. For inpatient utilization, all three racial groups experienced a statistically significant drop following the introduction of MDPCP. Results for ED utilization and AH event incidence are not statistically significant.

There was no significant effect of MDPCP on total spending for beneficiaries included in the "other" race group. This may be due to the relatively small size of the group (i.e., 8% of total

sample), as well as the fact that the "other" group combines multiple racial and ethnic groups. See Table 15 for full results.

Table 15. Effect of MDPCP by Race Group

Outcome	Quarterly Effect Estimate	95% Confidence Interval	
		Lower	Upper
Race Group: Black			
Total Medicare Parts A & B expenditure	**-$170.84**	**-$241.19**	**-$100.49**
Any inpatient utilization	**-0.0030**	**-0.0049**	**-0.0012**
Any ED utilization	-0.0019	-0.0048	0.0010
Any AH event	0.0003	-0.0011	0.0017
Race Group: White			
Total Medicare Parts A & B expenditure	**-$115.29**	**-$154.84**	**-$75.74**
Any inpatient utilization	**-0.0030**	**-0.0041**	**-0.0019**
Any ED utilization	-0.0011	-0.0026	0.0004
Any AH event	-0.0004	-0.0010	0.0002
Race Group: other†			
Total Medicare Parts A & B expenditure	-$38.46	-$126.66	$49.74
Any inpatient utilization	**-0.0024**	**-0.0047**	**-0.0002**
Any ED utilization	-0.0015	-0.0047	0.0017
Any AH event	-0.0004	-0.0018	0.0010

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

†"Other" includes beneficiaries who are Asian, North American Native, Hispanic/Latinx, Native Hawaiian, Other Pacific Islander, or any other racial or ethnic group (e.g., Middle Eastern, North African).

Medicare-Medicaid Dual Status

Next, Hilltop examined the impact of MDPCP for individuals that were dually eligible for Medicare and Medicaid, relative to those who were not dually eligible during the first quarter of attribution. We found that dually eligible individuals experienced a larger spending reduction upon the onset of MDPCP than non-dually eligible individuals: -$176.58 compared to -$110.14 per beneficiary per quarter on average, although the 95% CIs for both subgroups overlap meaning that the difference is not statistically significant. Dually eligible individuals also experienced a greater reduction in inpatient admissions relative to non-duals: -0.0043 compared to -0.0028, although, again, the difference is not statistically significant. Finally, dually eligible individuals experienced a reduction in ED utilization, while non-duals did not. However, the 95% CIs for the effect of MDPCP on ED utilization for the two groups overlap, meaning that they are not statistically different from each other. Neither group experienced a statistically significant reduction in AH events. Table 16 presents the full results.

Table 16. Effect of MDPCP by Medicare-Medicaid Dual Eligibility Status

Outcome	Quarterly Effect Estimate	95% Confidence Interval	
		Lower	Upper
Dually Eligible at Time of Attribution			
Total Medicare Parts A & B expenditure	**-$176.58**	**-$270.88**	**-$82.29**
Any inpatient utilization	**-0.0043**	**-0.0070**	**-0.0017**
Any ED utilization	**-0.0038**	**-0.0075**	**-0.0001**
Any AH event	0.00002	-0.0018	0.0019
Not Dually Eligible at Time of Attribution			
Total Medicare Parts A & B expenditure	**-$110.14**	**-$143.58**	**-$76.71**
Any inpatient utilization	**-0.0028**	**-0.0037**	**-0.0018**
Any ED utilization	-0.0009	-0.0022	0.0004
Any AH event	-0.0003	-0.0009	0.0003

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

Rural-Urban Status

Hilltop assessed the extent to which MDPCP impact estimates differed by whether a beneficiary's county of residence is considered rural or urban. We operationalized this as whether the modal location over the study period (2019Q1–2022Q4) was urban or rural as defined by the Maryland Department of Health (additional details presented in the Appendix). We found that the general pattern of results for expenditure and inpatient utilization holds for both urban and rural residents and that they do not differ by whether the beneficiary lived in an urban or rural county. However, we found that urban residents experienced a statistically significant reduction in ED utilization, while rural residents did not. It is worth noting that the 95% CIs for the effect of MDPCP on ED utilization overlap for rural and urban residents though, meaning that they are not statistically different from each other. Neither group experienced a statistically significant reduction in AH events. Table 17 presents the full results.

Table 17. Effect of MDPCP by Rural-Urban Status

Outcome	Quarterly Effect Estimate	95% Confidence Interval	
		Lower	Upper
Modal Location is Rural			
Total Medicare Parts A & B expenditure	**-$101.65**	**-$179.06**	**-$24.23**
Any inpatient utilization	**-0.0030**	**-0.0051**	**-0.0009**
Any ED utilization	0.0006	-0.0028	0.0040
Any AH event	-0.0007	-0.0022	0.0008
Modal Location is Urban			
Total Medicare Parts A & B expenditure	**-$122.81**	**-$157.27**	**-$88.36**
Any inpatient utilization	**-0.0030**	**-0.0040**	**-0.0020**
Any ED utilization	**-0.0017**	**-0.0029**	**-0.0004**
Any AH event	-0.0002	-0.0008	0.0003

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

Area Deprivation Index

Hilltop estimated the impact of MDPCP separately for individuals by ADI level. The ADI provides rankings of geographic areas (in this case, ZIP codes) by socioeconomic disadvantage in the state of Maryland that are based on areal-level income, education, employment, and housing quality (University of Wisconsin School of Medicine and Public Health, 2024; Kind & Buckingham, 2018). These data were not available for 2019, so we used the 2020 ADI state rankings for Maryland. We created ADI-based subgroups by partitioning the sample into individuals who had ever, and never, lived in a "high" ADI area. Similar to the methodology used for MDPCP Health Equity Advancement Resource and Transformation (HEART) payments, we defined a ZIP code as being "high ADI" if it ranked in the top quintile of ADI levels among all ZIP codes in Maryland (Maryland Department of Health, 2022). We found that individuals who had ever resided in a high-ADI ZIP code experienced larger spending reductions and inpatient utilization reductions than individuals who had never lived in a high-ADI area; however, they have overlapping 95% CIs, meaning that the impact estimates did not statistically differ. Table 18 presents the full results.

Table 18. Effect of MDPCP by Area Deprivation Index Status

Outcome	Quarterly Effect Estimate	95% Confidence Interval	
		Lower	Upper
Ever Lived in ZIP Code with ADI in Highest Quintile			
Total Medicare Parts A & B expenditure	**-$162.21**	**-$245.07**	**-$79.34**
Any inpatient utilization	**-0.0035**	**-0.0057**	**-0.0012**
Any ED utilization	-0.0023	-0.0055	0.0009
Any AH event	0.0007	-0.0010	0.0023
Never Lived in ZIP Code with ADI in Highest Quintile			
Total Medicare Parts A & B expenditure	**-$110.20**	**-$145.65**	**-$74.76**
Any inpatient utilization	**-0.0029**	**-0.0038**	**-0.0019**

Any ED utilization	-0.0011	-0.0024	0.0002
Any AH event	-0.0005	-0.0011	0.0001

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

CTO Status

Next, Hilltop estimated MDPCP impact estimates for individuals who had ever been attributed to an MDPCP-participating practice that partnered with a CTO, compared with participating beneficiaries in MDPCP practices not affiliated with a CTO. Similar to results for the full sample, both spending and inpatient utilization were reduced in both subgroups, but we did not observe any significant effect of MDPCP on ED utilization or AH event incidence. Further, the 95% CIs in both the CTO and non-CTO subgroups overlapped for all outcomes, meaning that we did not detect systematic differences in impact by CTO status. See Table 19 for full results.

Table 19. Effect of MDPCP by Practice CTO Status

Outcome	Quarterly Effect Estimate	95% Confidence Interval	
		Lower	Upper
Ever Attributed to Practice with CTO			
Total Medicare Parts A & B expenditure	**-$121.03**	**-$156.47**	**-$85.60**
Any inpatient utilization	**-0.0028**	**-0.0037**	**-0.0018**
Any ED utilization	-0.0012	-0.0025	0.0001
Any AH event	-0.0002	-0.0008	0.0003
Never Attributed to Practice with CTO			
Total Medicare Parts A & B expenditure	**-$113.00**	**-$174.69**	**-$51.31**
Any inpatient utilization	**-0.0039**	**-0.0057**	**-0.0020**
Any ED utilization	-0.0017	-0.004	0.0005
Any AH event	-0.0005	-0.0016	0.0004

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level. This analysis includes 441,707 individuals attributed to MDPCP practices that had ever partnered with a CTO, with matched comparators, and 105,683 individuals attributed to MDPCP practices that had never partnered with a CTO, with matched comparators.

Practice Ownership Status

Finally, we estimated MDPCP impact estimates by practice ownership status. Due to data availability limitations, we were only able to use practice ownership status as of 2023, which falls outside the study period. This has two consequences: first, certain practices in the analytic data set no longer participate in MDPCP as of 2023, meaning that we were unable to use the individuals attributed to these practices in this sub-analysis. Second, practice ownership can and does change over time; practices owned by a health system in 2023 are not necessarily owned by a health system at other points in the study period. To the extent that this is correlated with

practice ownership during the study period, however, our results may be suggestive of the true underlying impact of practice ownership.

As above, the pattern of results is identical for practices owned by a health system in 2023 relative to practices not owned by a health system: total spending and inpatient utilization both fell and are statistically significant. Further, the 95% CIs for both groups on these outcomes overlap, meaning that they are not statistically different. ED utilization and AH events fell but are not statistically different from either zero or each other. Table 20 presents the full results.

Table 20. Effect of MDPCP by Practice Ownership Status

Outcome	Quarterly Effect Estimate	95% Confidence Interval	
		Lower	Upper
Practice owned by health system (as of 2023)			
Total Medicare Parts A & B expenditure	**-$132.19**	**-$176.55**	**-$87.83**
Any inpatient utilization	**-0.0029**	**-0.0040**	**-0.0017**
Any ED utilization	-0.0014	-0.0030	0.0002
Any AH event	-0.0003	-0.0010	0.0004
Practice not owned by health system (as of 2023)			
Total Medicare Parts A & B expenditure	**-$120.23**	**-$163.07**	**-$77.39**
Any inpatient utilization	**-0.0033**	**-0.0045**	**-0.0021**
Any ED utilization	-0.0009	-0.0025	0.0006
Any AH event	-0.0004	-0.0011	0.0003

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level. This analysis includes 265,479 individuals attributed to MDPCP practices that were owned by a health system as of 2023, with matched comparators, and 248,853 individuals attributed to MDPCP practices that were not owned by a health system as of 2023, with matched comparators. This analysis excludes 33,058 individuals attributed to MDPCP practices for which practice ownership information as of 2023 was unavailable, with matched comparators.

COVID-19 Impact

Research indicates that MDPCP practices responded to COVID-19 in such a way that MDPCP-attributed beneficiaries experienced fewer COVID-19-related infections, hospitalizations, and deaths, and experienced greater use of telehealth services. While the swift response by MDPCP practices is laudable, it lends interpretational nuance to an evaluation of MDPCP: any effect of MDPCP may incorporate not just the difference in patient-level experience between MDPCP and non-MDPCP practices, but also the differential effects of COVID-19 response. While the use of the difference-in-differences methodology should account for common shocks experienced by both the treatment and comparison group, a differential response by MDPCP practices implies that the MDPCP effects could have changed during the pandemic.

In order to understand the impact of the COVID-19 pandemic, we estimated whether the effect of MDPCP on the four main outcomes differed during the COVID-19 period (2020Q2–2021Q4) relative to the pre-COVID-19 period (2019Q1–2020Q1). See Table 21 for results.

Table 21. Effect of MDPCP by Calendar Quarter, Pooled by COVID-19 Status

Outcome	Pre-COVID *2019Q1 – 2020Q1*	During COVID *2020Q2 – 2021Q4*	Difference (95% CI)
Total Medicare Parts A & B expenditure	**-$162.58**	**-$99.25**	**$63.32 ($15.96, $107.79)**
Any inpatient utilization	**-0.0042**	**-0.0023**	**0.0019 (0.0010, 0.0028)**
Any ED utilization	-0.0007	-0.0013	-0.0006 (-0.0015,0.0002)
Any AH event	0.0001	-0.0002	-0.0003 (-0.0010,0.0003)

Notes: The adjusted effect estimates presented are average changes in the measure level (expenditure) or percentage point change in the probability of occurrence of the event (utilization outcomes) per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. Results are stratified temporally into pre-pandemic and intra-pandemic periods by occurrence of the COVID-19 pandemic operationalized as beginning during the second quarter of 2020 (April 2020). **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

We found that the negative spending effect of MDPCP was moderated during the COVID-19 period, falling from an average quarterly effect of -$162.58 during the pre-pandemic period to an average quarterly effect of -$99.25 during the COVID-19 period. We saw similar results for inpatient utilization: the MDPCP effect was larger in magnitude in the pre-COVID-19 period relative to during COVID-19 (-0.0042 compared to -0.0023). Notably, the difference between the pre-pandemic and intra-pandemic effect estimates was statistically significant for both outcomes.

We documented the opposite impact for ED utilization: the small negative MDPCP effect prior to the pandemic almost doubled in magnitude during the pandemic. Finally, we estimated that the AH effect marginally increased in magnitude; however, neither of these differences was statistically significant.

These findings are broadly consistent with the research documented above on the impact of COVID-19 on MDPCP practices. In principle, two effects are possible: through effective outreach and messaging, it is possible that MDPCP practices might have successfully deterred unneeded utilization, which would, all else equal, have increased the (negative) magnitudes of the MDPCP effect during the COVID-19 pandemic. However, it is also possible that MDPCP practices successfully maintained beneficiaries' connection with the health system—for example, through the use of telehealth as documented in Gruber et al., 2023—which would, all else equal, tend to reduce the (negative) magnitudes of the MDPCP effect. These results indicate that the latter effect dominates the former: relative to the 2019Q1–2020Q1 period, beneficiaries attributed to MDPCP-participating practices incurred *relatively* more spending and inpatient utilization in the 2020Q2–2021Q4 period. However, it is important to note that we continue to estimate a negative MDPCP program effect during this period for total spending and inpatient utilization, albeit smaller in magnitude than in the 2019Q1–2020Q1 period.

Section 6. Limitations

This analysis has limitations that should be considered when interpreting the results. First, as noted above, selection into MDPCP is non-random: that is, practices that believe that they will benefit from MDPCP may be the first to join, and practices that do not anticipate benefits under MDPCP may elect not to join. This does not threaten the validity of the results of this evaluation but should be considered when interpreting the results: it is not the case that if all primary care practices across Maryland participated in MDPCP their attributed beneficiaries would experience effects similar to the practices that voluntarily joined. This is also suggested by the heterogeneity of impact estimates by joining cohort: the reductions in spending and inpatient hospital utilization are strongest for the 2019Q1 joiners (who were in the first wave of joining practices). The later primary cohorts, however, experienced smaller, statistically insignificant effects on spending and utilization. Thus, extrapolating effects more broadly should be conducted with caution.

Second, it is never possible to confirm that the comparison group is a true counterfactual for the treatment group: that is, that the comparison group is a valid approximation of what the experience of the treatment group would have been had they not been treated. Any evaluation using difference-in-differences must acknowledge the possibility that non-parallel counterfactual trends could contribute to the results, independent of the treatment effect. For this evaluation, we employed several techniques to minimize this risk: we used a within-Maryland research design, meaning that the effects of the Maryland TCOC Model would not be conflated with the effect of MDPCP; our comparison pool consisted of individuals who were attributed to a non-participating primary care provider at least once during the study period using the same attribution algorithm used for MDPCP, meaning that our program estimates should not be picking up a "primary care" effect; we created a matched comparison group that has strong balance based on observable characteristics; and we demonstrated that, overall, there is no evidence of a violation in the parallel trends assumption using statistical tests for differential linear trends in the outcomes for the four quarters leading up to the start of MDPCP, nor is there evidence of pre-treatment effects.

Additionally, by using variation in individual-level exposure to MDPCP, rather than the timing of practice-level participation, we exploited two sources of variation: the timing of when individuals are attributed to practices, and the timing of when practices join MDPCP. To the extent that individuals do not self-select into MDPCP practices on the basis of anticipated spending or utilization reductions, this may leverage relatively more exogenous variation in MDPCP exposure than using variation in practice exposure alone. Additionally, for the 2019 joining practices, applications for MDPCP were accepted from August 1, 2018–August 31, 2018;[10] thus, it is unlikely that practices differentially selected into MDPCP on the basis of anticipated beneficiary spending and outcomes starting in 2019Q1. Finally, to the extent that individual exposure to

[10] See graphic depiction of the MDPCP Timeline: https://health.maryland.gov/mdpcp/Documents/MDPCP%20Timeline_for%20web.pdf

MDPCP is driven by practice-level decision to join MDPCP (or not), the comparison group should not be endogenously sorting out of MDPCP over time.

Third, the COVID-19 pandemic occurred during the study period and likely influenced the study results. Research has shown that MDPCP practices systematically differed from non-MDPCP practices in their treatment of COVID-19 (Gruber et al., 2023; Perman et al., 2021). While the use of difference-in-differences should control for common shocks across the treatment and comparison groups, it is possible that differences in outcomes may reflect a differential response to COVID-19 by MDPCP and non-MDPCP practices, as well as an underlying true effect. We examined the extent to which the observed MDPCP effects differ during the COVID-19 period and found that, while the negative spending and inpatient utilization effects were moderated during the COVID-19 period relative to the pre-COVID-19 period, these differences were statistically significant. Thus, it is possible that our overall effects understate the magnitude of the true spending and inpatient utilization impacts of MDPCP because they necessarily include the differential response.

Fourth, Hilltop did not use Medicare standardized hospital spending in this analysis because standardized payment amounts from CMS[11] were not available for use in this evaluation. The Maryland health care landscape is unique among states in that hospitals are regulated though a global budgeting system: that is, hospital revenue is prospectively set each year, allowing for certain adjustments; as a result, hospitals are incentivized to reduce unnecessary utilization in order to retain revenue as net income. As a consequence, the underlying "price" of a procedure at a given hospital can change throughout the course of the year. If a hospital is running under its global budget, then it has the latitude to increase its unit prices; if a hospital is running over its global budget, then it is required to reduce its unit prices. This could introduce bias into this evaluation if MDPCP-attributed beneficiaries disproportionally received services at hospitals for which prices increased (or decreased), relative to the comparison population. Our matching strategy is designed to mitigate this. By requiring exact match on ZIP3 in the first quarter of attribution, each treated individual was paired with a geographically proximal comparator. To the extent that geographically proximal individuals are in the same hospital catchment areas, then the risk of bias due to systematic differentials in hospital pricing should be minimal. Additionally, Hilltop derived individual-level spending from CCLF claims, which, for selected Part B services, mechanically reflect lower reimbursement rates due to the structure of the CPCP. While this likely artificially inflates the magnitude of the estimated person-level expenditure savings, we accounted for this in the calculation of the overall net program costs by including the CPCP prospective payments in MDPCP's operational costs.

Section 7. Discussion

This report describes Hilltop's evaluation of the impact of MDPCP on total Medicare spending, inpatient admissions, ED visits, and AH events. We leveraged an advanced difference-in-difference approach optimized for programs with staggered start times to model the effect of

[11] See overview of CMS Payment Standardization: https://resdac.org/articles/cms-payment-standardization-overview

the program on Medicare FFS beneficiaries that were attributed to MDPCP practices from 2019 to 2022 relative to a matched comparison group. In addition to examining the effect of MDPCP on the full cohort, we estimated whether the effect of MDPCP differed by beneficiary (i.e., race, dual status, geography) or practice (i.e., practice ownership, use of a CTO) characteristics, as well as whether the COVID-19 pandemic moderated any program effects.

This section presents a summary of the results from this evaluation, analyzes their internal consistency, presents net cost estimates of MDPCP, reconciles our results with those of other studies, and discusses other possible unmeasured costs and benefits of MDPCP.

Summary of Results

Overall, Hilltop found that the introduction of MDPCP was associated with beneficiary-level significant savings and reductions in both inpatient utilization and moderate reductions in ED utilization. Moreover, given that these impact estimates are based on a within-Maryland treatment and comparison group, they should be interpreted as occurring over and above the additional components of the TCOC.

Specifically, this analysis found that the introduction of MDPCP led to a 4.33% reduction in total spending relative to baseline, or a -$119.60 reduction per person per quarter on average. Annualized, this translates to a reduction of $424.68, accounting for the fact that not all MDPCP-attributed beneficiaries were continuously attributed for all quarters of a calendar year.[12] Analysis of the three largest joining cohorts indicates that the negative spending effect was largest for the 2019Q1 cohort (-$123.15 per quarter), and smaller and statistically insignificant for the 2020Q1 joining cohort (-$29.10 per quarter) and the 2021Q1 joining cohort (-$32.27 per quarter). Further event-study analysis indicated that, for the 2019Q1 joiners, the spending reduction was apparent immediately, although the magnitude of effects fell during the COVID-19 pandemic. Subgroup analysis indicated that the spending effects appear to be larger for Black beneficiaries, individuals that were dually eligible for Medicare and Medicaid as of the first quarter of attribution, and individuals who ever lived in an area with high values of ADI. While these are not statistically different from the overall impact estimates, they suggest that MDPCP is impacting traditionally underserved groups and, in doing so, advancing health equity in Maryland.

Hilltop also documented a significant reduction in inpatient utilization. Following introduction of MDPCP, individuals attributed to MDPCP-participating practices experienced a 7.18% reduction in inpatient utilization relative to baseline, net of effects for comparable matched individuals. As with the results for expenditure, this result is largest for the 2019Q1 joiners but smaller and insignificant for the 2020Q1 and 2021Q1 joiners. The results began immediately for the 2019Q1 cohort, with some moderation during the COVID-19 pandemic, and the effect was most

[12] In order to annualize the average quarterly impact estimate, we calculated the average number of quarters per year that the MDPCP treatment group is in the analytic sample in the post-treatment period and scale the quarterly estimate by this amount (3.55 quarters).

pronounced for individuals that were dually eligible at the time of attribution (although this subgroup effect does not statistically differ from the overall impact estimate).

Unsurprisingly, this evaluation noted smaller impacts of MDPCP on ED utilization. By definition, individuals in the treatment and comparison groups are connected to primary care practices, which should help mitigate unnecessary usage of EDs. As a consequence, there may have been limited scope for MDPCP to reduce non-emergent ED utilization.

Finally, this evaluation found no statistical evidence that MDPCP reduced the incidence of avoidable hospital events. This may be a function of the rarity of this outcome, which, in our data, are roughly 1/5th as prevalent as ED utilization and 1/3rd as prevalent as inpatient utilization. We performed a subgroup analysis to test whether MDPCP had a meaningful impact on AH events by source—either ED or inpatient—and found no discernible impact. We documented moderately stronger impacts using an alternative specification focusing on the annual level, which suggests that this event may be too rare to detect any impacts using the current methodology. Additionally, we noted that this lack of detected effect may reflect the nature of this outcome: recent commentary calls into question the validity of the AHRQ definition of "avoidable hospital event" as a quality indicator for ambulatory and inpatient care (Berenson & Skopec, 2024).

Internal Consistency

When considered together, there is internal consistency among the results. Overall, we found that inpatient utilization fell by 0.0039 inpatient claims (i.e., admissions) per person per quarter on average and that total Medicare spending fell by $119.60 per quarter, or 4.33% relative to baseline. We estimated that, as of 2020, the average Medicare cost per inpatient stay in Maryland was $19,956.80.[13] Taking the point estimates at face value, we estimated the spending reduction due to the reduction in inpatient stays as 0.0039 * $19,956.80 = $77.83 per quarter. That is, the reduction in inpatient utilization accounts for almost two thirds of the overall reduction in expenditure.

There are other potential mechanisms that may drive the spending result. We documented a reduction in ED utilization for which we do not account in this exercise, and individuals with inpatient utilization may also experience a reduction in the intensity of hospital care, which could lead to cost savings. Additionally, the structure of the CPCP implies that reimbursement amounts on Part B claims for SPCS are artificially lower than standard FFS rates for MDPCP-attributed beneficiaries in Track 2 practices due to the structure of the CPCP. This may result in an inflated estimate of marginal or person-level expenditure impacts. However, this evaluation accounts for the issue by adding prospective CPCP payments to MDPCP operational costs in calculating the

[13] We calculated this using two sources. First, we linearly trended HCUP data from 2009-2017 on average cost per inpatient stay for traditional Medicare beneficiaries (https://hcup-us.ahrq.gov/reports/statbriefs/sb262-Medicare-Advantage-Costs-2009-2017.pdf). We then inflated these projections by 30% to account for the fact that, in Maryland, Medicare reimbursements are higher than other states due to the Total Cost of Care Model (https://www.crisphealth.org/wp-content/uploads/2023/03/MD-Model-Analytics-Inpatient-spending-2023-03-10.pdf, page 23).

program's aggregate net cost estimates. Taken together, we believe that these results are internally consistent.

Net Cost Estimates

The results from the baseline specification are in terms of person-quarter: that is, we found that the introduction of MDPCP led to an average reduction of $119.60 in total spending per quarter per person. We created net cost estimates by scaling the point estimate for spending by the total number of person-quarters in the analytic data set following the start of treatment (N = 5,622,377) and then incorporating the program's operational costs. We based these operational costs on the net program payments from 2019-2022, which include all unrecouped CMF, PBIP, CPCP, and HEART payments. The inclusion of CPCPs in total MDPCP operational costs is intended to correct for any artificial inflation in the magnitude of person-level expenditure estimates due to the reduced fee schedule Medicare applies to FFS reimbursements for SPCS in Track 2 practices. See Table 22 for details.

Table 22. Estimated Medicare Expenditure Savings and Operational Costs during the First Four Year of Implementation of the Maryland Primary Care Program, 2019–2022

Outcome	Aggregate Effect	95% Confidence Interval (Aggregate)	
		Lower	Upper
Expenditure	-$672.4 million	-$852.7 million	-$492.2 million
Program Cost	$510.5 million	$510.5 million	$510.5 million
Net Program Cost	**-$161.9 million**	**-$342.2 million**	**$18.3 million**

Notes: Net program cost is calculated as the difference between the total aggregate savings on Medicare Parts A & B expenditure among participating beneficiaries, and the total of the program's payments (net of recoupments) to participating providers in primary care practices and care transformation organizations for per member per month care delivery fees and incentives over the four-year period. The aggregate program estimate is calculated as the overall impact estimate multiplied by the total number of person-quarters in the treatment period (N = 5,622,377). A positive value indicates that program operational costs exceeded estimated savings, while a negative value indicates that estimated savings on Medicare expenditure were greater than the program's payments to participating providers.

Hilltop found that, based on the point estimate of our baseline specification, MDPCP results in a net cost savings of -$161.9 million across the years 2019–2022, but with a 95% confidence interval from -$342.2 million to $18.3 million. While the point estimate suggests that MDPCP led to overall net cost savings, we were unable to rule out that MDPCP did not lead to small net total costs. Taken together, these results imply that MDPCP was revenue-neutral at the program-level, although with suggestive evidence of net cost savings.

Due to the voluntary enrollment of practices in MDPCP, we were unable to extrapolate our impact estimates to simulate the cost savings if all Medicare FFS enrollees were attributed to MDPCP practices. In other words, it is possible that practices that have not joined MDPCP have decided not to participate due to minimal anticipated benefits or challenges with meeting

program requirements; thus, even if the program were spread across the state, "additional" practices may experience minimal benefits. Our evaluation documented some support for this: the positive effects are concentrated among the first (2019Q1) joining cohort. The 2020Q1 and 2021Q1 joining cohorts experienced small reductions in total spending and inpatient utilization, but not statistically differentiated from zero.

Reconciliation of Results with Existing Studies

Previous studies of the effect of MDPCP have tended to find modest savings that do not outweigh the direct program costs. This section of our report situates the findings from our evaluation in the context of the discussion in the "Other Studies" subsection of Section 1.

MDPCP Performance Report

The headline results from this evaluation accord with the results presented in the January 2023 MDPCP performance report. That report found that, relative to matched comparison group of equivalent, non-participating population, the MDPCP population experienced a 4.3 percentage point spending reduction from 2019–2022; a 3.0 percentage point reduction in inpatient utilization; a 0.6 percentage point reduction in ED utilization; and 0.0 percentage point reduction in PQI-like events.[14] In contrast, this evaluation finds a 4.33% reduction in total spending, a 7.18% reduction in inpatient utilization, and a smaller, 1.70% reduction in ED utilization, all of which are statistically significant. We estimated a small and statistically insignificant reduction in the rate of AH events.

Since this MDPCP report does not present confidence intervals for its estimates, we were unable to assess the extent to which we are able to reject the hypothesis that these quantities are statistically distinguishable; however, we interpret this pattern of results as broadly consistent with the results presented in this evaluation.

Joint Chairmen's Reports

The HSCRC program evaluation from 2023 documents that the program led to $114.1 million in hospital expenditure savings but cost $198.6 million, for a net cost of $84.5 million in 2022. It is important to acknowledge two missing pieces of information: this does not indicate the effect of MDPCP on total costs (both hospital and non-hospital) and does not provide confidence intervals. Thus, it is possible that these results might be statistically indistinguishable from the results presented in this study. Additionally, the 2023 report includes gross program costs, rather than program costs net of recoupments as are included in this evaluation.

TCOC Evaluation – Overall

The recent TCOC progress report found that, relative to the 2017-2018 period, the Maryland Medicare FFS population experienced a 1.0 percentage point reduction in total spending and a

[14] Definition of PQI-like events in the MDPCP Performance Report is similar to the AH event definition used in this evaluation.

5.6 percentage point reduction in all-cause acute care admissions from 2019-2022 (Peterson et al., 2024a). Given that MDPCP is the second largest component of the TCOC and began in 2019, it is possible that some of the additional aggregate savings accrued from 2019-2022 can be attributed to MDPCP.

TCOC Evaluation – MDPCP Sub-Analysis

The recent TCOC progress report documented savings of approximately $6 million for individuals who joined 2019 practices, for a net cost of approximately $90 million (Peterson et al., 2024a). As noted above in Section 1, there are three reasons this result likely differs from the results provided in this evaluation: first, the authors of the TCOC evaluation created their own attribution that only partially overlaps with the official payment-based attribution. Second, the study examined only practices that joined MDPCP in 2019 using yearly cross-sections of beneficiaries. Third, the study used practice-level fixed effects to control for practice-specific, time-invariant factors; this effectively uses within-practice variation over time (i.e., change in average spending at the practice level) to identify the program effect. That is, this methodology compares people within each practice and calculates the program effect based on average outcomes across practice participants over time.

Crucially, this empirical strategy allows for changing practice composition to affect the estimated effects. For example, suppose that MDPCP practices, by virtue of their advanced primary care, retain high-cost enrollees at a higher rate than non-MDPCP practices. Thus, over time, average practice-level spending may appear to rise for MDPCP practices due to the changing composition (consisting of relatively more high-cost individuals) relative to non-MDPCP practices. This, in turn, would be interpreted as a positive effect of MDPCP on spending.

The use of individual fixed effects, however, means that this evaluation identifies the MDPCP impact by comparing within-person changes in outcomes over time between individuals in the treatment and comparison groups, before and after the start of MDPCP. That is, all time-invariant, unobserved characteristics of individuals are effectively accounted for, leaving only within-person change in outcomes over time, and not between-person comparison of outcomes. Essentially, the use of individual fixed effects builds the impact estimate from the individual upward rather than from the practice downward. This, in turn, isolates the individual-level impact of MDPCP, independent of any practice-level impacts.

A simple example illustrates this point. Consider two practices, A and B, each with two patients: A.1 and A.2, and B.1 and B.2. Assume that practice A is in MDPCP and that practice B is not in MDPCP. Suppose, further, that there is no causal impact of the introduction of MDPCP on spending, and that spending is flat for each individual. Individual A.1 incurs $1,000 per period, and individual A.2 incurs $10,000 per period. Individual B.1 incurs $2,000 per period, and individual B.2 incurs $12,000 per period. Finally, suppose there are two periods: pre-MDPCP and post-MDPCP.

First, assume no differing sample composition: that is, all four individuals remain in the analytic sample for the entire study period. The average spending per person for Practice A before and

after MDPCP is $5,500; the average spending for practice B before and after MDPCP is $7,000. The change in spending, at the practice level, is 0 for both practices A and B, and, thus, the estimated MDPCP practice-level effect is 0. Similarly, the change in spending for each person within practice A and B is also 0, indicating that the estimated individual-level effect is also 0.

Next, suppose that individual B.2 dies shortly after MDPCP was implemented and thus is not in the analytic sample for the "post-MDPCP" period for practice B. This change in practice composition changes the average practice-level spending: now, the average practice-level spending for practice B is $7,000 before MDPCP and $2,000 after MDPCP (since this is driven only by individual B.1). The average practice-level spending for Practice A is still $5,500 per period. The new practice-level effect of MDPCP is now an increase of $5,000: average practice-level spending did not change for the MDPCP practice, but fell by $5,000 for the non-MDPCP practice, indicating that, relative to the non-MDPCP practice, spending rose for the MDPCP practice by $5,000.

However, the individual-level effect is still 0. This effect is now driven by only three individuals: A.1, A.2, and B.1, who all experience 0 change in their spending. Individual B.2 does not contribute to the estimate because they were not in the analytic data set for both periods. This example, while simple, illustrates a more fundamental fact: the practice-level impact of MDPCP on outcomes is not necessarily identical to the individual-level impact of MDPCP on those same outcomes. Differentially changing sample composition for the treatment and comparison group can impact results, especially to the extent that these changes occur within individuals with high values of outcomes (for example, expenditure).

This methodological difference can potentially explain the differences in MDPCP impact estimates between the current evaluation and those conducted as part of the TCOC progress report. Hilltop conducted two checks in order to assess the possibility of differential sample composition over time. First, we conducted additional balance checks to assess the balance of our treatment and comparison groups on condition history. We found that the groups are well-balanced on eight selected conditions. See additional results in the Appendix (Table A.1). Second, we assessed the distribution of date of death in the analytic data set overall, and for individuals in the 2019Q1 joining cohort. Given that the treatment and comparison groups are well-balanced on conditions, individuals in the MDPCP cohort exiting the analytic sample due to death less often than individuals in the comparison group could potentially indicate differing retention due to the advanced primary care at MDPCP-participating practices. This, in turn, could imply that average practice-level spending may rise over time for MDPCP practices due to a rising proportion of high-cost beneficiaries in the practice panel. Table 23 presents the results of this analysis.

Table 23. Fraction of Cohort with a Death Date in the Study Window

Population	% Died
Entire Cohort	
Ever MDPCP	11.85%
Never MDPCP	14.62%
2019 Q1 Joiners	
Ever MDPCP	17.34%
Never MDPCP	21.49%

Note: This table shows the fraction of the study cohort that died within the study window following the start of MDPCP (i.e., 2019Q1 through 2022Q4) by treated status, both overall and for 2019Q1 joiners.

The Hilltop team found that, across all joining cohorts, 11.85% of individuals who are ever attributed to MDPCP died at some point during the study period (2019Q1 – 2022Q4), while 14.62% of the comparison population died. This difference is larger for the 2019Q1 joining cohort: 17.34% of MDPCP-attributed beneficiaries died during the study period (2019Q1 – 2022Q4), compared to 21.49% of comparison individuals from that cohort.

Taken together—the strong balance of the treatment and comparison groups on baseline condition history, and the differential rate at which beneficiaries exit the analytical sample due to death—we believe that changing practice composition over time may be generating the differences between the results from the federally funded TCOC sub-analysis of MDPCP and the results presented in this evaluation. Both methodological approaches—building impact estimates from the individual upward as well as from the practice downward—are methodologically valid but seek answers to slightly different questions.

Other Unmeasured Costs and Benefits of MDPCP

It is possible that MDPCP affected other populations. Three such effects are possible. First, operational changes that practices implemented as a result of MDPCP for their Medicare FFS beneficiaries may have led to within-practice spillover effects for beneficiaries of other payers. It is possible that Medicaid- or commercially insured beneficiaries may have experienced reductions in spending or utilization to the extent that practice-level initiatives caused by MCPCP did not differ for patients by payer.

Second, it is possible that MDPCP may have cross-practice spillover effects. In the health care space, competition plays a key role in driving innovation between practices (Rivers & Glover, 2008). If practices enrolled in MDPCP offer enhanced primary care services (e.g., care management, social needs referrals), non-MDPCP practices may start providing similar advanced primary care services to retain their patients.

Finally, there might be long-term effects of MDPCP that will only become apparent with the passage of time. To the extent that MDPCP practices have undertaken structural changes in patient management, patterns of care, or built other long-term capacity, these changes might

accrue benefits for practices and patients in the future. Interventions for the young-old can likely improve the health of those beneficiaries when old-old, but these effects will necessarily require years to manifest.

One such potential mechanism for long-term effects is the recent announcement of the continuation of Maryland's innovative health services delivery landscape. The States Advancing All-Payer Health Equity Approaches and Development (AHEAD) Model builds on the TCOC Model and other successful accountable care programs in Pennsylvania and Vermont with additional emphases on screening and referrals of patients to community resources that address social drivers of health, as well as a focus on strategies to reduce disparities in health outcomes (Centers for Medicare & Medicaid Services, 2024b). Maryland was recently selected among the first cohort of states to participate in CMMI's AHEAD model launching in January 2026 (Raths, 2024). As with the TCOC, states participating in AHEAD will be accountable for limiting the growth of total health care spending by public and private payers. The primary care component of the AHEAD model will involve increasing the proportion of health care spending allocated to primary care at the statewide level, and supporting practices that choose to participate with CMFs for advanced primary care functions (Burns, 2024).

Conclusion

In this evaluation, Hilltop used matching and difference-in-differences analyses to assess the effect of the introduction of MDPCP as the change in outcomes for individuals attributed to MDPCP practices, net of changes in comparable individuals not attributed to MDPCP practices. Overall, we found that the introduction of MDPCP was associated with significant beneficiary-level savings and reductions in both inpatient utilization and moderate reductions in ED utilization and no discernible impact on AH events. In a sub-analysis, we found that this overall result appears to be largest among 2019Q1 joiners, with smaller, statistically insignificant effects for 2020Q1 and 2021Q1 joiners. Subgroup analyses indicated that the spending effects appear to be larger for Black beneficiaries, individuals that were dually eligible as of the first quarter of attribution, and individuals that ever lived in an area with high values of ADI. While these estimates are not statistically differentiable across groups, they suggest that MDPCP is impacting traditionally underserved groups and, in so doing, advancing health equity in Maryland.

Scaling the baseline specification point estimates, we were unable to reject the hypothesis that MDPCP was revenue neutral. The program incurred $510.5 million in operating costs from 2019-2022 and led to a total savings of $672.4 million, with a 95% confidence interval for the savings from $852.7 million to $492.2 million. Accounting for the program operating costs yields net savings of $161.9 million, although with a 95% confidence interval of $342.2 million in net savings to $18.3 million in net costs. That is, the balance of the evidence indicates that MDPCP was likely revenue neutral at the program-level, although with suggestive evidence of net costs savings.

Conducting this evaluation entailed addressing several challenges. MDPCP is a voluntary program, meaning that participating practices likely differ from non-participating practices in

non-random ways; the staggered rollout of MDPCP required the use of novel estimation techniques to isolate the relevant variation for use in the differences-in-differences specification; COVID-19 occurred starting in 2020Q2, leaving only five quarters of pre-COVID experience; and limited data availability regarding comparison practices led us to adopt an estimation strategy that focused on estimating the effect of MDPCP on the basis of within-individual changes in outcomes over time.

Our pattern of results largely aligns with other evaluations of MDPCP. We documented significant reductions for spending and inpatient utilization, smaller reductions for ED utilization, and no statistical evidence of reductions for AH events. Notably, a recent sub-analysis in the TCOC Progress Report found minimal savings as a result of MDPCP (Peterson et al., 2024a). We believe that these differing results are due to a matter of perspective. This evaluation estimates the individual-level impact on outcomes, whereas other evaluations focus on the practice level. To the extent that practice composition changes over time, a practice-downward impact estimate may differ from an individual-upward estimate. Both perspectives are valid; they simply ask slightly different questions.

These results are especially germane given the recent announcement that Maryland was recently selected among the first cohort of states to participate in CMMI's AHEAD Model launching in January 2026 (Raths, 2024). The AHEAD Model aims to curb the growth of health care expenditure and improve population health outcomes by increasing investments in primary care and providing financial stability for hospitals (Centers for Medicare & Medicaid Services, 2024b). Specifically for primary care, the AHEAD Model promotes improved care management, behavioral health integration, and a focus on health-related social needs of patients. Given the similarities in goals and strategies between the TCOC and AHEAD Models, this evaluation of MDPCP demonstrating reductions in expenditure and inpatient utilization suggests that Maryland is well-positioned to capitalize on the opportunity presented by the AHEAD Model to continue to implement innovative health services programming that reduce overall cost growth and improve population health.

References

Agency for Healthcare Research and Quality. (2022). *AHRQ Quality Indicators™ Prevention Quality Indicators* (15-M053-3-EF).

Barosso, G. (2015). *How to Identify hospital claims for emergency room visits in the Medicare claims data*. ResDAC. https://resdac.org/articles/how-identify-hospital-claims-emergency-room-visits-medicare-claims-data

Berenson, R. A., & Skopec, L. (2024, June 3). How preventable hospitalizations became a widely used but flawed quality measure. *Health Affairs Forefront*. https://doi.org/10.1377/forefront.20240531.688339

Burns, A. (2024). *What is the Centers for Medicare and Medicaid Services' new AHEAD model?* [Issue Brief]. Kaiser Family Foundation. https://www.kff.org/affordable-care-act/issue-brief/what-is-the-centers-for-medicare-and-medicaid-services-new-ahead-model/

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Center for Medicare and Medicaid Innovation. (2019). *Maryland Total Cost of Care Model Maryland Primary Care Program request for applications, version 2.0*. https://www.cms.gov/priorities/innovation/files/x/mdtcocm-rfa.pdf

Center for Medicare and Medicaid Innovation. (2023). *Maryland Total Cost of Care Model Maryland Primary Care Program Request for Applications, version 5.0*. https://www.cms.gov/priorities/innovation/media/document/mdpcp-rfa-yr6

Centers for Medicare & Medicaid Services. (2024a). *Maryland Total Cost of Care Model*. [Model Summary]. https://www.cms.gov/priorities/innovation/innovation-models/md-tccm

Centers for Medicare & Medicaid Services. (2024b). *States Advancing All-Payer Health Equity Approaches and Development (AHEAD) Model*. [Model Summary]. https://www.cms.gov/priorities/innovation/innovation-models/ahead

Centers for Medicare & Medicaid Services. (n.d.-a). *Maryland Primary Care Program Track 3 Fact Sheet*. https://www.cms.gov/priorities/innovation/media/document/mdpcp-model-track3-fs

Centers for Medicare & Medicaid Services. (n.d.-b). *Comprehensive Primary Care Plus*. https://www.cms.gov/priorities/innovation/innovation-models/comprehensive-primary-care-plus

Chairmen of the Senate Budget and Taxation Committee and House Appropriations Committee. (2024). *Report on the Fiscal 2025 State Operating Budget (SB 360) And the State Capital Budget (SB 361) And Related Recommendations* [Joint Chairman's Report]. https://mgaleg.maryland.gov/Pubs/BudgetFiscal/2024rs-budget-docs-jcr.pdf

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014

Gruber, E., Perman, C., Grisham, R., Adashi, E. Y., & Haft, H. (2023). Association of participation in the Maryland Primary Care Program with COVID-19 outcomes among Medicare beneficiaries. *JAMA Network Open*, *6*(1), e2249791–e2249791.

Henderson, M., Goetschius, L. G., Han, F., & Stockwell, I. (2023). *Risk score specifications and codebook for The Hilltop Institute's pre- models*. The Hilltop Institute at UMBC. https://www.hilltopinstitute.org/wp-content/uploads/publications/RiskScoreSpecificationsCodebookForHilltopPre-Models-Version2-Dec2023.pdf

Kind, A. J. H., & Buckingham, W. R. (2018). Making neighborhood-disadvantage metrics accessible—The neighborhood atlas. *The New England Journal of Medicine*, *378*(26), 2456–2458. https://doi.org/10.1056/NEJMp1802313

Marcus, M., & Sant'Anna, P. H. C. (2021). The role of parallel trends in event study settings: An application to environmental economics. *Journal of the Association of Environmental and Resource Economists*, *8*(2), 235–275. https://doi.org/10.1086/711509

Maryland Department of Health. (2022). *MDPCP HEART payment playbook*. https://health.maryland.gov/mdpcp/Documents/MDPCP_HEART_Payment_Playbook.pdf

Maryland Department of Health. (2023a). *Maryland Primary Care Program (MDPCP) 2022 annual report*.

Maryland Department of Health. (2023b). *Maryland Primary Care Program report*. [Dashboard Report]. https://health.maryland.gov/mdpcp/Documents/Maryland%20Primary%20Care%20Dashboard%20Report.pdf

Maryland Health Services Cost Review Commission. (2021). *Evaluation of the Maryland Primary Care Program: 2021*. [Joint Chairman's Report].

Maryland Health Services Cost Review Commission. (2023). *Evaluation of the Maryland Primary Care Program*. [Joint Chairman's Report].

Neall, R. R., Haft, H., Perman, C., Sowinski-Rice, A., Bowden, S., & Gruber, E. (2019). *Maryland Primary Care Program (MDPCP) 2019 annual report* [Executive Summary]. Maryland Department of Health. https://health.maryland.gov/mdpcp/Documents/Executive%20Summary_MDPCP%202019%20Annual%20Report.pdf

O'Malley, A. S., Singh, P., Fu, N., Duda, N., McCall, N., Geonnotti, K., Petersen, D., Ghosh, A., Morrison, N., Keith, R., Tu, H., Orzol, S., Dale, S., Shanmugam, P., Peebles, V., Lee, K. M., Au,

M., Felland, L., Embry, J., … Little, J. S. (2023). *Independent evaluation of Comprehensive Primary Care Plus (CPC+): Final report*. Mathematica.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373–1379. https://doi.org/10.1016/S0895-4356(96)00236-3

Perman, C., Adashi, E., Gruber, E., & Haft, H. (2021). *Improving COVID-19 outcomes for Medicare beneficiaries: A public health-supported advanced primary care paradigm* [Issue Brief]. The Milbank Memorial Fund. https://www.milbank.org/publications/improving-covid-19-outcomes-for-medicare-beneficiaries-a-public-health-supported-advanced-primary-care-paradigm/

Peterson, G., Rotter, J., Machta, R., Calkins, K., Lee, K. M., Markovitz, A., Sarwar, R., Stewart, K., Vogler, J., Platt, I., Whicher, D., & McCall, N. (2024a). *Evaluation of the Maryland Total Cost of Care Model: Progress report*. Mathematica.

Peterson, G., Rotter, J., Machta, R., Calkins, K., Lee, K. M., Markovitz, A., Sarwar, R., Stewart, K., Vogler, J., Platt, I., Whicher, D., & McCall, N. (2024b). *Evaluation of the Maryland Total Cost of Care Model: Progress report appendices*. Mathematica.

Raths, D. (2024, July 3). Connecticut, Hawaii Joining Maryland, Vermont in CMS' AHEAD Model. *Healthcare Innovation*. https://www.hcinnovationgroup.com/policy-value-based-care/alternative-payment-models/news/55093542/connecticut-hawaii-joining-maryland-vermont-in-cms-ahead-model

Rios-Avila, F., Sant'Anna, P. H. C., & Callaway, B. (2021). *CSDID: Stata module for the estimation of difference-in-difference models with multiple time periods* (1.72) [Stata]. Boston College Department of Economics. https://ideas.repec.org/c/boc/bocode/s458976.html

Rivers, P. A., & Glover, S. H. (2008). Health care competition, strategic mission, and patient satisfaction: Research model and propositions. *Journal of Health Organization and Management*, *22*(6), 627–641.

Roth, J., Sant'Anna, P. H. C., Bilinski, A., & Poe, J. (2023). What's trending in difference-in-differences? A synthesis of the recent econometrics literature. *Journal of Econometrics*.

Rubin, D. B. (2005). Causal Inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, *100*(469), 322–331. https://doi.org/10.1198/016214504000001880

Schrader, D., Haft, H., Perman, C., Sowinski-Rice, A., Anand, R., Gruber, E., Morrison, C., Hope, K., & Grishman, R. (2022). *Maryland Primary Care Program (MDPCP) 2021 annual report*. Maryland Department of Health. https://health.maryland.gov/mdpcp/Documents/2021%20Annual%20Report.pdf

Schrader, D., Haft, H., Perman, C., Sowinski-Rice, A., Bowden, S., Gruber, E., Anand, R., Hope, K., & Roa, J. (2021). *Maryland Primary Care Program (MDPCP) 2020 annual report*. Maryland Department of Health. https://health.maryland.gov/mdpcp/Documents/MDPCP_2020_Annual_Report.pdf

Stevens, N. (2006). *Changing postal ZIP code boundaries* [Report for Congress]. Congressional Research Service. https://apps.dtic.mil/sti/tr/pdf/ADA462043.pdf

The Lewin Group. (2024, July 8). *MDPCP Methodologies: Beneficiary Attribution Payments, and Performance Assessment – Model Year 2023*.

University of Wisconsin School of Medicine and Public Health. (2024). *2020 Area Deprivation Index* (v.4.0.1). https://www.neighborhoodatlas.medicine.wisc.edu/

U.S. Postal Service Office of Inspector General. (2013). *The untold story of the ZIP code* (RARC-WP-13-006).

Appendix

This appendix provides additional detail on several technical items referenced in the main report.

CCLF Data Aggregation

The data source for these analyses were the Claim and Claim Line Feed (CCLF) data files. These files include Part A claims headers, Part A claims revenue centers, Part A procedure codes, Part A diagnosis codes, Part B physician claim lines, Part B durable medical equipment (DME) claims, Part D claims, and beneficiary demographics. These files are created and shared with the Hilltop team monthly, but they reflect a rolling, 36-month lookback timeframe from the most recent month. Because the rolling data extracts reflect final action claims and current information (if fields have changed), the monthly files have many duplicates of the same exact data or records where some information has been updated. As a result, the team combined the monthly data extracts to remove any duplicates and retain the most current information.

For each file, records were matched on the unique identifier (e.g. claim identifier for the Part A claims header files). For the claim records, we compared the previous version of the claim with the next version of the claim received per data extract (e.g., January 2019 version was compared to the previous version from December 2018) across all monthly extracts, and if any information changed, then the most current information for that field was retained. Notably, the Medicare individual unique identifier (the Medicare beneficiary identifier [MBI]) was not invariant across extracts. The change of MBI is identified when the same claim identifier has different MBI values on two consecutive monthly extracts. We catalogued all changes to the MBI across the claims compared using the claim identifier. The new MBI observed was applied to all records, including the beneficiary demographics file to ensure that we were combining all records for the same individual, even when their MBI changed over the course of the study.

Variable Definitions

- Inpatient hospital admissions are defined as Part A claims headers with a claim type code of 60 or 61.
- ED visits (those that do not result in an inpatient hospital admission) are defined as a hospital outpatient claims (claim type code of 40) with at least one revenue center line item from the Part A revenue center file in the following: 0450, 0451, 0452, 0453, 0454, 0455, 0456, 0457, 0458, 0459, and 0981 (Barosso, 2015).
- Avoidable hospital events: Avoidable inpatient admissions and ED visits are defined using the AHRQ 2022 Prevention Quality Indicators (PQI) definitions.[15] Specifically, these create a 0/1 indicator for whether an individual incurred an AH event in quarter *q* of year *y* by searching inpatient admissions and outpatient ED visits—using the definitions described above—for the presence of select procedure and/or diagnosis codes. This comprises 10 constituent underlying potentially avoidable conditions: diabetes short-term

[15] AHRQ Quality Indicators™ Prevention Quality Indicators, 2022

complications, diabetes long-term complications, COPD or asthma in older adults, hypertension, heart failure, community-acquired pneumonia, urinary tract infections, uncontrolled diabetes, asthma in younger adults, and lower-extremity amputation among patients with diabetes. All constituent conditions are, theoretically, preventable given the timely deployment of primary care, and this outcome is also used to create the Pre-AH scores that are sent to practices each month (Henderson et al., 2023).

- Total expenditure is calculated as all Parts A and B Medicare expenditures using the claim line payment amounts (CLM_LINE_NCH_PMT_AMT) for Part B physician and CLM_LINE_PRFNL_NCH_PMT_AMT for DME claims lines, and the claim payment amount for Part A and outpatient claims (CLM_PMT_AMT) from the CCLF Part A claims headers. We summed these three spending variables across all claims occurring during the quarter for each person-quarter; we then winsorized this variable at the 99th percentile for all person-quarters and recoded negative amounts to 0. This is consistent with the recent TCOC evaluation, which winsorizes expenditures and focuses on Medicare payments, which excludes "the amounts that third parties and beneficiaries paid for deductibles, coinsurance, and copayments" (Peterson et al., 2024b, p. 86). Specifically,
    - For Part B physician claim lines: The variable, CLM_LINE_NCH_PMT_AMT, is the "amount of payment made from the trust funds (after deductible and coinsurance amounts have been paid) for the line item service on the non-institutional claim.[16]"
    - For Part B DME claim lines: The variable, CLM_LINE_PRFNL_NCH_PMT_AMT, is the "amount of payment made from the Medicare trust fund (after deductible and coinsurance amounts have been paid) for the line item service on the non-institutional claim.[24]"
    - For Part A and outpatient claims: the variable, CLM_PMT_AMT, is the "Amount of payment made from the Medicare trust fund for the services covered by the claim record. [24]"
- Medicare-Medicaid dual eligibility is defined for the person-quarter using the monthly dual-eligibility flags in the CCLF beneficiary demographics file. Specifically, we imputed quarterly dual eligibility status for each individual as of quarter *q* of year *y* if they have a non-missing dual eligibility flag from any of the constituent months (e.g., January, February, or March for quarter 1) not equal to "N/A" using the variable DUAL_YYYY_MM. We categorize beneficiaries with full or partial Medicaid benefits as duals in this indicator.
- Race group is aggregated into three coarse groupings: White, Black, and all other races. This is operationalized as BENE_RACE_CD equals to 1, 2, or all other values, respectively. This covariate is non-time varying.

[16] See CCLF Information Packet (IP) for more detail: https://www.cms.gov/files/document/cclf-information-packet.pdf

- ZIP code is calculated from the variable BENE_ZIP_CD. For each of the underlying monthly CCLF source files, we used the first available value of BENE_ZIP_CD within a given calendar year quarter. These data are unavailable for quarters from 2018; as such, we imputed the 2018 values using ZIP as of 2019Q1. This covariate varies over time.
- Rural-urban status is operationalized using county FIPS codes, which are calculated for each individual in each quarter, using the values BENE_FIPS_STATE_CD and BENE_FIPS_CNTY_CD. For each of the underlying monthly CCLF source files, we used the first available value of BENE_FIPS_STATE_CD and BENE_FIPS_CNTY_CD within a given calendar year quarter. These data are unavailable for quarters from 2018; as such, we imputed the 2018 values using 2019Q1 values.

  We operationalized rural and urban status using the following logic that is consistent with specifications released by the Maryland Department of Health[17]: if county is Allegany, Garrett, Washington, Caroline, Dorchester, Kent, Queen Anne's, Somerset, Talbot, Wicomico¸ or Worcester, code as "rural." For Maryland counties not coded as rural, code as "urban." For all other counties, code as missing. For the purposes of the subgroup analysis, we assigned the modal value of either rural or urban status, so that it is time-invariant.
- Area Deprivation Index (ADI) was obtained from The University of Wisconsin's Center for Health Disparities Research Neighborhood Atlas (*2020 Area Deprivation Index*, 2024; Kind & Buckingham, 2018). Lower scores indicate lower levels of disadvantage. We used ADI from 2020 and calculate ZIP code ADI levels as the median Census block group level. We used the "state" ADI rank for this analysis and define "high ADI" as being in the top quintile of state ADI rank across Maryland ZIP codes. In order to facilitate the subgroup analysis, we assigned an individual to "high ADI" status if they have ever had a value of ZIP codes that matches to "high ADI" ZIP codes as of 2020 (MDPCP HEART Payment Playbook, 2022). This covariate, like others used in the subgroup analysis, is time-invariant.
- CTO status is operationalized using the "program data" file from the 2023 Q2 MDPCP attribution files. This data file contains longitudinal data by quarter on MDPCP participating practices and includes an identifier for CTO for a given practice in a given quarter. We created a 0/1 flag for whether a given practice has ever been affiliated with a CTO, and then link this to the analytic data set using the aggregated beneficiary attribution files. While individuals can change practices, we used the practice at which the individual has been attributed for the longest time period for this analysis. Where an individual has changed practices and is attributed to different practices for the same duration, we randomly selected a practice.

[17] See page 6 of the Office of Long Term Services and Supports Provider Solicitation—Request for Responses for the Program of All-Inclusive Care for the Elderly, https://health.maryland.gov/mmcp/longtermcare/Documents/Maryland%20PACE%20Expansion%20Solicitation%20%28September%201%2C%202021%29.pdf

These data are only available for treated individuals; as such, we conducted an imputation in order to operationally facilitate the subgroup analysis. Specifically, for each treated individual, we attributed the CTO status to the corresponding matched individual in the comparison group. Thus, the subgroup analysis compared changes in outcomes for ever-treated and never-treated individuals in quarter *q* relative to a baseline period conditional on the treated individual's *primary* practice being ever affiliated with a CTO. This covariate, like others used in the subgroup analysis, is time-invariant.

- Practice affiliation with a health system as of 2023 was provided by the MDPCP PMO, based on data reported by practice coaches. These data are only available as of 2023. Even though hospital ownership status as of 2023 does not necessarily apply to prior years, this should be tantamount to "attenuation bias," and thus bias *against* detecting significant effects. That is, using 2023 practice ownership will contain measurement error in prior years: certain practices indicated as being owned by a health system will not, in fact, be owned by a health system in year *y*; conversely, certain practices not indicated as being owned by a health system in 2023 might, in fact, be owned by a health system in year *y*. As above, this measure is only available for MDPCP participants; thus, we performed the imputation mentioned above in order to operationally facilitate this subgroup analysis. Despite the limitations of this subgroup analysis, the results should be directionally accurate. This covariate, like others used in the subgroup analysis, is time-invariant.

Additionally, we used eleven covariates for the logistic propensity score model:

- Age in years as of quarter *q* in year *y*.
- Sex with female as compared to male.
- A 0/1 indicator for whether an individual's original Medicare eligibility was due to old age.
- Average quarterly Medicare Parts A and B expenditure (defined as above, except without Part B DME expenditures) for the prior calendar year.
- Average quarterly inpatient utilization (defined as above) for the prior calendar year.
- Average quarterly ED utilization (defined as above) for the prior calendar year.
- Average quarterly AH event incidence (defined as above,) for the prior calendar year.
- Change in average quarterly Medicare Parts A and B expenditure (without Part B DME expenditures) from year t-2 to year t-1.
- Change in average quarterly inpatient utilization from year t-2 to year t-1.
- Change in average quarterly ED utilization from year t-2 to year t-1.
- Change in average quarterly AH event incidence from year t-2 to year t-1.

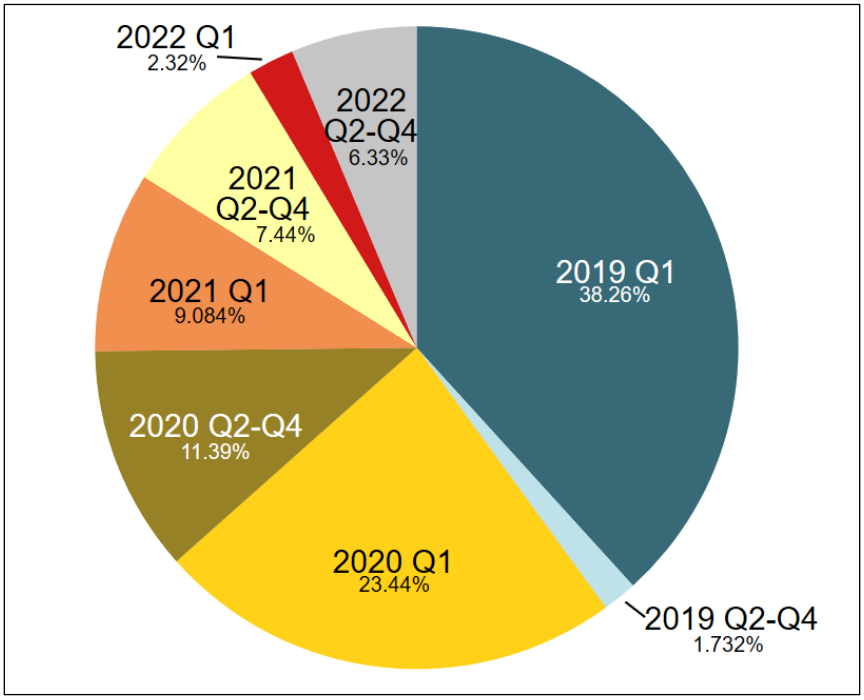
We note that the measures of total spending and change in total spending did not include Part B DME payments in the matching (but the final spending outcome does include Part B DME payments). The average amount of Part B DME spending only constitutes 1.21% of total

spending per quarter; given the small magnitude of DME expenditure, we do not anticipate that the omission of Part B DME spending in the matching algorithm materially affected the overall study results.

Analytic Data Set

This section presents additional details on the construction of the analytic data set. First, we appended the quarterly MDPCP attribution files to identify all beneficiaries who have ever been attributed to an MDPCP-participating provider from 2019Q1-2022Q4, as well as their first attribution date. We used an "intent-to-treat" interpretation and assume that attribution is an absorbing state: that is, once an individual is attributed once, they are considered "treated" for the remainder of the sample period. This is a commonly used approach to minimize selection bias, such as in the federally funded evaluation of the CPC+ program (O'Malley et al., 2023). We excluded beneficiaries that are missing MBI. The total number of beneficiaries in this population is 574,080. Figure A.1 shows the distribution of beneficiaries by quarter of joining.

Figure A.1. MDPCP Beneficiaries by Joining Quarter



Notes: Joining quarter is the earliest quarter that the beneficiary was attributed to a participating MDPCP practice for the treated group or attributed to a primary care practice in Maryland, if in the comparison group.
Q – quarter of the calendar year.

Next, we appended the quarterly non-participating attribution files to identify all beneficiaries who have ever been attribution to a non-MDPCP primary care provider from 2019Q1-2022Q4.

The total number of beneficiaries in this population is 612,035—some of whom may have been attributed to an MDPCP-participating provider, at some point, too.

Hilltop then merged these two populations together by MBI to identify two key groups: "ever treated" and "never treated." We define the "ever treated" population as those individuals who have ever been attributed to an MDPCP-participating provider. As noted above, the total number of beneficiaries in this population is 574,080. The individuals from the non-participating attribution files who have never appeared in the MDPCP attribution files is the "never treated" population. The total number of beneficiaries in this population is 340,294. In sum, the entire treatment and comparison populations consist of 914,374 unique individuals. This population is then restricted and filtered in several ways.

Eligibility Restriction

Hilltop first restricted the analytic sample based on eligibility. For any individual in the never-treated or ever-treated populations to enter the final analytic data set in a given quarter, they must have traditional, fee-for-service Medicare for both Parts A and B, reside in Maryland, and be alive. If they do not have traditional A and B eligibility, we will not observe their full claims history in the CCLF data, and our estimates of the effect of MDPCP on spending and utilization would reflect incomplete data.

We assembled eligibility information from the CCLF beneficiary demographics files. These files contain monthly eligibility information based on Medicare buy-in (indicating whether an individual has Parts A and/or B), Maryland residency, HMO use (indicating whether an individual has Medicare Part C), and date of death. As noted above, we imputed quarterly eligibility for each individual as of quarter *q* of year *y* if they have an eligibility flag from any of the constituent months (e.g., January, February, or March for quarter 1) equal to "AB" using the variable ELIG_YYYY_MM. We calculated this for all individuals from the period 2018Q1 – 2022Q4. With these restrictions applied, this amounts to 14,072,424 person-quarters across 884,683 individuals from the ever treated and never-treated populations. Of this total, 569,219 are ever treated and 315,464 are never treated.

Prior to matching, we imposed four additional restrictions. First, individuals in the ever-treated group were dropped if they have no post-treatment experience in the eligibility data (for example, if an individual attributed to MDPCP as of 2019Q1 does not have Medicare Part A and B FFS eligibility while residing in Maryland for any quarters after 2018Q4). Second, individuals in the ever-treated group were dropped if they do not have at least one pre-treatment quarter of Medicare Part A and B FFS eligibility while residing in Maryland (for example, if an individual attributed to MDPCP as of 2019Q1 does not have Medicare Part A and B FFS eligibility prior to 2019Q1). Third, individuals in the ever-treated group were restricted to only having up to 4

quarters of pre-treatment experience. Finally, any person-quarters after an individual's quarter of death were removed from the analytic data set.[18]

We imposed the analogous restrictions on the matching pool for the comparison group: individuals must have at least two quarters' experience in the non-participating provider attribution file, and we match using individuals' second year-quarter so that each matched comparison individual is guaranteed to have at least one quarter of pre-treatment experience. As above, we remove any person-quarters after an individual's quarter of death.

Finally, we exclude individuals who are missing data for any matching covariates or who do not have a valid Maryland ZIP3 (i.e., outside the range 206-219). The final sample upon which the matching algorithm is deployed consists of 865,271 total individuals: 556,240 ever-treated and 309,031 never-treated.

Condition History Balance Check

In order to assess the extent to which the population of MDPCP-attributed beneficiaries resembled the comparison group in terms of acuity, we conducted a balance check using eight condition history flags available in the CCLF beneficiary demographics file. Each flag represents whether an individual met the requirements of the chronic condition algorithm for that time period. In this analysis, we used the chronic condition flag as of the year prior to attribution (or, for the comparison group, the attribution for the matched treated individual).

Chronic condition flags in the CCLF take four possible values: 0, to denote if the beneficiary either did not meet the claims criteria or have sufficient fee for service coverage; 1, if the beneficiary met the claims criteria for the condition but did not have sufficient fee for service coverage; 2, if the beneficiary did not meet the claims criteria but did have sufficient fee for service coverage; and 3, if the beneficiary met the claims criteria and had sufficient claims coverage.[19] For the purposes of this exercise, we collapsed this into a 0/1 flag, where the flag takes a value of 1 if the beneficiary meets the claims criteria (original values 1 and 3), and 0 otherwise.

Hilltop assessed balance for heart failure, chronic kidney disease, COPD, lung cancer, acute myocardial infarction, hip/pelvic fracture, stroke, and female/male breast cancer. We found that the treatment and comparison groups are well-balanced along these baseline conditions. While the treatment group exhibits marginally higher rates of chronic kidney disease, female/male breast cancer, and stroke, the comparison group exhibits marginally higher rates of COPD, heart failure, hip/pelvic fracture, and lung cancer. Table A.1, below, presents the results.

[18] While this is largely incorporated via the ELIG_YYYY_QQ variable, we observed instances where who died in, for example, December 2020 is still present in the analytic data set as of 2021 Q1. This final sample restriction is intended to remove these edge cases.

[19] See Master Beneficiary Summary File (MBSF) data dictionary: https://resdac.org/cms-data/variables/mbsf-27-cc/alzheimers-disease-end-year-indicator

Table A.1. Condition History Balance Tests for Matched Treatment and Comparison Group

Condition	Treatment Mean	Matched Comparison Mean	Standardized Difference
Acute Myocardial Infarction	0.0062	0.0062	0.0003
Chronic Kidney Disease	0.2173	0.2101	0.0176
COPD	0.0879	0.0889	-0.0036
Female/Male Breast Cancer	0.0373	0.0366	0.0040
Heart Failure	0.0932	0.1021	-0.0302
Hip/Pelvic Fracture	0.0047	0.0052	-0.0077
Lung Cancer	0.0091	0.0094	-0.0028
Stroke	0.0390	0.0383	0.0035

Notes: This table indicates the balance on eight selected conditions for the individuals attributed to MDPCP-participating practices and their matched comparators, including the mean for the treated group, mean for the matched comparison group, and standardized difference. For this analysis, condition variables from the CCLF beneficiary demographics file are collapsed into 0/1 flags, where the flag takes a value of 1 if the beneficiary meets the claims criteria (original values 1 and 3), and 0 otherwise.

Balance By Joining Cohort

This section presents additional details on matched sample balance by primary cohort. Table A.2 provides the balance results.

Table A.2. Characteristics of Treatment and Comparison Group by Three Primary Joining Cohorts

Baseline Characteristic	Treatment Mean	Matched Comparison Mean	Standardized Difference
2019 Q1 Joiners (38.3% of ever-treated)			
Average baseline quarterly Medicare Part A & B expenditure*	$3506.8	$3485.4	0.0031
Average baseline quarterly inpatient utilization	0.168	0.167	0.0029
Average baseline quarterly ED utilization	0.287	0.285	0.0029
Average baseline quarterly avoidable hospitalization event incidence	0.059	0.06	-0.0032
Change in average baseline quarterly Medicare Part A & B expenditure†	$608.9	$623.2	-0.0019
Change in average baseline quarterly inpatient utilization	0.021	0.021	0.0002
Change in average baseline quarterly ED utilization	0.025	0.025	0.0003
Change in average baseline quarterly avoidable hospitalization event incidence	0.009	0.009	-0.0003
Original Medicare Entitlement due to Age	0.843	0.845	-0.0059
Sex (female)	0.598	0.599	-0.0018

Baseline Characteristic	Treatment Mean	Matched Comparison Mean	Standardized Difference
Age	74.1	74.2	-0.0078
Race (Black)	0.203	0.203	0.0000
2020 Q1 Joiners (23.6% of ever-treated)			
Average baseline quarterly Medicare Part A & B expenditure*	$3861.3	$3865	-0.0005
Average baseline quarterly inpatient utilization	0.173	0.173	-0.0001
Average baseline quarterly ED utilization	0.296	0.298	-0.0042
Average baseline quarterly avoidable hospitalization event incidence	0.063	0.062	0.0035
Change in average baseline quarterly Medicare Part A & B expenditure†	$899.1	$906.6	-0.0009
Change in average baseline quarterly inpatient utilization	0.032	0.033	-0.0011
Change in average baseline quarterly ED utilization	0.030	0.032	-0.0023
Change in average baseline quarterly avoidable hospitalization event incidence	0.014	0.013	0.0034
Original Medicare Entitlement due to Age	0.822	0.82	0.0042
Sex (female)	0.595	0.593	0.0023
Age	73.0	73.1	-0.0099
Race (Black)	0.220	0.220	0.0000
2021 Q1 Joiners (9.1% of ever-treated)			
Average baseline quarterly Medicare Parts A & B expenditure*	$3677.4	$3706.4	-0.0035
Average baseline quarterly inpatient utilization	0.178	0.179	-0.0007
Average baseline quarterly ED utilization	0.231	0.226	0.0088
Average baseline quarterly avoidable hospitalization event incidence	0.049	0.047	0.0066
Change in average baseline quarterly Medicare Parts A & B expenditure†	$1060.1	$1077.3	-0.0021
Change in average baseline quarterly inpatient utilization	0.055	0.055	0.0000
Change in average baseline quarterly ED utilization	-0.01	-0.01	0.0012
Change in average baseline quarterly avoidable hospitalization event incidence	0.002	0.001	0.0055
Original Medicare Entitlement due to Age	0.779	0.782	-0.0079
Sex (female)	0.580	0.590	-0.0199
Age	70.9	71.0	-0.0091
Race (Black)	0.284	0.284	0.0000

Notes: *Spending is winsorized at the 99th percentile, with negative values re-coded to 0.
†Change in spending in winsorized at the 99th and 1st percentiles.

Overall, these results indicate strong balance on baseline observable characteristics. Notably, Table A.2 indicates that the composition of the MDPCP population changes in meaningful ways over time. For individuals first attributed in 2019Q1, the average age is 74.1 years old, 20.3% of the population is Black, and 84.3% of individuals have an original Medicare entitlement reason of old age. For individuals first attributed in 2021Q1, however, the average age is 70.9 years, 28.4% of the population is Black, and 77.9% of individuals have an original Medicare entitlement reason due to age.

Difference-in-Differences Methodological Details

Hilltop estimated the impact of MDPCP on spending and utilization using a difference-in-differences approach designed and validated from data with multiple time periods and staggered treatment onsets (Callaway & Sant'Anna, 2021). This method is conceptually similar to a traditional difference-in-differences analysis in that it quantifies the impact of an intervention by comparing treatment and comparison groups over time before and after starting the intervention. However, it differs in that it isolates "clean" comparisons between individuals in the treatment and comparison groups over time and excludes "forbidden" comparisons that occur when members of the treatment group start receiving the treatment at different times and that can result in biased, inaccurate causal estimates (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021; Roth et al., 2023).

The difference-in-differences method used in this evaluation is based on the group-time average treatment effect ($ATT(g,t)$ p), which estimates the average treatment effect (ATT) for members of a particular group (g) at a particular time period (t). This effect can be denoted by

$$ATT(g,t) = \mathrm{E}\left[Y_t(g) - Y_t(0) \middle| G_g = 1\right].$$

This means that the model estimated the effect of MDPCP separately for Medicare beneficiaries in each joining cohort (e.g., 2019Q1) for each quarter in the analytic data set (e.g., 2019Q1, 2019Q2, etc.). This method considers treatment to be an "absorbing state," meaning that once a beneficiary has been attributed to MDPCP for at least one quarter, they are considered treated for the remainder of the study. We used our matched sample of Medicare beneficiaries who were never attributed to MDPCP practices for the comparison group (see the Methodology: Analytic Populations section for more details).

Results Aggregation

Estimating the group-time average treatment effect allowed for the flexibility to aggregate the treatment effect parameter at multiple levels to answer different research questions (Callaway & Sant'Anna, 2021). First, we aggregated the $ATT(g,t)$ parameter to the group level, which represents the effect of MDPCP for Medicare beneficiaries in each joining cohort across all treated time periods:

$$\theta_{S(g)} = \frac{1}{T-g+1}\sum_{t=2}^{T} 1\{g \leq t\} ATT(g,t),$$

and then aggregated the average effects for each group to estimate the overall effect of MDPCP for each outcome that was weighted based on group size:

$$\theta_S^O := \sum_{g=2}^{T} \theta_s(g) P(G = g).$$

Second, similar to an event-study design, we aggregated the $ATT(g,t)$ parameter based on length of exposure to the treatment (e) (i.e., how many quarters a beneficiary was "treated" with MDPCP):

$$\theta_D(e) := \sum_{g=2}^{T} 1\{g + e \leq T\} ATT(g, g+e) P(G = g | G + e \leq T).$$

Last, to determine whether the effect of MDPCP was different during the height of the COVID-19 pandemic (2020Q2 – 2021Q4), we aggregated the $ATT(g,t)$ parameter based on calendar time to quantify the average effect of MDPCP for all treated beneficiaries in each quarter:

$$\theta_c(t) = \sum_{g \in G} 1\{t \geq g\} P(G = g | G \leq t) ATT(g,t).$$

This analytic approach resulted in an unbiased estimate of the causal effect of MDPCP overall ($\theta_{S(g)}$) while also quantifying the average effect of MDPCP for each joining cohort of Medicare beneficiaries ($\theta_{S(g)}$), the average effect of MDPCP based on how many quarters a beneficiary was considered "treated" ($\theta_D(e)$), and the average effect of MDPCP for each calendar quarter (θ_c^O). These different ways of aggregating $ATT(g,t)$ enabled us to evaluate whether there were heterogeneous treatment effects by joining quarter, time attributed, or calendar time.

Subgroup Analyses

To determine whether there were differential effects of MDPCP on utilization or expenditures for subgroups of beneficiaries, we estimated six sets of subgroup regressions where we ran a separate difference-in-differences model for each subgroup (i.e., Black beneficiaries) and each outcome (i.e., total Medicare Part A and B spending) resulting in 13 subgroup regression models per outcome for four outcome 52 subgroup models total). We aggregated each subgroup's regression results to estimates the overall treatment effect and compared the pattern of results across subgroups rather than testing for interactions between MDPCP attribution and subgroup membership because interpretation of significant or non-significant interaction estimates across all $\boldsymbol{ATT(g,t)}$ estimates was not feasible or likely meaningful.

Computational Details

All difference-in-differences analyses were operationalized using the "CSDID: Stata module for the estimation of Difference-in-Difference models with multiple time periods" (Rios-Avila et al., 2021 v1.72) in Stata (v18.0). Data were treated as panel data and therefore included individual-level fixed effects. We used the linear outcome regression ("reg") estimator, clustered on the individual MBI level, and used a multiplicative WildBootstrap procedure with 999 repetitions using a mammen approach to estimate standard errors (the default bootstrap options). Consistent with standard event-study effects, pre-treatment ATT effects were estimated using "long2" gaps, which uses T-1 as base period, and G-1 as the post period, where G is the first period a unit received treatment.

Results from Sensitivity Analyses

Annual Analyses

The analytic plan for this evaluation was designed to match the outcome specifications with the quarterly MDPCP attribution cadence to precisely model the impact of the program on spending and utilization. To assess the extent to which intra-year, inter-quarter variation in spending and utilization may influence our impact estimates, we conducted a robustness check to model the impact of MDPCP on annual outcomes aggregating to the yearly level. We operationalized this using the sum of quarterly spending and indicator variables indicating any inpatient, ED, or AH event utilization, respectively, in a given year. In our regressions, we used the Callaway & Sant'Anna 2021 estimator, as in our baseline specification, but weighted the beneficiaries in the year-level data set by the percentage of treatment quarters per year a beneficiary was present in the quarterly analytic data set. Furthermore, the difference-in-differences approach used in this evaluation requires all individuals treated at the same time (e.g., 2020) to have the same pre-treatment period (e.g., 2019). Therefore, in this sensitivity check, we only included beneficiaries who had a full year (four quarters) of pre-treatment data (73% of beneficiaries).

The Hilltop team found that, similar to the baseline specification, the introduction of MDPCP led to significant declines in expenditure and inpatient utilization and no difference in incidence of AH events. In contrast with the headline impact estimates, there was no longer a statistically significant reduction in emergency department utilization. Table A.3 presents the results. It is worth noting that the analytic strategy in this evaluation was not optimized for year-level data and thus, it is not surprising that these results are less statistically precise than the quarterly results.

Table A.3. Annual Effect of MDPCP

Outcome	Annual Effect Estimate	95% Confidence Interval	
		Lower	Upper
Total Medicare Parts A & B expenditure	**-$314.88**	**-$425.57**	**-$204.18**
Any inpatient utilization	**-0.0095**	**-0.0115**	**-0.0074**
Any ED utilization	-0.0007	-0.0031	0.0017
Any AH event	-0.0008	-0.0020	0.0004

Notes: The adjusted effect estimates presented are average changes in the measure level per attributed Medicare beneficiary per calendar year, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

Additional Outcomes

As another robustness check, we evaluated the impact of MDPCP on four additional outcomes: number of inpatient admissions (as defined by claim counts), number of ED visits, AH events in the inpatient setting, and AH events in the ED. We used our baseline, person-quarter specification for this analysis. See Table A.4 for results.

Table A.4. Overall Quarterly Effect for Alternative Outcome Definitions

Outcome	Quarterly Effect Estimate	95% Confidence Interval	
		Lower	Upper
Number of inpatient admissions	**-0.0039**	**-0.0052**	**-0.0026**
Number of ED visits	-0.0015	-0.0032	0.0003
Any avoidable inpatient hospital event	-0.0003	-0.0008	0.0003
Any avoidable ED event	-0.0002	-0.0007	0.0002

Notes: The adjusted effect estimates presented are average changes in the measure level per attributed Medicare beneficiary per calendar quarter, after accounting for changes among matched comparators and controlling for individual fixed effects. **Bold text** indicates that the estimated effect is statistically significant at the 5% confidence level.

Similar to the baseline specification, results show that attribution to an MDPCP-participating provider was associated with a significant reduction in the number of inpatient claims per quarter. We found that the effect of MDPCP on the number of ED visits per quarter was not statistically significant. However, the overall pattern of results is similar to what we observed when using the quarterly, any ED visit outcome where the effect of MDPCP was statistically significant. We see that the effect estimate is similar (any ED: -0.0013, number of ED: -0.0015); however, the 95% CI for the number of events is slightly wider, suggesting more variability when using the number of ED visits as an outcome. Additionally, the small, negative impact on AH events is not differentiated between site of care: either the ED or the inpatient setting. Both effects are modest in magnitude and statistically insignificant. Overall, these alternative outcome definitions were intended as a validity check for the main outcome specifications and yield consistent findings with the main analysis.

The Hilltop Institute

UMBC

Sondheim Hall, 3rd Floor
1000 Hilltop Circle
Baltimore, MD 21250
410-455-6854
www.hilltopinstitute.org