Evaluation of the Maryland Total Cost of Care Model: Quantitative-Only Report for the Model's First Three Years (2019 to 2021)

December 2022

Jason Rotter, Keri Calkins, Kate Stewart, Isabel Platt, Rachel Machta, Keith Kranker, Nancy McCall, Greg Peterson

Submitted to:

U.S. Department of Health and Human Services
Centers for Medicare & Medicaid Services
7500 Security Blvd.
Baltimore, MD 21244-1850
Contracting Officer's Representative: Caroline Ly
Contract Number: HHSM-500-2014-00034I

Submitted by:

Mathematica
1100 1st Street, NE 12th Floor
Washington, DC 20002-4221
Telephone: (202) 484-9220
Project Director: Greg Peterson
Reference Number: 50865

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2014-00034I, Task Order 75FCMC19F0003. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. Mathematica assumes responsibility for the accuracy and completeness of the information contained in this report.

Acknowledgments

The authors wish to thank Sheena Flowers, Donovan Griffin, Sandi Nelson, Christopher Palo, Margaret Raskob, and Megan Baker for their contributions to this report. We are grateful to the people from the Centers for Medicare & Medicaid Services and the Maryland Health Services Cost Review Commission who shared their knowledge and experience with us. We also appreciate the guidance from Centers for Medicare & Medicaid Services staff, including Julia Zucco, Chaz Aloe, Katherine Giuriceo, and Suzanne Wensky.

Contents

Tables

Figures

Acronyms

ACO	Accountable care organization
ACS	American Community Survey
AHRQ	Agency for Healthcare Research and Quality
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CCW	Chronic Condition Data Warehouse
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
CMS	Centers for Medicare & Medicaid Services
CRISP	Chesapeake Regional Information System for our Patients
CRP	Care Redesign Program
CTI	Care Transformation Initiatives
DPP	Diabetes Prevention Program
ED	Emergency department
ESRD	End-stage renal disease
FFS	Fee for service
HCC	Hierarchical Condition Category
HPP	Hospital Payment Program
HRSA	Health Resources and Services Administration
HSCRC	Health Services Cost and Review Commission
IPPS	Inpatient prospective payment system
MD TCOC	Maryland Total Cost of Care
MDAPM	Maryland All-Payer Model
MDPCP	Maryland Primary Care Program
MD-PPAS	Medicare Provider Practice and Specialty
MPA	Medicare Performance Adjustment
NPI	National provider identifier
PBPY	Per beneficiary per year
PCP	Primary care physician
pp	Percentage point
PQI	Prevention Quality Indicator
PUMA	Public Use Microdata Area
SIHIS	Statewide Integrated Health Improvement Strategy
SVI	Social Vulnerability Index
TIN	Taxpayer identification number

1. Introduction and key results

In 2018, the Centers for Medicare & Medicaid Services (CMS) and the state of Maryland signed an agreement that established the Maryland Total Cost of Care (MD TCOC) Model. MD TCOC builds on the Maryland All-Payer Model (MDAPM), which ran from 2014 to 2018 and created hospital all-payer global budgets for regulated Maryland hospitals. MD TCOC continues using hospital global budgets and extends that transformation beyond hospital walls by expanding statewide accountability for cost and quality outcomes and broadening the incentives and supports to providers to transform care (Sapra et al. 2019; Machta et al. 2021).

This report presents estimates of MD TCOC impacts in its first three years (2019 to 2021). Future reports will cover more years and combine qualitative and quantitative data on model implementation and impacts.

Although MDAPM and MD TCOC are distinct models (and established by separate legal agreements), we conceptualize them as parts of an overarching and evolving Maryland Model for the purpose of estimating impacts. We estimate impacts each year, starting in 2014, relative to an estimate using a nationally matched sample of what would have occurred in the state if Maryland and CMS had not made any of the changes they did starting in 2014. We describe impacts during the MD TCOC period (2019–2021) and how those impacts compare with impacts during earlier periods, specifically the last two years of the MDAPM period (2017–2018). We do not, however, attempt to separate the effects of the new components MD TCOC started in 2019 from the hospital global budgets. Hospital global budgets have had a strong and growing influence on hospital outcomes that cannot be isolated from the new model components.

Overall, we find that the Maryland Model had the following favorable effects during the first three years of the MD TCOC period (2019 to 2021; see Table 1, column 3, for impact estimates for all outcomes):

- It substantially reduced rates of all-cause acute care hospital admissions (by 16.1 percent).
- It moderately reduced total Medicare fee-for-service (FFS) spending (Part A and Part B) by 2.5 percent. The Maryland Model increased non-hospital spending (by 2.7 percent) but reduced hospital spending by more (6.6 percent), leading to a $781 million reduction in total spending.
- It improved several quality-of-care measures, including reducing potentially preventable admissions (by 16.1 percent), reducing the likelihood of an unplanned readmission to the hospital (9.5 percent), and increasing timely follow-up after hospital discharge (2.5 percent).

The model did not, however, measurably affect patients' ratings of their personal doctor or the hospitals in which they received care. The personal doctor rating came from the FFS and Medicare Advantage Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys and the hospital ratings came from the Hospital CAHPS surveys. Although the model did not improve either of these patient rating measures, the results suggest that hospitals' efforts to improve efficiency have not come at the expense of lower patient ratings.

For most outcomes, the impacts were larger and more favorable during the MD TCOC period than they were at the end of the MDAPM period (2017–2018), indicating further improvement. For example, all-cause admissions impacts were 6.1 percentage points larger (16.1 versus 10.0 percent), total Medicare spending impacts were 1.5 percentage points larger (2.5 versus 1.0 percent), and impacts on the likelihood of readmission were 1.6 percentage points larger (9.5 versus 7.9 percent). The larger, more favorable effects during the MD TCOC period might be due to (1) the growing influence of global budgets that began in 2014; (2) the broader accountability, incentives, and supports that Maryland and CMS introduced in 2019 and that continue to evolve; and (3) the synergies between the two.

Impacts were mostly consistent across the first three years of the MD TCOC period with one notable exception: non-hospital spending. In 2019, the Maryland Model only modestly increased non-hospital spending (2.2 percent), and, in 2020, the model's impact estimate (0.3 percent) was not statistically different from zero. However, the model increased non-hospital spending substantially in 2021 (5.5 percent). Future impact estimates will indicate whether 2021 was an aberration.

Findings in this report were generally robust to a range of sensitivity tests, including those designed to limit the likelihood that COVID-19 could bias impact estimates.

Table 1. Summary of impacts of the Maryland Model during the MD TCOC period and earlier

		Percentage impact of the Maryland Model[a]		
Outcome	Favorable direction of effect	During the MD TCOC period (2019–2021)	During the last two years of the MDAPM period (2017–2018)	Difference[b]
Domain: Health care utilization				
All-cause acute care hospital admissions	↓	-16.1%*	-10.0%*	-6.1pp*
Outpatient ED visits and observation stays	↓	-3.8%*	-2.2%*	-1.6pp*
Intensity of hospital care (measured by standardized hospital spending)[c]	↓	-8.0%*	-4.5%*	-3.5pp*
Domain: Medicare FFS spending				
Total Medicare FFS spending (Parts A and B)	↓	-2.5%*	-1.0%*	-1.5pp*
Hospital spending (inpatient and outpatient)	↓	-6.6%*	-4.6%*	-2.0pp*
Non-hospital spending	↓ or ↑[d]	2.7%*	3.7%*	-1.0pp*
Total Medicare FFS spending + non-claims payments[c]	↓	-2.9%*	-1.2%*	-1.7pp*
Post-acute care spending	↓	-7.1%*	-2.8%*	-4.3pp*
Domain: Quality of care				
Potentially preventable admissions	↓	-16.1%*	-9.2%*	-6.9pp*
30-day post-discharge unplanned readmission	↓	-9.5%*	-7.9%*	-1.6pp*
Timely follow-up after acute exacerbation of chronic conditions	↑	2.5%*	2.0%*	0.5pp
Patients' rating of their personal doctor[c]	↑	0.4%	0.1%	0.3pp
Patients' rating of their hospital[d]	↑	0.7%	0.9%	-0.2pp
Domain: Population health				
Use of Diabetes Prevention Program services[e]	↑	Not shown[d]	Not shown[d]	Not shown[d]

* Asterisks indicate statistical significance at the $p < 0.10$ level.

[a] We calculate the percentage impact as the impact estimate divided by the unadjusted Maryland mean minus the impact estimate.

[b] We calculate the difference as the percentage impact in 2019–2021 minus the percentage impact in 2017–2018.

[c] Total spending with non-claims payments, standardized spending, and patients' rating of their hospital were only available at the time of this report through 2020. These estimates exclude the model impact in 2021. Patients' rating of their personal doctor was only available at the time of this report through 2019, so the report excludes model impacts in 2020 and 2021.

[d] The model could have favorable effects on non-hospital spending if it either reduced non-hospital spending or increased it by less than the decreases in hospital spending, leading to total Medicare savings.

[e] The impact estimate as a percentage of the mean for the Diabetes Prevention Program was unstable because the use of (or billing for) Diabetes Prevention Program services in Maryland was very low. See Appendix A, Table A.5.3 for detailed impact estimates.
ED = emergency department; FFS = fee for service; pp= percentage point.

2. Background on the Maryland Model

The Maryland Model builds on a unique hospital financing system in Maryland that stretches back decades. In the 1970s, Maryland—in response to rising hospital spending and rising uncompensated care—began regulating hospital prices (Murray and Berenson 2015). An independent commission, the Health Services Cost Review Commission (HSCRC), began setting the rates that all payers paid for care in most hospitals in the state. CMS waived the payment systems that govern Medicare hospital payments in the rest of the country, allowing Maryland to set the prices that Medicare paid for hospital care. Maryland could maintain this waiver so long as the Medicare price per hospital stay did not grow faster in Maryland than it did nationally. Maryland met this requirement for decades, but hospitals compensated for low price growth by increasing volume, which raised total hospital spending (Murray and Berenson 2015).

The Maryland Model began in 2014 with MDAPM, which focused on hospital spending and quality. The state committed to (1) limiting the growth in *total* hospital spending per Maryland resident (combining price and volume), and (2) improving several hospital-focused quality measures, such as reducing preventable complications that develop during a hospital stay. To help meet these commitments, HSCRC switched from setting prices per service to setting prospective global budgets across all payers for each hospital in the state. At the start of each state fiscal year, HSCRC sets a global budget across all payers for each hospital, with the size of the budget tied to historical volume, performance on quality measures, and other factors. Hospitals continue to bill Medicare and other payers per service, but the hospitals continually adjust the prices they charge so that, by the end of the of the year, their total revenue (price times volume) matches their budgets. These global budgets have created strong incentives for hospitals to invest in strategies to reduce avoidable or low-value inpatient and outpatient services. These reductions in hospital service use lower operating expenses and—because revenues are fixed—generate net revenues. Hospitals can keep these revenues as increased margins or reinvest them in other initiatives, such as those to improve population health. Further, the global budgets allow HSCRC to set the rate of growth in hospital spending in the state, helping the state meet its savings commitments.

The Maryland Model began its next phase—which is legally distinct from, but building on, the first phase—in 2019, with the start of the MD TCOC period. This phase (2019 to present) expanded accountability and incentives beyond the hospital. As of the end of 2021, the Maryland Model includes the following:

- **The state's commitment to limit growth in *total* (hospital plus non-hospital) Medicare FFS spending (Part A and Part B)** per Medicare beneficiary, generating $2 billion in savings over 8 years (2019 to 2026; relative to what Medicare would have spent during that period if Medicare spending per beneficiary in 2013 had grown at the national rate). If the state fails to generate these savings, CMS can remove the waiver that allows Maryland to set Medicare hospital prices.
- **The state's commitment in the Statewide Integrated Health Improvement Strategy (SIHIS) to improve quality** in three areas: hospital quality of care, care transformation across the health system, and population health. These commitments include decreasing hospital admissions that could be prevented with better ambulatory care, increasing follow-up after hospital discharge, decreasing the mean body mass index among Maryland residents, and reducing deaths because of drug overdoses.
- **The Hospital Payment Program** (HPP), which continues the hospital global budgets that began in 2014, adjusting them based on the hospital's performance on quality measures. Starting in 2019,

HSCRC also adjusts payments based on the hospital's performance on total cost of care for attributed beneficiaries through the Medicare Performance Adjustment (MPA). Almost all hospitals in Maryland receive all-payer global budgets (52 hospitals). The only hospitals that do not are those for which HSCRC does not set Medicare payment rates, including federal hospitals, children's hospitals, and some specialty hospitals.

- **The Care Redesign Program** (CRP), which (1) allows hospitals to pay providers for interventions to help the hospital perform well under quality-adjusted global budgets through the portion of the CRP called the Health Care Improvement Program (HCIP), and (2) rewards hospitals and their partners for improving the efficiency and quality of care for episodes beyond the hospital stay through the portion of the CRP called the Episode Care Improvement Program (ECIP). In 2021, 4 hospitals participated in HCIP and 21 participated in ECIP.[1]
- **Care Transformation Initiatives** (CTIs), which reward hospitals for efficient episodes of care but give them more flexibility (compared with ECIP) in defining the episodes and interventions. In 2021, 42 hospitals participated in CTIs for one or more types of episodes of care.
- **The Maryland Primary Care Program** (MDPCP), which pays primary care practices and associated Care Transformation Organizations for improving the quality and comprehensiveness of primary care. In 2021, 524 primary care practices were participating in MDPCP, accounting for 27 percent of all primary care practices in the state.
- **Outcomes-Based Credits**, which incentivize Maryland to improve targeted population health outcomes. If Maryland limits the incidence of diabetes or other agreed-upon conditions in Maryland relative to a comparison group, CMS will deduct the expected lifetime savings to Medicare associated with the improvements from the state's savings targets.
- **The Regional Partnership Catalyst Program,** funded through the hospital rate-setting system, which supports hospitals and their community partners in efforts to improve population health. So far, this new funding to hospitals has focused on reducing diabetes incidence and improving behavioral health.

The MD TCOC Implementation Report (Machta et al. 2021) provides more details about model design and implementation.

3. Outcomes and methods for estimating impacts

3.1. Outcomes and how the model could improve them

Together with CMS, we selected outcomes for this report that capture important dimensions of what the Maryland Model aims to accomplish during the MD TCOC period and are feasible to construct with Medicare FFS claims, enrollment data, and patient experience surveys. Table 2 lists the selected outcomes, the direction of effect that would be favorable, and how the model's incentives and supports could lead to these favorable effects. We selected outcomes based on the model logic (described in Machta et al. 2021), the financial commitments that CMS and Maryland set in the MD TCOC state agreement establishing the model (CMS 2018), and the quality goals that CMS and Maryland set in SIHIS (HSCRC 2020a). To the extent feasible, we aligned quality outcomes in this report with SIHIS, but not all SIHIS outcomes can be measured using claims. Further, SIHIS typically focuses on improvements

[1] In 2022, Maryland also launched a specialty-based episode program under CRP called the Episode Quality Improvement Program (EQIP). However, we do not include EQIP in this report because it began in 2022, after the period covered by the impact estimates in this report.

for all Maryland residents, and this report focuses specifically on Medicare FFS beneficiaries. Appendix Section B.3 compares the measures in this report with those in the state agreement and SIHIS.

Although we grounded our hypotheses in the state agreement and SIHIS, we are not formally assessing whether the model met the terms of the state agreement or met SIHIS goals. The state agreement and SIHIS set their own targets and methods for assessing progress. For example, the state agreement uses a national benchmark to assess savings, and we use a matched comparison group to estimate the path Maryland would have been on if not for the Maryland Model (see Section 3.2). CMS, through a separate model implementation contractor, is assessing whether the state has met its savings commitments.

Table 2. Outcomes, the directions of effect that would be favorable, and how the model could improve these outcomes

Outcome	Favorable direction of effect	How the Maryland Model could improve this outcome
Domain: Health care utilization		
All-cause acute care hospital admissions	↓	The hospital global budgets create strong incentives for hospitals to reduce admissions because doing so generates net revenues that hospitals can keep as margins or reinvest in other initiatives. Hospitals can improve care to reduce preventable admissions or shift care to lower-acuity settings. Further, MDPCP—by improving the management of chronic conditions—could prevent the need for some admissions.
Outpatient ED visits and observation stays	↓	The global budget incentives, MDPCP, and population health initiatives could all lower the need for ED visits or shift them to lower-acuity, non-hospital settings (for example, to primary care).
Intensity of hospital care (measured by standardized hospital spending)[a]	↓	This follows the same logic as admissions and ED visits. Further, the standardized spending measure captures non-emergency care in hospital outpatient departments (for example, for elective procedures). Hospital global budgets—which cover both inpatient and outpatient care—also incentivize hospitals to shift non-emergency outpatient care to lower acuity, non-hospital settings.
Domain: Medicare FFS spending		
Total Medicare FFS spending	↓	HSCRC can directly lower spending by lowering hospital global budgets, a main mechanism for achieving total cost of care savings targets. Additional model incentives and supports—including the MPA, MDPCP, and episode programs—aim to improve the efficiency of non-hospital services, lowering non-hospital spending as well.
Hospital spending (inpatient and outpatient)	↓	HSCRC can directly lower growth in all-payer hospital spending by lowering growth in hospital budgets. This is the main lever HSCRC has to ensure it meets commitments on limiting growth in all-payer hospital spending and to help the state meet its total cost of care savings commitments to Medicare.
Non-hospital spending	Not increase by more than hospital spending decreases	The hospital global budgets incentivize hospitals to shift care to lower-acuity, non-hospital settings, which could increase non-hospital spending. Several other model components, however, could decrease non-hospital spending (or offset what would otherwise be larger increases). These components include episode-based programs (ECIP and CTIs) that encourage efficient episodes of care that span both the hospital and non-hospital settings, hospital accountability for the total cost of care (via the MPA), and the MDPCP that could reduce spending on specialty care or duplicative tests.
Total Medicare FFS spending + non-claims payments[a,b]	↓	The state agreement requires Maryland to generate Medicare FFS Part A and B savings after factoring in non-claims-based payments such as MDPCP payments. The model could generate these savings so long as the various model components reduce total Medicare FFS spending enough to fully offset the added non-claims costs to Medicare for supporting the Maryland Model.[c]
Post-acute care spending	↓	The episode-based programs (CTIs and ECIP) could lower post-acute care spending. Further, fewer initial admissions (see above) could prevent the need for post-acute care.
Domain: Quality of care		
Potentially preventable (PQI) admissions	↓	These are admissions that might be prevented with higher quality ambulatory care (for example, admissions for complications from diabetes or for heart failure). Hospitals and their systems could invest in ambulatory care to reduce PQI admissions to perform well under global budgets. Further, the quality adjustments to global budgets directly incentivize reductions in PQIs. In addition, MDPCP investments aim to improve ambulatory care to reduce PQIs.
30-day post-discharge unplanned readmission	↓	Global budgets incentivize hospitals to reduce all admissions, including readmissions. HSCRC's quality programs and episode programs also directly incentivize hospitals to reduce readmission rates. MDPCP encourages greater post-discharge follow-up that could reduce readmissions.

Outcome	Favorable direction of effect	How the Maryland Model could improve this outcome
Timely follow-up after acute exacerbation of chronic conditions	↑	SIHIS set a goal to improve timely follow-up after an ED visit or hospitalization. HSCRC recently incorporated this measure into its quality-based incentive payments to hospitals. Further, the readmission incentives, MDPCP, episode programs, and global budgets all create incentives and opportunities for a range of providers to improve follow-up after discharge.
Patients' rating of their personal doctor (from FFS and MA CAHPS)[d]	↑	Through increased coordination of care, access, or comprehensiveness of care, MDPCP might improve patients' experience with their personal doctor. Further, MDPCP practices are directly incentivized to improve this measure because it factors into whether practices can retain their performance-based incentive payments.
Patients' rating of their hospital (from HCAHPS)[a]	↑	The quality adjustments to hospital global budgets encourage hospitals to improve patients' ratings because those ratings factor into the size of the budget adjustments.
Domain: Population Health		
Use of Medicare Diabetes Prevention Program services	↑	SIHIS set a goal to reduce mean body mass index statewide, and the state is incentivized to reduce the incidence of diabetes through its Outcomes-Based Credits. Maryland expects to achieve the long-term population health improvement, in part, through increased use of Medicare Diabetes Prevention Program services.[e]

[a] Data are only available through 2020 for this report.
[b] The non-claims-based payments include shared savings payments for ACOs, payments for primary care programs (MDPCP within Maryland and CPC+ and CPC classic nationwide), and the 5 percent Part B bonuses for providers participating in advanced alternative payment models (several components of the MD TCOC can qualify providers for these bonuses). CMS includes ACO, MDPCP, and CPC+ payments in its calculation of whether the state meets its annual savings requirements. Per discussions with CMS, CMS is precluded from including 5 percent bonuses in its savings calculations, but would like us to include the 5 percent bonuses in impact estimates so we can fully assess the model's effects on Medicare spending.
[c] The added costs to Medicare to support the Maryland Model are less than the total non-claims based payments for MDPCP, ACOs, and the 5 percent Part B bonuses. We assume that, even absent the Maryland Model, CMS would still be making some non-claims payments in Maryland, which are approximated by non-claims payments in the comparison group.
[d] Data are only available through 2019 for this report.
[e] For example, HSCRC established a Regional Partnership Catalyst Program to hospitals to expand the Medicare Diabetes Prevention Program and improve the statewide health information exchange (CRISP) to permit easier referrals to the program.

ACO = accountable care organization; CAHPS = Consumer Assessment of Healthcare Providers & Systems; CMS = Centers for Medicare & Medicaid Services; CPC+ = Comprehensive Primary Care Plus; CRP = Care Redesign Program; CRISP = Chesapeake Regional Information System for our Patients; CTI = Care Transformation Initiative; ECIP = Episode Care Improvement Program; ED = emergency department; FFS = fee for service; HCAHPS = Hospital Consumer Assessment of Healthcare Providers & Systems; HSCRC = Health Services Cost Review Commission; MA = Medicare Advantage; MDPCP = Maryland Primary Care Program; MPA = Medicare Performance Adjustment; PQI = Prevention Quality Indicator; SIHIS = Statewide Integrated Health Improvement Strategy.

3.2. Overview of impact methods

To estimate the impacts of the evolving Maryland Model, we used a difference-in-differences analysis with a matched comparison group selected from outside Maryland. This approach estimates impacts as the changes in outcomes for Medicare FFS beneficiaries in Maryland over time minus contemporaneous changes for the comparison group. Our underlying assumption is that the changes observed in the comparison group approximate the changes that would have occurred in Maryland absent the Maryland Model. To construct the comparison group, we matched Public Use Microdata Areas (PUMAs). PUMAs are similar to counties, but they break populous counties into smaller units and aggregate sparsely populated counties into larger units. PUMAs are made up of roughly 100,000 people. We matched Maryland's 44 PUMAs with 553 comparison PUMAs drawn from the rest of the country. We matched on basic characteristics of the area (such as demographics), characteristics of the health systems, and outcome levels and trends for Medicare FFS beneficiaries.[2] To permit estimates of model impacts in each year during the full Maryland Model period, we defined the matching variables in 2011 to 2013, before MDAPM began. We estimated impacts in each year (from 2014 to 2021) as the differences in outcomes in Maryland versus the comparison group during that year minus the average differences during the 2011–2013 baseline period. We also contrasted the effects during the MD TCOC period (2019–2021) with the last two years of the MDAPM period (2017 and 2018) to identify how effects during the MD TCOC period compare to a period right before changes made during the MD TCOC period took effect.

Although the focus of this report is the impacts of the Maryland Model during the MD TCOC period, we intentionally chose a baseline period (2011–2013) that predated MDAPM because it allowed us to estimate two different quantities of interest. First, we estimated the effect of the Maryland Model—including global budgets and its newer MD TCOC period innovations in those years—relative to an estimated counterfactual of the path Maryland would have been on had it not adopted any of the changes it did starting in 2014. This estimate is important to be able to capture the effects of the Maryland Model and all its components, including growth in potential effects over time. The second quantity of interest estimates the effects of the Maryland Model during the MD TCOC period minus the effects achieved by the end of the MDAPM period. This second quantity thus represents an estimate of the *improvements* made during the MD TCOC period relative to where MDAPM left off.

It is important to note that we cannot estimate the impact of new components introduced during the MD TCOC period alone. This is because global budgets might continue to have growing effects that we cannot separate from new elements added during the MD TCOC period. Broadly, any changes in impacts from the end of the MDAPM period (2017–2018) to the MD TCOC period could be because of (1) the growing influence of global budgets; (2) the broader accountability, incentives, and supports that Maryland and CMS introduced in 2019 and that continue to evolve; and (3) synergies between the two.

The regression models we used to estimate impacts controlled for beneficiary-level demographic characteristics, such as age, race, and gender; an area-level measure of social vulnerability from the CDC; and a specific set of time-varying chronic health conditions. By time-varying conditions, we mean that we updated a beneficiary’s set of health conditions in each year of the analysis. We chose to include time-varying health conditions because Medicare Advantage enrollment has been increasing faster nationally than it has in Maryland and because we observe that beneficiaries who leave for Medicare Advantage are, on average, healthier than FFS beneficiaries. The lower rates of exit to Medicare Advantage over time in Maryland could make the beneficiaries who remain in the Maryland analytic sample healthier relative to the comparison group for reasons not related to model impacts. Including time-vary health condition controls help to control for those health status differences due to greater movement into Medicare Advantage in the comparison group that could bias estimates of model impacts. In Appendix Section A.6, we detail our rationale for including these conditions in our primary models and report the robustness of our results to models that remove time-varying health conditions.

[2] Rather than match individual PUMAs, we implemented an optimization-based approach called stable balancing weights that seeks to reweight the rest of the country to look like Maryland on specified key characteristics. Matching and weighting, which have similar objectives, are based on similar principles: matching methods select a subset of potential comparison regions to form the comparison group, whereas weighting methods use all comparison regions but give different regions different weights. One benefit of the approach we used is that it allows us to define balance constraints for specific variables based on their relative importance. The 553 PUMAs we identify as part of the comparison group are those that received a non-trivial weight; the other 1,754 PUMAs in the country were poor matches, so they received extremely small weights.

3.3. Accounting for COVID-19 in our estimates

COVID-19 affected health outcomes in all parts of the country, including Maryland, and—especially early in the pandemic—led to significantly reduced use of hospital care and other services. A risk to the Maryland Model's evaluation is that the pandemic might have affected Maryland differently from the comparison-group regions in ways unrelated to the model. More specifically, impact estimates in 2020 or 2021 could be subject to bias if Medicare beneficiaries, including those who do not get COVID-19, respond differently to the pandemic in Maryland versus the comparison group in ways that influence outcomes.

We aimed to limit this risk of bias in four ways. First, we matched on characteristics, such as housing density, that are likely to affect the spread and severity of disease outbreaks (as defined by elements of the Social Vulnerability Index; CDC 2021). Second, we assessed—and confirmed—that drops in hospital utilization rates from 2019 to 2020 were similar in Maryland and the matched comparison group, which suggests that the initial response to COVID-19 in early 2020 was similar in Maryland and our comparison group. Third, we ran sensitivity tests that controlled for the rates of COVID-19 emergency department (ED) visits and hospitalizations rates in 2020 and 2021 in the intervention and comparison groups. We did not include these COVID-19 controls in our main regressions because the Maryland Model could influence the spread and severity of COVID-19 in the state (Haft et al. 2020), so the controls could unintentionally remove a true effect of the model on outcomes (see Appendix C for details). These sensitivity results were generally similar to the main estimates for most outcomes, and they became even more similar in 2021 than they were in 2020. Finally, we ran an additional sensitivity test that changed how we defined health condition flags in 2021. We use the Chronic Condition Data Warehouse chronic condition flags (see Appendix B), which use claims one to three years before to identify conditions. We were concerned that the flags in 2021—which rely on claims from 2018 to 2020—might be understating true condition prevalence because people avoided visiting their providers early in the pandemic (and so they did not generate the claims need to identify conditions). We reestimated impacts assuming the conditions that beneficiaries had in 2021 were the same ones we identified for them in 2020 (based on claims from 2017 to 2019). This sensitivity tests showed modestly larger total savings in 2021 than in our main model and a modestly smaller increase in non-hospital spending (see Section 4.4 and Appendix C for details).

Because of these four approaches for mitigating bias risk from COVID-19, we do not see strong evidence that COVID-19 biases the results in this report in a way that would change the primary conclusions (such as a change from favorable to unfavorable results). Nonetheless, we recommend interpreting estimates in 2020 and 2021 with caution because we cannot fully mitigate all risk of bias to estimates and because true model impacts might be unusually large or small during the pandemic.

4. Results

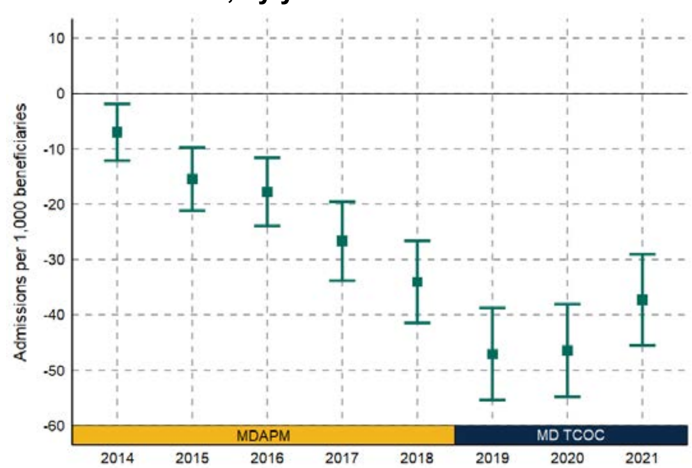
In this section, we report the impacts of the Maryland Model, focusing on the those during the MD TCOC period (2019–2021). For a complete picture, we estimate impacts each year, starting in 2014, relative to an estimate of what would have happened in the state if Maryland and CMS had not made any of the changes they did starting in 2014. We describe the impacts during the MD TCOC period and how they compare with impacts during earlier periods—especially the end of the MDAPM period (2017–2018). For each outcome, we also report impacts on the percentage scale (defined as the impact estimate divided by the unadjusted Maryland mean minus the impact estimate) to help interpret the size of impacts across outcomes. The figures for each outcome show the yearly impact estimates. The tables (one for each domain) show the average impact over the first three years of the MD TCOC period (2019–2021), the average impact from the last two years of the MDAPM period (2017–2018), and the difference between the two. We conclude that the Maryland Model had an impact in a year or period if the impact estimate is statistically different from zero (using a $p < 0.10$ cutoff for significance) and that the Maryland Model had a different impact in the MD TCOC period than the later MDAPM period if the difference is statistically significant. The figures and tables in this section focus on the impact estimates, but Appendix Section A.3 shows the unadjusted mean outcomes over time for the intervention and comparison groups. These trends underlie the impact estimates.

4.1. Impacts on utilization

4.1.1. All-cause acute care hospital admissions

- The Maryland Model reduced hospital admissions by an average of 44 admissions per 1,000 beneficiaries (90% CI: -52, -35; 16.1 percent) in the first three years of the MD TCOC period (2019–2021) (Figure 1 and Table 3).
  - Reductions were similar in 2019 and 2020 and slightly smaller in 2021.
  - Hospitalization rates fell during the entire eight-year period (2014–2021) in Maryland and the comparison group, but rates fell faster in Maryland (see Appendix A, Figure A.3).
- The model decreased hospital admissions in 2019–2021 by about 13 admissions per 1,000 beneficiaries (90% CI: -17, -9; 6.1 percentage points) more than it did at the end of the MDAPM period.
  - Favorable reductions during the MD TCOC period were largely consistent with a trend of increasing reductions that began at the start of the MDAPM period.

Figure 1. Estimated impact of the Maryland Model on all-cause acute care hospital admissions for Medicare FFS beneficiaries, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval; FFS = fee for service.

– These further reductions relative to the end of MDAPM could be attributable to the growing effects of hospital global budgets, new components, or synergies between them.

– The fact that the MD TCOC period estimates were consistent with the general trend downward during MDAPM suggest that at least some, and perhaps much, of the growth in impacts were due to growing effects of hospital global budgets.

Table 3. Estimates of the Maryland Model's effects on health care utilization during the MD TCOC period (2019–2021) and the end of the MDAPM period (2017–2018)

Outcome	Maryland mean during the MD TCOC period (2019–2021)	Estimated impact of the Maryland Model: In the first three years of the MD TCOC period (2019–2021) (90% CI)	Estimated impact of the Maryland Model: In the last two years of the MDAPM period (2017–2018) (90%CI)	Difference[a]
All-cause acute care hospital admissions				
Number per 1,000 beneficiaries per year	229	-44*** (-52; -35)	-30*** (-37; -23)	-13*** (-17; -9)
Percentage impact[b]		-16.1%	-10.0%	-6.1 pp
Outpatient ED visits and observation stays				
Number per 1,000 beneficiaries per year	405	-16*** (-25; -8)	-11** (-19; -3)	-5* (-11; 0)
Percentage impact[b]		-3.8%	-2.2%	-1.6pp
Intensity of hospital care (measured by standardized hospital spending; 2019–2020 only)				
$ per beneficiary per year	$4,731	-$414*** (-$491; -$337)	-$225*** (-$283; -$166)	-$189*** (-$236; -$141)
Percentage impact[b]		-8.0%	-4.5%	-3.5pp

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$

[a] We calculate the difference as the percentage impact in 2019–2021 minus the percentage impact in 2017–2018.

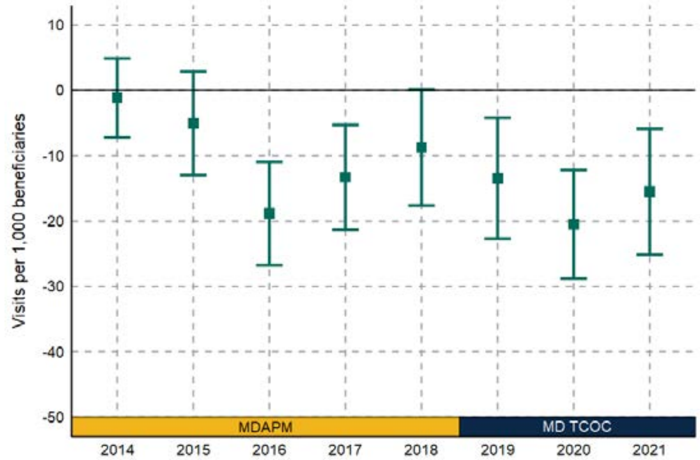
[b] We calculate the percentage impact as the impact estimate divided by the unadjusted Maryland mean minus the impact estimate.

CI = confidence interval; ED = emergency department; pp = percentage point.

4.1.2. Outpatient ED visits and observation stays

- The Maryland Model reduced outpatient ED visits and observations stays by an average of 16 visits per 1,000 beneficiaries (90% CI -25, -8; 3.8 percent) in the first three years of the MD TCOC period (Figure 2 and Table 3).
  - Estimates were similar from 2019 to 2021.
- Reductions in outpatient ED visits and observation stays were 1.6 percentage points larger than they were at the end of the MDAPM period (-5 [90% CI: -11, 0]) (Table 3).
  - Recent trends during the MD TCOC period appear to have reversed an earlier trend towards smaller reductions in ED visits and observations stays estimated at the end of the MDAPM period.

Figure 2. Estimated impact of the Maryland Model on outpatient ED visits and observation stays, by year



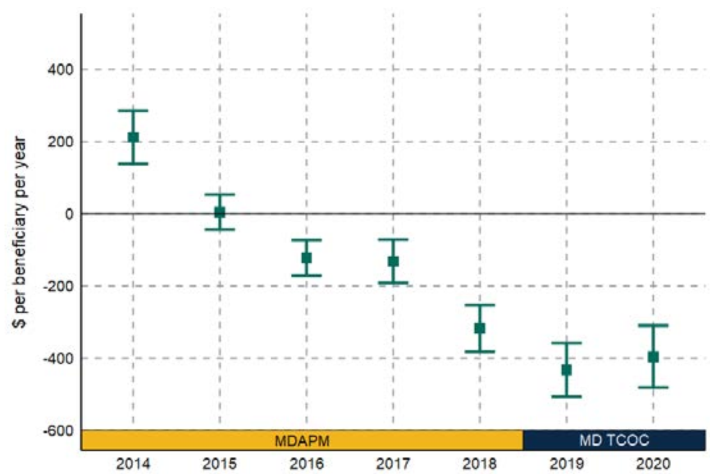
Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval; ED = emergency department.

4.1.3. Intensity of hospital care (measured by standardized hospital spending)

- Standardized hospital spending is an aggregate measure of the intensity of hospital care. It includes spending for inpatient and outpatient care but removes differences in spending across hospitals for reasons other than utilization (for example, the differences in prices set by HSCRC and those set by the Inpatient Prospective Payment System/Outpatient Prospective Payment System [IPPS/OPPS]). We only have standardized spending data through 2020 for this report.
- The Maryland Model reduced standardized hospital spending by \$414 per beneficiary per year (PBPY) (90% CI: -\$491, -\$337; 8.0 percent) in the first two years of the MD TCOC period (Figure 3 and Table 3).

Figure 3. Estimated impact of the Maryland Model on standardized hospital spending, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold. Data for standardized spending were only available through 2020.

CI = confidence interval.

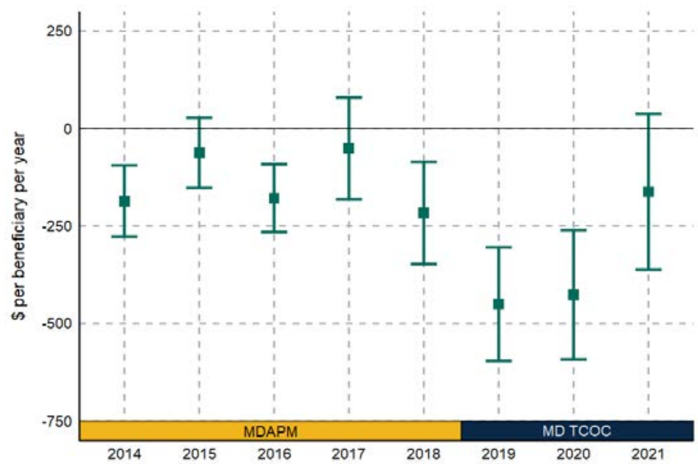
  - Estimates were similar in 2019 and 2020.

- The Maryland Model lowered standardized hospital spending in 2019 and 2020 by about $189 (90% CI: -$236, -$141; 3.5 percentage points) more than it did at the end of the MDAPM period (Table 3).
  - Similar to hospital admissions, reductions in standardized hospital spending in 2019 are consistent with a trend of increasing reductions that began during the MDAPM period.

4.2. Impacts on Medicare FFS spending

4.2.1. Total Medicare FFS spending

Figure 4. Estimated impact of the Maryland Model on total Medicare FFS spending, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval.

- The Maryland Model reduced total Medicare spending by an average of $348 PBPY (90% CI: -$504, -$192; 2.5 percent) in the first three years of the MD TCOC period (Figure 4 and Table 4).
  - Reductions were similar in 2019 and 2020, but noticeably smaller in 2021 (-$162, 90% CI: -$362, $38).
  - The model increased non-hospital spending (Section 4.2.3) but decreased hospital spending by more (Section 4.2.2), leading to a $781 million reduction in total spending from 2019 to 2021.[3]
  - The reduction in total Medicare spending was smaller in 2021 than in 2019 and 2020, largely because the increases in non-hospital spending were larger that year (see Section 4.2.3).

- On average, the Maryland Model decreased total Medicare spending during the MD TCOC period by about $214 PBPY (90% CI: -$306, -$121) more than it did at the end of the MDAPM period (Table 4).

[3] We calculated the total Part A and B spending reduction of $781 million as the impact estimate in each year from 2019 to 2021 (Appendix A.5, Table A.9) multiplied by the total number of beneficiaries (weighted) in Maryland in each year from 2019 to 2021 (Appendix A.3, Table A.6), and summed across all three years. This figure does not account for non-claims spending in those years because data for non-claims spending was only complete through 2020 (see Section 4.2.4).

Table 4. Estimates of the Maryland Model's effects on spending during the MD TCOC period (2019–2021) and the end of the MDAPM period (2017–2018)

		Estimated impact of the Maryland Model		
Outcome	Maryland mean during the MD TCOC period (2019-2021)	In the first three years of the MD TCOC period (2019–2021) (90% CI)	In the last two years of the MDAPM period (2017–2018) (90% CI)	Difference[a]
Total Medicare FFS spending				
$ per beneficiary per year	$13,467	-$348*** (-$504; -$192)	-$134* (-$258; -$9)	-$214*** (-$306; -$121)
Percentage impact[b]		-2.5%	-1.0%	-1.5pp
Hospital spending (inpatient and outpatient)				
$ per beneficiary per year	$7,262	-$510*** (-$633; -$387)	-$341*** (-$434; -$248)	-$169*** (-$241; -$97)
Percentage impact[b]		-6.6%	-4.6%	-2.0pp
Non-hospital spending				
$ per beneficiary per year	$6,205	$162*** ($95; $229)	$207*** ($148; $265)	-$45* (-$84; -$5)
Percentage impact[b]		2.7%	3.7%	-1.0pp
Total Medicare FFS spending + non-claims payments				
$ per beneficiary per year (2019–2020 only)	$13,300[c]	-$400*** (-$547; -$252)	-$157** (-$280; -$34)	-$243*** (-$326; -$159)
Percentage impact[b]		-2.9%	-1.2%	-1.7pp
Post-acute care spending				
$ per beneficiary per year	$996	-$76*** (-$108; -$44)	-$30** (-$55; -$6)	-$46*** (-$66; -$25)
Percentage impact[b]		-7.1%	-2.8%	-4.3pp

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$.

[a] We calculate the difference as the percentage impact in 2019–2021 minus the percentage impact in 2017–2018. For Total FFS spending + non-claims payments, which is measured through 2020, we calculate the difference as the percentage impact in 2019–2020 minus the percentage impact in 2017–2018.

[b] We calculate the percentage impact as the impact estimate divided by the unadjusted Maryland mean minus the impact estimate.

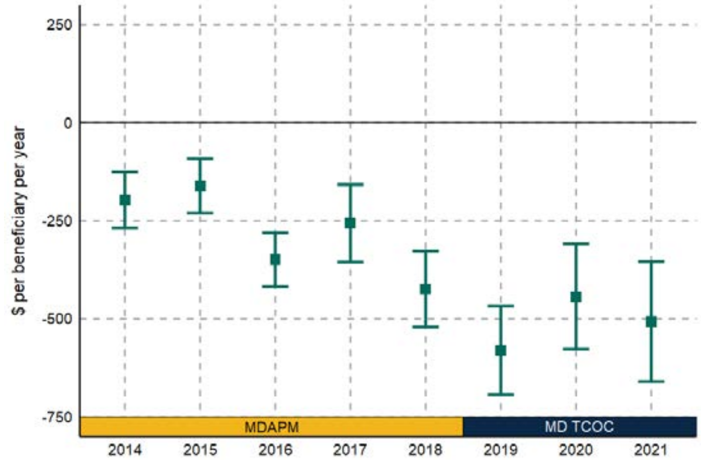
[c] The average yearly total spending including non-claims payments in 2019 and 2020 ($13,300) is lower than the total spending without non-claims payments in 2019–2021 ($13,467) because of the differences in years covered. The total spending without non-claims payments in 2019 and 2020 was $13,159, indicating that the non-claims payments added an average of $141 per beneficiary per year in 2019 and 2020.

CI = confidence interval; FFS = fee for service; pp = percentage point.

4.2.2. Hospital spending (inpatient and outpatient)

- The Maryland Model reduced hospital spending by an average of $510 PBPY (90% CI: -$633; -$387; 6.6 percent) in the first three years of the MD TCOC period (Figure 5 and Table 4).
  - Reductions in hospital spending were driven by substantially slower growth in hospital spending in Maryland than in the comparison group (Appendix Section A.3).
- The Maryland Model reduced hospital spending during the MD TCOC period by about $169 PBPY (90% CI: -$241, -$97; 2.0 percentage points) more than it did at the end of the MDAPM period (Table 4).
- The Maryland Model reduced hospital spending in 2020 and 2021 by less than it did in 2019, which might be partly attributable to how global budgets operated during the COVID-19 pandemic.[4]
  - In 2020, global budgets protected hospitals in Maryland from drops in revenue that occurred elsewhere in the country as hospital volumes dropped because of COVID-19 (Levy et al. 2021). Supporting this finding, we estimated that the model continued to reduce all-cause admissions and standardized spending by similar amounts in 2020 as in 2019 (Figures 1 and 3), suggesting smaller reductions in hospital spending in 2020 were related to higher prices per hospital stay.
  - In 2021, some hospitals continued to charge higher rates to recoup their full budget from 2020 (HSCRC 2022a). This means that, because of limits in how much hospitals could increase their prices (even limits that were relaxed as part of COVID-19 policies in 2020 to compensate for lower volume in 2020), some hospitals did not receive their full budget allocation in 2020, so they could continue to charge higher rates in 2021 to recoup the difference. This might have led to hospital spending levels that were higher than they would have been under the 2021 budget allocation alone.

Figure 5. Estimated impact of the Maryland Model on hospital spending (inpatient and outpatient), by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

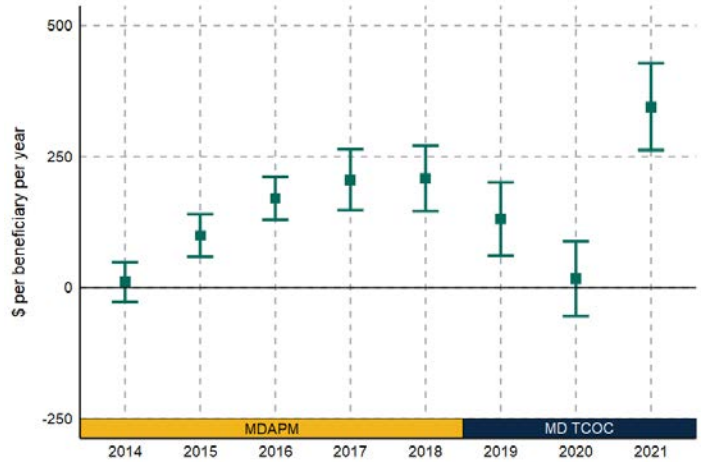
CI = confidence interval.

[4] Our measure of hospital spending does not include federal Coronavirus Aid, Relief, and Economic Security (CARES) Act funding that went to hospitals in Maryland or the comparison group.

4.2.3. Non-hospital spending

- The Maryland Model increased non-hospital spending by an average of $162 PBPY (90% CI: $95; $229; 2.7 percent) in the first three years of the MD TCOC period (Figure 6 and Table 4).
  - In 2019, the Maryland Model only modestly increased non-hospital spending ($131), and, in 2020, the model did not measurably increase spending at all.
  - But in 2021, the model increased non-hospital spending more substantially, by $345 (90% CI: $262; $428). Non-hospital spending grew substantially from 2020 to 2021 in Maryland and the comparison group, as utilization recovered from drops early in the COVID-19 pandemic. But non-hospital spending grew by more in Maryland (Appendix Section A.3).

Figure 6. Estimated impact of the Maryland Model on non-hospital spending, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

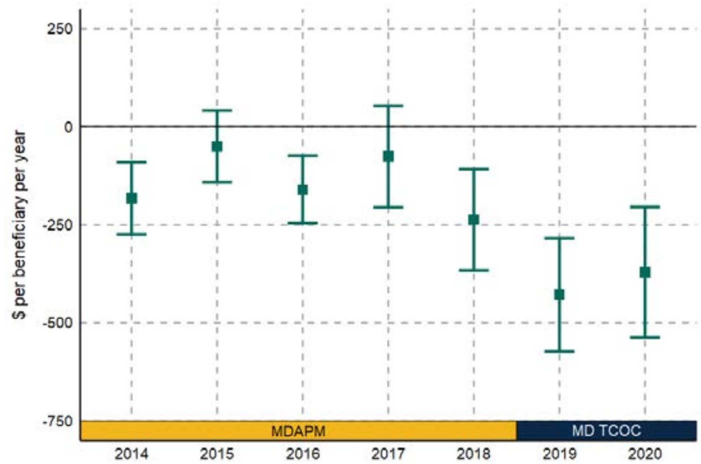
CI = confidence interval.

- On average, over the three-year MD TCOC period, the Maryland Model increased non-hospital spending by about $45 PBPY (90% CI: $84; $5, 1.0 percentage points) *less* than it did at the end of the MDAPM period, which was a favorable result.
  - But in 2021, the model increased non-hospital spending by $138 PBPY (90% CI: $72; $204) *more* than it did at the end of MDAPM, driving smaller (less favorable) reductions in total spending (section 4.2.1)
  - Results from 2022 and beyond will help determine whether 2021 impacts represent an aberration or part of a new trend.

4.2.4. Total Medicare FFS spending + non-claims payments

- Non-claims payments include payments for MDPCP, as well as national primary care programs such as the Comprehensive Primary Care and Comprehensive Primary Care Plus initiatives, accountable care organization, and bonuses to providers for participating in advanced payment models. We only have data on non-claims payments through 2020.
- The Maryland Model reduced total Medicare spending, including non-claims-based payments, by $400 PBPY (90% CI: -$547, -$252; -2.9 percent) in the first two years of the MD TCOC period (Figure 7 and Table 4).
  - Estimates are similar to spending estimates that do not account for non-claims-based payments, suggesting that these additional payments do not change the conclusion that the model reduced total Medicare spending.
  - Impact estimates are similar with and without the non-claims-based payments because, relative to total Medicare spending, the non-claims-based payments were relatively small in Maryland in 2019 and 2020 (at $104 and $183 per Medicare beneficiary, respectively) and because the comparison group spending also increased by a similar amount ($87 and $129) when we include non-claims-based payments for alternative payments models.

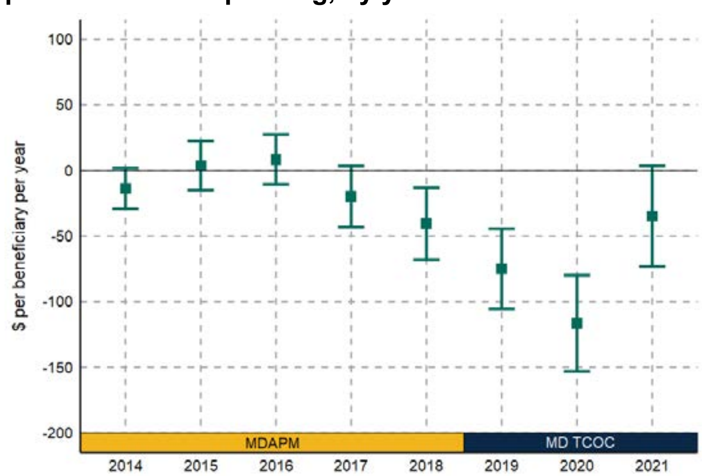
Figure 7. Estimated impact of the Maryland Model on total Medicare FFS spending + non-claims payments, by year (data through 2020 only)



Notes: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold. Data for standardized spending were only available through 2020.

CI = confidence interval.

Figure 8. Estimated impact of the Maryland Model on post-acute care spending, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval.

4.2.5. Post-acute care spending

- The Maryland Model reduced post-acute care spending by $76 PBPY (90% CI: -$108, -$44; -7.1 percent) during MD TCOC period (Figure 8 and Table 4).

- Estimates were trending toward larger reductions in post-acute care spending largely because of a relatively flat trend in post-acute care in our comparison group alongside consistent reductions in post-acute care spending in Maryland. In 2020, post-acute care spending dipped sharply in Maryland before rebounding in 2021, which explains the pattern in impacts (See Appendix A, Figure A.2, panel E).
- Similar to non-hospital spending, impacts on post-acute care spending in 2021 were substantially different from earlier years during the MD TCOC period and not significantly different from the end of the MDAPM period (-$35 PBPY, 90% CI: -$73, $4). Post-acute care spending is a subset of non-hospital spending, but the smaller reduction in post-acute care spending in 2021 only represents about a quarter of the increase in non-hospital spending we observed from 2020 to 2021, suggesting that other factors contributed to increases in non-hospital spending in 2021. Future reports may explore the specific categories of non-hospital spending that increased under the model.

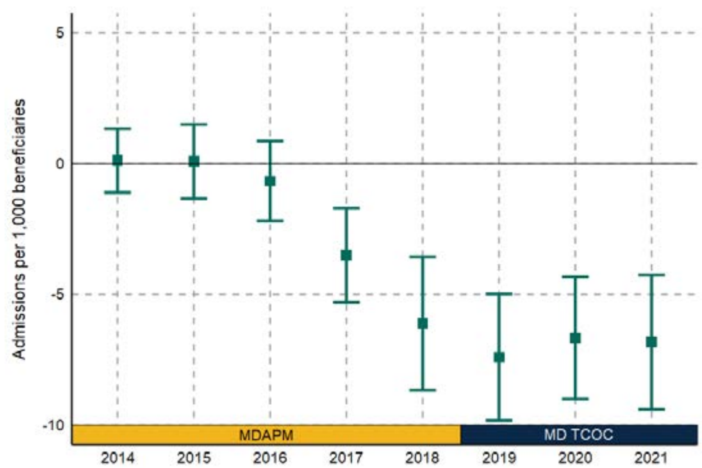
- Despite smaller reductions in 2021, on average, during the three-year MD TCOC period, the Maryland Model reduced post-acute care spending by about $46 PBPY (90% CI: -$66, -$25) more than it did at the end of the MDAPM period.

4.3. Impacts on quality and population health

4.3.1. Potentially preventable admissions

- The Maryland Model reduced potentially preventable admissions by 7.0 admissions per 1,000 beneficiaries (90% CI: -9.3, -4.6; -16.1 percent) in the first three years of the MD TCOC period (Figure 9 and Table 5).
  - Estimates were similar in all three years of the MD TCOC period.
- The reductions in potentially preventable admissions during the MD TCOC period were 2.1 admissions per 1,000 beneficiaries (90% CI: -3.1, -1.2) larger than they were at the end of the MDAPM period (Table 5).
  - Reductions in potentially preventable admissions were consistent with a trend of increasing reduction that began in the middle of the MDAPM period.

Figure 9. Estimated impact of the Maryland Model on potentially preventable admissions, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval.

Table 5. Estimates of the Maryland Model's effects on quality and population health during the MD TCOC period (2019–2021) and the end of the MDAPM period (2017–2018)

Outcome	Maryland mean during the MD TCOC period (2019-2021)	Estimated impact of the Maryland Model: In the first three years of the MD TCOC period (2019–2021) (90% CI)	Estimated impact of the Maryland Model: In the last two years of the MDAPM period (2017–2018) (90%CI)	Difference[a]
Potentially preventable admissions				
Number per 1,000 beneficiaries per year	36.6	-7.0*** (-9.3; -4.6)	-4.8*** (-6.9; -2.7)	-2.1*** (-3.1; -1.2)
Percentage impact[b]		-16.1%	-9.2%	-6.9pp
30-day post-discharge unplanned readmissions				
Percentage of discharges	16.2%	-1.7pp*** (-2.0; -1.4)	-1.4pp*** (-1.7; -1.1)	-0.3pp** (-0.5; -0.1)
Percentage impact[b]		-9.5%	-7.9%	-1.6pp
Timely follow-up after acute exacerbation of chronic conditions				
Percentage of discharges	69.1%	1.7pp*** (1.1; 2.3)	1.4pp*** (0.8; 2.0)	0.3pp (-0.1; 0.7)
Percentage impact[b]		2.5%	2.0%	0.5pp
Patients' rating of their personal doctor[c]				
Rating on a scale from 0 to 100	90.9	0.4 (-0.3; 1.0)	0.1 (-0.5; 0.7)	0.3 (-0.5; 1.1)
Percentage impact[b]		0.4%	0.1%	0.3pp
Patients' rating of their hospital[d]				
Percent with a score of 9 or 10 out of 10	67.8%	0.5pp (-1.0; 2.0)	0.6pp (-0.7; 1.9)	-0.1pp (-0.9; 0.6)
Percentage impact[b]		0.7%	0.9%	-0.2pp
Use of Medicare Diabetes Prevention Program services				
Number of beneficiaries using services per 10,000 beneficiaries	0.2	-0.4** (-0.7; -0.1)	-0.9*** (-1.4; -0.5)	0.5* (0.1; 1.0)
Percentage impact[e]		Not shown	Not shown	Not shown

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$.

[a] We calculate the difference as the percentage impact in 2019–2021 minus the percentage impact in 2017–2018.

[b] We calculate the percentage impact as the impact estimate divided by the unadjusted Maryland mean minus the impact estimate.

[c] Data are only available through 2019 for this report

[d] Data are only available through 2020 for this report

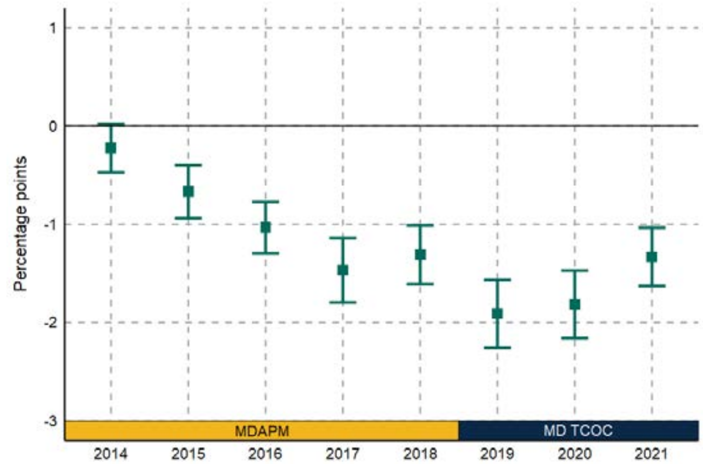
[e] We do not report the percentage impact for use of Diabetes Prevention Program services because rates are so low that dividing by small such numbers creates unstable percentage impacts that look misleading.

CI = confidence interval; pp= percentage point.

4.3.2. 30-day post-discharge unplanned readmissions

- The Maryland Model reduced the probability of 30-day unplanned readmissions by 1.7 percentage points (90% CI: -2.0, -1.4; -9.5 percent) in the first three years of the MD TCOC period (Figure 10 and Table 5).
  - Reductions were similar in 2019 and 2020, and smaller in 2021.
- On average, the model reduced readmissions during the MD TCOC period by about 0.3 percentage points (90% CI: -0.5, -0.1) more than it did at the end of the MDAPM period (Table 5).
  - The reductions in readmissions in 2019 were consistent with a favorable trend of increasing reductions that began with the start of MDAPM in 2014 (Figure 10). But reductions in 2021 were very similar to reductions at the end of the MDAPM period.

Figure 10. Estimated impact of the Maryland Model on 30-day post-discharge unplanned readmissions, by year



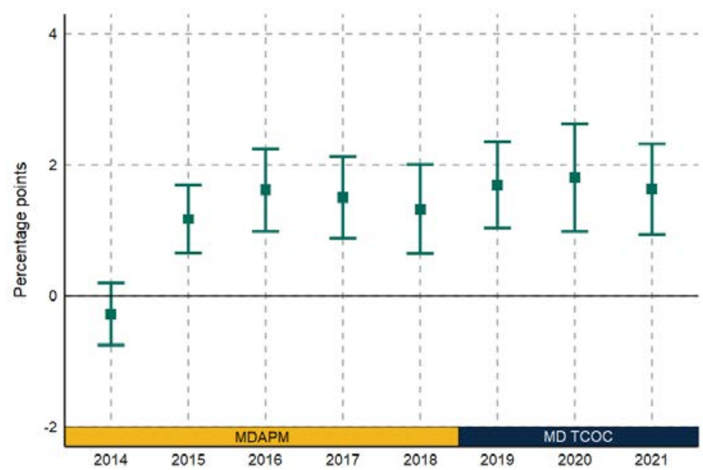
Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval.

4.3.3. Timely follow-up after acute exacerbation of chronic conditions

- The Maryland Model increased the probability of follow-up after acute exacerbation by 1.7 percentage points (90% CI: 1.1, 2.3; 2.5 percent) in the first three years of the MD TCOC period (Figure 11 and Table 5), which was a favorable result.
  - Estimates were similar in all three years of the MD TCOC period.
- The Maryland Model increased the probability of follow-up in 2019 and 2020 by nearly the same amount as it did at the end of the MDAPM period, suggesting that the Maryland Model has so far shown little additional improvement on follow-up after acute exacerbation above what was achieved during MDAPM (Table 5).

Figure 11. Estimated impact of the Maryland Model on timely follow-up after acute exacerbation of chronic conditions, by year



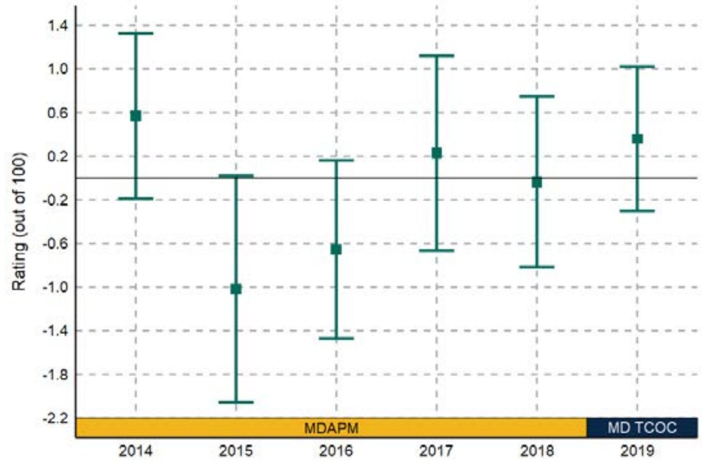
Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval.

4.3.4. Patients’ rating of their personal doctor

- In 2019, the Maryland Model had no measurable effects on patients’ rating of their personal doctor according to ratings from FFS and Medicare Advantage beneficiaries collected via the FFS and Medicare Advantage Consumer Assessment of Healthcare Providers & Systems (CAHPS) surveys (0.4 unit increase on a scale from 0 to 100, 90% CI: -0.3, 1.0, 0.4 percent) (Figure 12 and Table 5).
  - Patients generally rated their primary doctors highly in Maryland and the comparison group, with a mean rating of about 90 percent in both groups in 2019 (Appendix Figure A.4). Therefore the lack of measured impacts might be, in part, due to relatively little room for improvement on this measure.
  - Lack of measured impacts might also be because the survey is not limited to beneficiaries seen by an MDPCP practice (one of the primary expected mechanisms for improvement), which could make impacts harder to detect.
  - Data were unavailable for 2020 (because of COVID-19) and 2021 (unreleased).
  - This measure is limited to beneficiaries who, on the CAHPS surveys, said they have a personal doctor and had visited that doctor at least once in the past six months.
- Effects in 2019 were similar to effects during the MDAPM period, when no impacts on patients’ rating of their personal doctor were expected.

Figure 12. Estimated impact of the Maryland Model on patients’ rating of their personal doctor, by year



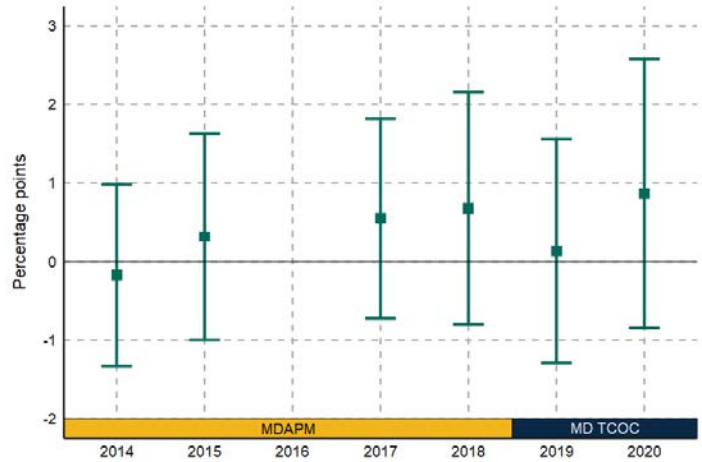
Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold. Data were only available through 2019.

CI = confidence interval.

4.3.5. Patients' rating of their hospital

- Relative to the comparison group, the Maryland Model also did not improve patients rating of their hospital collected via Hospital CAHPS in 2019 and 2020 (0.5pp increase in the percentage of patients giving an overall hospital rating of 9 or 10 out of 10, 90% CI: -1.0, 2.0, 0.7 percent) (Figure 13 and Table 5). Hospital rating data was only available through 2020.
  - Maryland ratings were slightly lower (66.6 percent of patients giving an overall hospital rating of 9 or 10 out of 10) than the comparison group (67.8 percent) throughout the three-year baseline period (2011–2013) (Appendix Figure A.4).
  - In both Maryland and our comparison group, hospital ratings improved from 2011 to 2017 and then began to decline (Appendix Figure A.4). A similar rate of improvement and decline led to no measurable impacts in the MDAPM or MD TCOC periods.

Figure 13. Estimated impact of the Maryland Model on patients' rating of their hospital, by year



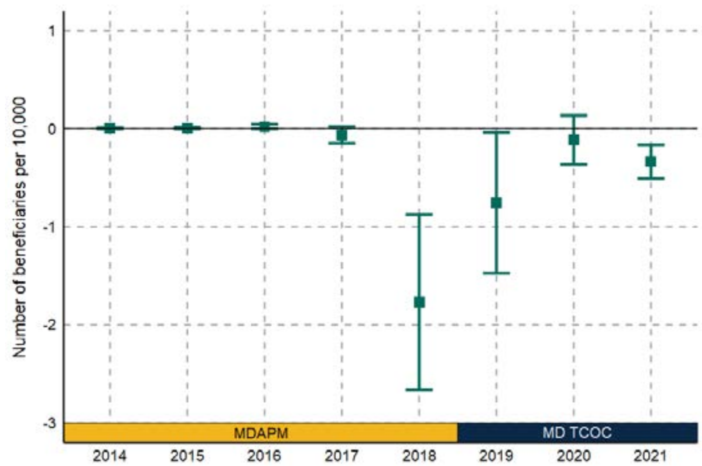
Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold. 2016 was excluded from analyses because several large hospitals in Maryland did not report scores in that year, potentially skewing results. Data were only available through 2020.

CI = confidence interval.

4.3.6. Use of Medicare Diabetes Prevention Program Services

- During the TCOC period, the Maryland Model appeared to have slightly reduced the use of Medicare Diabetes Prevention Program (DDP) services (-0.4 per 10,000 beneficiaries, 90% CI: -0.7, -0.1) (Figure 14 and Table 5).
  - According to Medicare FFS claims, use of this program is extremely low in Maryland and the comparison group, peaking at about 3 per 10,000 beneficiaries in the comparison group in 2018 (the first full year of the program) and 1 per 10,000 beneficiaries in Maryland (Appendix Figure A.4). By 2021,

Figure 14. Estimated impact of the Maryland Model on use of Medicare Diabetes Prevention Program services, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold. The impact estimates in 2014 to 2016 are zero because Diabetes Prevention Program Services did not become a funded benefit until 2017.

CI = confidence interval.

rates were much lower (0.4 and 0.1 per 10,000 in the comparison group and Maryland, respectively).

  - This impact estimate suggests that the model reduced the total number of Medicare FFS beneficiaries throughout the state using DPP services during the MD TCOC period by about 30 people (out of about 750,000 Medicare FFS beneficiaries in Maryland).
  - Because the number of beneficiaries receiving DDP services is so small, this estimate might be spurious, reflecting small differences in the use of DPP services in Maryland and the comparison group that are unrelated to model impact. Conversely, it's possible that the Maryland Model has unintentionally suppressed the use of (or billing for) these services, which only became a funded benefit in Medicare in 2018.
- Because of the timing of model incentives, it is unsurprising that we do not see a favorable impact on DPP services by 2021. In 2021, HSCRC awarded $86 million in Regional Partnership Catalyst Program funding to hospitals and their partners to help prevent and manage diabetes, including increasing the supply of DPP services (HSCRC n.d.). This funding aims to slowly ramp up the percentage of Medicare beneficiaries with pre-diabetes living in awardees' service areas who receive DPP services to 7 percent over five years. HSCRC did not, however, anticipate any increases in 2021 since the funding was newly awarded.

4.4. Sensitivity tests

- The primary impact estimates control for time-varying health conditions to account for movement into Medicare Advantage (and out of FFS) that is happening more frequently in the comparison group than in Maryland over time. But, if the Maryland Model itself affected the rates of health conditions, controlling for these conditions could remove some of the model's effects. Sensitivity analyses that remove these controls were qualitatively consistent with the main impact findings, though moderately larger in all years. We believe the risk of bias from a changing population because of Medicare Advantage enrollment outweighs the risk of overcontrolling for the Maryland Model's impacts (see Appendix Section A.6 for more details on this decision). Therefore, we believe the main results presented above provide the best estimate of model effects, but we have included results without these controls in Appendix Section A.6.

- The COVID-19 pandemic has had a profound impact on all aspects of health care use and delivery since early 2020. We took several steps to mitigate bias in our impact estimates because of COVID-19, including in matching, and in our regression models (see Appendix C for additional details). In general, we do not find evidence that COVID-19 directly affected our results in ways that change our conclusions about the effects of the Maryland Model in 2020 and 2021.

  - Our main regression models do not control for COVID-19-related variables during the MD TCOC period because of the concern that the Maryland Model could have affected the rates or severity of COVID-19 (for example, through outreach or other support programs). As a sensitivity test, we ran separate regression models that do control for the direct effects of COVID-19 by including beneficiary-level indicators for a hospitalization or ED visits for COVID-19. These models generally show consistent results with models that do not control for COVID-19, suggesting that the direct effect of COVID-19 exposure does not play a large role in explaining impacts in 2020 or 2021.

– We observed declines in claims-based health diagnoses measured during COVID-19. Those declines are most likely because beneficiaries avoided care during the pandemic (and thus do not show up in claims). We tested whether our results were sensitive to these changes in health condition controls by replacing condition values in 2021 (that were largely based on claims in 2019–2020) with condition values as they were defined in 2020 (based on claims in 2018–2019). Results were largely consistent with our primary models, though spending measure estimates were moderately larger (but not as large as removing health conditions altogether). Appendix C provides more details.

5. Discussion

5.1. Summary of findings in relationship to model goals and logic

The Maryland Model had significant, favorable effects on utilization, spending, and quality-of-care outcomes during the MD TCOC period for most (11 of 14) of the outcomes listed in Table 2. The largest of these was a reduction in hospital admissions (16.1 percent), in which global budgets create a strong incentive for hospitals to reduce hospital care. Overall, we estimate that the model reduced Medicare Part A and B spending by $781 million (or 2.5 percent on average per year) for Medicare FFS beneficiaries in Maryland from 2019 to 2021. In other words, the amount Medicare actually spent from 2019 to 2021 for Part A and B services in Maryland was $781 million less than what we estimate Medicare would have spent on Part A and B services in those years if Maryland and CMS had not made of the changes they did starting with MDAPM in 2014. The model reduced total Medicare spending in 2019 and 2020 even after accounting for non-claims-based payments, which includes payments for MDPCP, accountable care organizations, and bonuses to providers for participating in advanced payment models in Maryland (and similar initiatives nationwide).[5] In addition to reducing Medicare spending for hospital care, the Maryland Model reduced standardized hospital spending, an aggregate measure of hospital use. This pattern indicates not only that Maryland, via HSCRC's rate setting, limited the growth in hospital budgets but also that the Maryland Model successfully incentivized hospital volume reduction, which should help hospitals improve or maintain their margins despite limits on global budget growth.

The Maryland Model increased non-hospital spending during the MD TCOC period, but we still consider this effect favorable overall because the increase was smaller than the decrease in hospital spending. The model creates strong incentives for hospitals to shift care to lower acuity, non-hospital settings, which, all else equal, will increase non-hospital spending. The results indicate that the model increased non-hospital spending during the MD TCOC period (2019–2021) by almost 3 percent. The reductions in hospital spending, however, were larger than the increases in non-hospital spending, leading to the modest savings to Medicare in total FFS spending. Although we did not investigate most individual categories leading to the increase in non-hospital spending, it's clear that post-acute care spending is not the root cause. We found that the model decreased post-acute care spending (including spending on skilled nursing facilities and Part A home health services), which could be because of incentives designed to improve the efficiency of episodes of care, a decline in admissions that trigger the need for post-acute care, or both.

For the remaining three outcomes in Table 2, we did not find evidence of favorable effects during the MD TCOC period.

[5] Because of lags in data availability, the impact estimates with non-claims-based payments does not include 2021. Similarly, the $781 million aggregate savings estimate from 2019 to 2021 does not include non-claims-based payments.

- The Maryland Model did not affect patients' rating of their personal doctor. We hypothesized that MDPCP would increase patients' rating of their personal doctor. The lack of effects might be because of patients generally being satisfied with their personal doctor (leaving less room for improvement); the relatively short follow-up period (improvement in patients' rating might take more time); or the fact that we're measuring this outcome among all Medicare FFS beneficiaries in the state but, to date, MDPCP has reached about half of these beneficiaries (Machta et al. 2021).
- The Maryland Model has not improved patients' rating of their hospital care. Hospital ratings are modestly lower in Maryland than they are in the nation, and Maryland has been working to improve hospital ratings. Through 2021, these efforts have not translated into statistically significant improvements relative to the comparison group. At the same time, the Maryland Model has not decreased patients' rating of their hospitals, helping to alleviate a potential concern that global budgets (by removing FFS incentives to attract patients) would decrease patients' rating of their hospitals.
- The Maryland Model slightly reduced the use of DPP services, as measured in claims.[6] Because the use of these services in the intervention and comparison groups is low, however, the impact estimate is very small on an absolute scale (reducing the probability that a Medicare FFS beneficiary in the state would use these services by less than 0.5 per 10,000 beneficiaries). Because of the small estimated model impact, this result does not raise concerns that the model is harming access to DPP services. Further, the very low rates of use of DPP services in Maryland (and elsewhere) indicate substantial room for improvement in future model years.

For most (10 of 14) outcomes, the Maryland Model not only had favorable effects during the MD TCOC period but also had effects that were larger and more favorable than they were at the end of the MDAPM period (2017–2018), indicating further improvements. These further improvements included outcomes in three domains: service use (for example, all-cause admissions), spending (for example, total Medicare spending), and quality of care (for example, unplanned readmissions). The growth in impacts could be due to the growing effects of hospital global budgets that began with MDAPM in 2014, the new components added with MD TCOC in 2019, or synergies between them. In general, it is not possible to separate the effects of global budgets from other model components active during the MD TCOC period. That said, comparing effects from the MD TCOC period with effects from earlier periods and following trends in impacts can give us a sense of whether impacts during the MD TCOC period are likely to be strongly influenced by the growing effects of global budgets or other elements. For example, trends in impacts in all-cause admissions suggest impacts during the MD TCOC period are plausibly related to a continuing trend that began during the MDAPM period.

The finding that the Maryland Model increased non-hospital spending substantially in 2021 illustrates a potential risk in the model, and it will be important to examine this outcome in future years. Under the Maryland Model, the state can directly control hospital spending through its global budgets. Therefore, HSCRC can largely ensure the state limits hospital spending growth to be at, or below, the national growth rate. Maryland cannot, however, control non-hospital spending directly. And although the state has introduced a variety of incentives and supports to limit growth in non-hospital spending, these incentives to date have been relatively more modest than global budget incentives (Machta et al. 2021). The global budgets create incentives to increase non-hospital spending (for example, shifting care from

[6] Providers in Maryland and in the comparison group may be providing some DPP services to Medicare beneficiaries without billing Medicare. Indeed, the Regional Partnership funding is designed, in part, to help providers build capacity in both providing and billing for DPP services.

the hospital to non-hospital settings). The magnitude of the increase in non-hospital spending in 2021, if it continued to grow, could represent a risk to the Maryland Model's goal of reducing the total cost of care. Future reports will help determine whether effects from 2021 are an aberration or part of a new trend.

5.2. Limitations

The analysis has several important limitations. First, though we attempted to mitigate the risk of bias from COVID-19, we cannot rule out that the COVID-19 pandemic's indirect effects, such as behavioral responses or disruptions to care or social supports, could have affected Maryland and the comparison group differently, potentially leading to some mismeasurement or bias in our results. Second, because the Maryland Model is implemented across the state, we drew our comparison group from outside the state. We achieved good balance between Maryland and our selected comparison group on baseline characteristics and outcomes, but no comparison group can be a perfect representation of how Maryland would have evolved without implementation of MDAPM or MD TCOC. For example, Maryland's all-payer rate-setting system that dates back to the 1970s might have influenced outcome trends in the state in ways not fully captured by the comparison group. Third, we cannot estimate the effects of the new components added to the Maryland Model in 2019 separate from the growing effects of global budgets that began in 2014. Some stakeholders might want to know the effects of just these new components. Fourth, we cannot measure some of the important outcomes for the evaluation well in Medicare FFS claims, such as SIHIS goals to reduce mean body mass index among all Maryland adults or to reduce overdose deaths. Finally, the analysis does not break out non-hospital spending into categories to identify which types of non-hospital spending have increased most in 2021. From the results in this report, we know that reductions in post-acute care spending (Section 4.2.5) were considerably smaller (less favorable) in 2021 than in 2019 or 2020, which could be part, but not most, of what we observe as an increase in all non-hospital spending. We might explore additional categories of non-hospital spending in future reports.

5.3. Conclusion

Overall, the Maryland Model in the first three years of the MD TCOC period (2019–2021) improved many of the targeted outcomes and did so more than the improvements observed at the end of the MDAPM period. The further improvements during the MD TCOC period could be due to growing effects of global budgets, the broader accountability and incentives introduced in 2019, and synergies between the two. Notably, the Maryland Model substantially increased non-hospital spending in 2021 despite state accountability and corresponding model incentives to hospitals and other providers to reduce the total cost of care. The evaluation will continue to examine impacts on non-hospital and total spending in future years to see whether 2021 was an aberration or part of a new trend.

References

Agency for Healthcare Research and Quality (AHRQ). "Prevention Quality Indicators Overview." n.d. Available at https://www.qualityindicators.ahrq.gov/modules/pqi_resources.aspx#techspecs. Accessed September 2021.

Centers for Disease Control and Prevention (CDC). "CDC/ATSDR Social Vulnerability Index." Washington, DC: CDC, 2021. Available at https://www.atsdr.cdc.gov/placeandhealth/svi/index.html.

CDC. "COVID-19 Vaccination Trends in the United States, National and Jurisdictional." CDC, 2022. Available at: https://data.cdc.gov/Vaccinations/COVID-19-Vaccination-Trends-in-the-United-States-N/rh2h-3yt2.

Centers for Medicare & Medicaid Services (CMS). "Maryland Total Cost of Care Model State Agreement." Baltimore, MD: CMS, 2018.

CMS. "Hospital Compare." Baltimore, MD: CMS. Updated December 1, 2021. Available at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalCompare. Accessed August 2022.

CMS. "Restructured BETOS Classification System." Baltimore, MD: CMS. Updated May 19, 2022. Available at https://data.cms.gov/provider-summary-by-type-of-service/provider-service-classifications/restructured-betos-classification-system. Accessed August 2022.

CMS. "Acute Inpatient PPS." Baltimore, MD: CMS. Updated August 2, 2022. Available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS. Accessed August 2022.

CMS. "Case Mix Index." Baltimore, MD: CMS, n.d. Available at https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Acute-Inpatient-Files-for-Download-Items/CMS022630. Accessed August 2022.

Chattopadhyay, Ambarish, Christopher H. Hase, and José R. Zubizarreta. "Balancing vs Modeling Approaches to Weighting in Practice." *Statistics in Medicine*, vol. 39, no. 24, October 2020, pp. 3227–3254.

Daw, Jamie R., and Laura A. Hatfield. "Matching and Regression to the Mean in Difference-in-Differences Analysis." *Health Services Research,* vol. 53, no. 6, 2018, pp. 4138–4156. doi:10.1111/1475-6773.12993.

Haft, Howard M., Chad Perman, and Eli Y. Adashi. "The Maryland Primary Care Program—A Blueprint for the Nation?" *JAMA,* vol. 1, no 10, 2020, e201326. Available at https://jamanetwork.com/journals/jama-health-forum/fullarticle/2772314.

Health Services Cost Review Commission (HSCRC). "Statewide Integrated Health Improvement Strategy Proposal." Baltimore, MD: HSCRC, December 2020a. Available at https://hscrc.maryland.gov/Documents/Modernization/SIHIS%20Proposal%20-%20CMMI%20Submission%2012142020.pdf.

HSCRC. "Medicare Advantage Partnership Grant Program." Baltimore, MD: HSCRC, March 2020b. Available at https://hscrc.maryland.gov/Documents/Modernization/Medicare%20Advantage%20Grants/Medicare%20Advantage%20Partnership%20Grant%20Program%20RFP%20-%20Final4%203-13-2020.pdf.

HSCRC. "Corridor Expansion Jan-June 2022." Baltimore, MD: HSCRC, February 4, 2022a. Available at https://hscrc.maryland.gov/Documents/pdr/PolicyClarification/2021/2022.02.04%20_%20Corridor%20Expansion%20Jan%20-%20Jun%202022.pdf?csf=1&e=LgaFOf. Accessed August 2022.

HSCRC. " Statewide Integrated Health Improvement Strategy Annual Report." Baltimore, MD: HSCRC, January 2022b. Available at https://hscrc.maryland.gov/Documents/Modernization/Statewide%20Integrated%20Health%20Improvement%20Strategy/SIHIS%202021%20Annual%20Report%20FINAL%20w%20appendix.pdf. Accessed November 3, 2022.

HSCRC. "Regional Partnership Catalyst Program." Baltimore, MD: HSCRC, n.d. Available at https://hscrc.maryland.gov/Pages/regional-partnerships.aspx. Accessed August 2022.

Hospital Consumer Assessment of Healthcare Providers and Systems. "Mode & Patient-Mix Adj." HCAHPS. Updated August 2022. Available at https://hcahpsonline.org/en/mode--patient-mix-adj/. Accessed August 2022.

IMPAQ Health. "Timely Follow-Up After Acute Exacerbations of Chronic Conditions." 2018. Available at https://impaqint.com/measure-information-timely-follow-after-acute-exacerbations-chronic-conditions. Accessed September 2021.

Levy, J.F., B.N. Ippolito, and A. Jain. "Hospital Revenue Under Maryland's Total Cost of Care Model During the COVID-19 Pandemic, March-July 2020." *JAMA*, vol. 325, no. 4, 2021, pp. 398–400. Available at https://jamanetwork.com/journals/jama/fullarticle/2775434.

Machta, Rachel, Greg Peterson, Jason Rotter, Kate Stewart, Shannon Heitkamp, Isabel Platt, Danielle Whicher, et al. "Evaluation of the Maryland Total Cost of Care Model: Implementation Report." Washington, DC: Mathematica, 2021. Available at https://innovation.cms.gov/data-and-reports/2021/md-tcoc-imp-eval-report.

Murray, Robert, and Robert Berenson. "Hospital Rate Setting Revisited." Washington, DC: Urban Institute, 2015. Available at https://www.urban.org/sites/default/files/publication/73841/2000516-Hospital-Rate-Setting-Revisited.pdf.

Perman, C., E. Adashi, E. Gruber, and H. Haft. "Improving COVID-19 Outcomes for Medicare Beneficiaries: A Public Health-Supported Advanced Primary Care Paradigm." New York, NY: The Milbank Memorial Fund, September 9, 2021. Available at https://www.milbank.org/publications/improving-covid-19-outcomes-for-medicare-beneficiaries-a-public-health-supported-advanced-primary-care-paradigm/.

Peterson, C., and D. Schumacher. "How Maryland's Total Cost Of Care Model Has Helped Hospitals Manage The COVID-19 Stress Test." Health Affairs Blog, October 7, 2020. Available at https://www.healthaffairs.org/do/10.1377/forefront.20201005.677034/full/.

Sapra, Katherine J., Katie Wunderlich, and Howard Haft. "Maryland Total Cost of Care Model: Transforming Health and Health Care." *JAMA,* vol. 321, no. 10, 2019, pp. 939–940. Available at https://jamanetwork.com/journals/jama/fullarticle/2726614.

Shimada, S. L., Alan M. Zaslavsky, Lawrence B. Zaborski, A. James O'Malley, Amy Heller, and Paul D. Cleary, "Market and beneficiary characteristics associated with enrollment in Medicare managed care plans and fee-for-service." *Medical care*, vol. 47, no. 5, May 2009, pp. 517–523. doi:10.1097/MLR.0b013e318195f86e.

Stuart, Elizabeth A. “Matching Methods for Causal Inference: A Review and a Look Forward.” *Statistical Science*, vol. 25, no. 1, February 2010, pp. 1–21.

Wang, Yixin, and José R. Zubizarreta. “Minimal Dispersion Approximately Balancing Weights: Asymptotic Properties and Practical Considerations.” *Biometrika*, vol. 107, no. 1, October 2019. doi:10.1093/biomet/asz050.

Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation. “2020 Readmission Measures Updates and Specifications Reports and Supplemental File.” 2020. Available at https://qualitynet.cms.gov/inpatient/measures/readmission/resources#tab3. Accessed August 2022.

Zubizarreta, José R. “Stable Weights that Balance Covariates for Estimation With Incomplete Outcome Data.” *Journal of the American Statistical Association*, vol. 110, no. 511, 2015, pp. 910–922. doi:10.1080/01621459.2015.1023805.

Appendix A: Detailed methods for estimating impacts and supplemental results

A.1. Design for estimating impacts

We used a difference-in-differences analysis to estimate the impacts of the Maryland Model on utilization, spending, quality of care, and population heath—for Medicare FFS beneficiaries throughout Maryland from 2014 to 2021. The difference-in-differences framework estimates impacts by comparing changes in outcomes over time for Medicare beneficiaries in Maryland with contemporaneous changes for a similar comparison group selected from outside Maryland. To select the comparison group, we used areas from across the nation, weighted to look like Maryland on many dimensions (including baseline levels and trends in key outcomes) so that the core assumption behind the difference-in-differences model is credible. That assumption is that the changes in outcomes for the comparison group accurately reflect the changes that would have occurred in Maryland absent the Maryland Model. We needed to draw the comparison group from outside Maryland because the Maryland Model is statewide with the potential to affect everyone in the state. Regression models improve the precision of the estimates and adjust for any observed differences between Maryland and the matched comparison group. The regression analyses use different units of analysis depending on the outcome (for example, the unit is the beneficiary-year for outcomes measured at the beneficiary level and discharge-year for those measured at the discharge level).

We are using 2011 to 2013 as the baseline period because doing so permits us to estimate impacts of the Maryland Model by year. By matching Maryland to a comparison group with similar outcome trends from 2011 to 2013, we aimed for the comparison group to reflect the path that Maryland would have been on if it had not introduced any of the changes starting in 2014—the counterfactual. These changes include the hospital global budgets that started with MDAPM in 2014, and the broader state accountability for cost and quality of care and corresponding broadening of incentives to providers that began with MD TCOC in 2019. Using this comparison group, we can directly estimate the accumulated effects of all changes since 2014. We can then use these yearly estimates to combine and compare estimates across time periods. For example, we can compare the average effects during the MD TCOC period (2019–2021) to the effects at the end of the MDAPM period (2017 and 2018) to comment on whether, and how much, effects have grown since the start of the MD TCOC period. For this report, we focused on this specific contrast—the MD TCOC period versus the end of the MDAPM period—because we think it represents a meaningful comparison, but the flexible yearly effect estimates approach allows the reader to make other comparisons as needed.

Because MDAPM and MD TCOC are statewide initiatives, the evaluation ultimately aims to measure population-level impacts for Maryland's entire Medicare FFS population. Thus, our primary impact analyses apply the difference-in-differences design to repeated cross sections of all observable Medicare FFS beneficiaries living in Maryland in each year.[7] The analytic file covers a pre-intervention period three years before MDAPM began (2011 to 2013), the MDAPM period (2014 to 2018), and a period after MD TCOC was implemented in 2019 and ending in 2021 for this report.

[7] We define a beneficiary as observable in the year if they are alive, enrolled in FFS Medicare with Part A and B, and have Medicare as the primary payer in at least one month of the year. We allow beneficiaries to be observable for only part of the year (as little as a single month based on meeting the criteria above). In those partial year observability cases, we annualize outcomes (projecting what outcomes would have been over a full year) and then weight by observability in the regressions, down-weighting beneficiaries who are observed for less than a full year proportional to the amount of time we observe them.

A.2. Developing the matched comparison group

We developed the matched comparison group in four steps:

1. Selected the unit of analysis for matching
2. Identified variables to match on and set criteria for what counts as sufficient balance
3. Used a reweighting method to create the matched comparison group
4. Assessed the quality of the matched comparison group in terms of balance, size, geographic spread, and statistical power

In the following sections, we describe each of these four steps. When we developed the comparison group, we explored many alternatives reflecting tradeoffs in different dimensions of quality for the comparison group. We discussed these alternatives with CMS and decided on a final comparison group that we agreed achieved the best balance on the various dimensions. In this section, we report only the results for the final selected comparison group.

A.2.1. Selecting the unit of analysis for matching

We selected PUMAs as the unit for matching. PUMAs are large enough to limit variation in outcomes attributable to random noise but small enough to capture meaningful variation within populous and diverse counties. Specifically, there are 44 PUMAs in Maryland, and the potential comparison group included 2,336 PUMAs from the remaining 49 states plus Washington, DC. PUMAs, defined by the U.S. Census Bureau, are built on census tracts and counties and contain at least 100,000 people. Larger counties such as Baltimore City (a county equivalent) are divided into multiple PUMAs, enabling finer resolutions for determining whether key contextual factors vary within the county. Sparsely populated counties are combined into a single PUMA to help ensure that any statistics calculated for this population are reliable.

A.2.2. Identifying variables to match on and setting criteria for what counts as sufficient balance

In close collaboration with CMS, we set priorities for matching variables to make the matching process feasible and on target (summarized in Table A.1). So that the matched comparison group would estimate Maryland's counterfactual, we set out to select a comparison group that had the following:

- Parallel trends for priority outcomes during the baseline period (2011–2013)
- Similar baseline levels for priority outcomes
- Similar beneficiary characteristics on aggregate, such as mean age or Hierarchical Condition Category (HCC) score
- Similar health care markets, such as Health Resources and Services Administration (HRSA) scores measuring the degree of health professional primary care shortage in the PUMA or the degree of hospital market concentration within the PUMA[8]

[8] We did not seek to match on participation in other alternative payment models (such as ACOs, CPC+, etc.) because most of these programs had not yet begun during our baseline period (2011–2013) or had low participation.

- Similar characteristics—such as percentage of people living in multi-unit homes—that can make areas more vulnerable to disease outbreaks (we included these variables to mitigate risk of bias because of COVID-19; Appendix C provides details)
- Similar proportions of beneficiaries who are Black and who live in urban versus rural areas, and similar levels and trends for select outcomes for these subgroups of beneficiaries (this similarity should help make future estimates by beneficiary subgroups more credible)

In addition, we identified what we would count as sufficient balance for each of the matching variables. The method we used to reweight comparison PUMAs allowed us to set balance standards for each individual variable. We chose tight standards (< 0.15 standardized differences between the intervention and control groups) for trends in many baseline outcomes (because tight balance underlies the parallel trends assumptions) and for some variables needed for face validity or subgroups. We chose more relaxed standards (0.25 standardized differences or larger) for other types of variables, or in cases in which tight balance was not feasible without substantially affecting the quality of the comparison group in other ways—mainly reducing the size or geographic distribution of the group.

Decision to include baseline outcomes in matching

We tried developing a comparison group without matching on baseline outcomes and trends to reduce the risk that regression-to the mean could bias impact estimates.[9] The resulting comparison groups, however, had substantially different levels and trends than the intervention group, which creates its own risk of bias if such non-parallel trends persist into the intervention period. We chose to match on outcomes (levels and trends) to improve balance on these variables and because several other aspects of the design help to mitigate the risk of regression-to-the mean. First, we used PUMAs with a large number of Medicare beneficiaries, substantially limiting the noise that underlies regression-to-the mean bias. Second, we matched on outcomes over three years, rather than a single year, further limiting noise. Finally, we assessed whether the outcome means for the comparison group in 2010 (the year before the baseline period), moved away from the baseline trend line, as you would expect it would if regression-to-the mean were biasing the estimates. We did not see any evidence that the outcome in 2010 diverged substantially from the 2011–2013 trend.

We did not match on baseline levels or trends for hospital spending or total Medicare FFS spending, even though these are important outcomes in the evaluation. Hospital spending is difficult to match on because hospital spending in Maryland during the baseline period was much higher than in the rest of the country because of the all-payer rate-setting system. If we tried to match on hospital spending, the comparison group would likely be very small and have high hospital spending for reasons quite different than those in Maryland. As a result, the trends in hospital spending for such a comparison group likely would not reflect a reasonable counterfactual for Maryland. Similarly, total Medicare spending cannot be a priority matching variable because hospital spending accounts for more than half of total spending. We did include standardized hospital spending in our matching, which is calculated for Maryland and the comparison group by re-pricing claims to a standardized national fee schedule. In this way, standardized

[9] In difference-in-differences analyses, matching on outcomes can unintentionally create biased estimates if (1) there is random variation in outcome levels in the intervention and comparison units, and (2) the selected comparison units have a long-term mean that differs from the intervention group, but they are selected because they—randomly—look like the intervention group units at the time of matching. In these cases, the mean for the comparison group can snap back to its long-term mean in the post-intervention period, leading to post-intervention outcome differences that would be misinterpreted as model impacts (Daw and Hatfield 2018).

hospital spending is more closely related to hospital utilization than spending because the pricing effects have been removed.

Table A.1. Baseline measures for selecting PUMAs into the matched comparison group

Domain and measure	Data source	SD required[a]	Time period: 2013	Time period: Trend 2011–2013
Characteristics of Medicare beneficiaries in a PUMA				
Average age	Medicare enrollment	0.25	X	
Percentage Black	Medicare enrollment	0.15	X	
Percentage Black, 5 category distribution	Medicare enrollment	0.15	X	
Percentage non-Hispanic White	Medicare enrollment	0.25	X	
Percentage Hispanic	ACS	0.25	X	
Percentage female	Medicare enrollment	0.25	X	
Percentage with rural residence	Medicare enrollment U.S. Census	0.25	X	
Percentage with rural residence, 5-category distribution	Medicare enrollment U.S. Census	0.15	X	
Percentage with original reason for Medicare entitlement: disability, ESRD	Medicare enrollment	0.25	X	
Average HCC risk score	Medicare claims	0.25	X	
Percentage of FFS beneficiaries with diabetes	Medicare claims	0.25	X	
Characteristics of a PUMA and its population[b]				
Cost of living adjusted percentage below poverty level	ACS U.S. Census	0.25	X	
Percentage living in multi-unit structure, mobile home, or group quarters	ACS	0.33	X	
Percentage older than 64	ACS	0.33	X	
Percentage younger than 18	ACS	0.33	X	
Percentage speaks English well	ACS	0.33	X	
Percentage living in crowded home	ACS	0.33	X	
Percentage without a vehicle	ACS	0.33	X	
Percentage with high school degree (or equivalent)	ACS	0.33	X	
Percentage of all adults with diabetes	BRFSS	0.25	X	
Obesity prevalence, 2011-2013[c]	BRFSS	0.25	X	

Domain and measure	Data source	SD required[a]	Time period: 2013	Time period: Trend 2011–2013
Characteristics of a PUMA's health care system and insurance market				
Health professionals shortage area index score (for primary care providers)	HRSA	0.25	X	
Herfindahl-Hirschman Index (a measure of market concentration)	Medicare claims	0.25	X	
Percentage of residents in Medicare	Medicare enrollment ACS	Medium	X	
Number of primary care providers per 1,000 beneficiaries	Medicare claims	0.25	X	
Characteristics of hospitals in a PUMA				
PUMA has one or more acute care hospitals	Hospital Compare	0.25	X	
Percentage of all discharges from a major teaching hospital	Hospital Compare and Medicare claims	0.25	X	
Percentage of all discharges from hospitals belonging to a health care system	AHRQ CHSP and Medicare claims	0.25	X	
Total number of beds	Hospital Compare IPPS	0.25	X	
Characteristics of practices and providers in the PUMA				
Percentage of PCPs in practices (TINs) that are small (1 NPI in TIN)	Medicare claims MD-PPAS	0.25	X	
Percentage of PCPs in practices (TINs) that are large (6+ NPIs in TIN)	Medicare claims MD-PPAS	0.25	X	
Characteristics related to key subgroups living within the PUMA				
Average HCC score among Black beneficiaries	Medicare enrollment Medicare claims	0.25	X	
Average HCC score among non-Hispanic White beneficiaries	Medicare enrollment Medicare claims	0.25	X	
Outcomes related to Medicare FFS spending				
Non-hospital spending, 2013	Medicare claims	0.25	X	
Non-hospital spending, 2011–2013 trend	Medicare claims	0.15		X
Post-acute care spending, 2013	Medicare claims	0.25	X	
Post-acute care spending, 2011–2013 trend	Medicare claims	0.10		X
Outcomes related to health care utilization				
Standardized hospital spending, 2013	Medicare claims	0.25	X	
Standardized hospital spending, 2011–2013 trend	Medicare claims	0.15		X
Number of all-cause acute care hospital admissions, 2013	Medicare claims	0.25	X	

Domain and measure	Data source	SD required[a]	Time period: 2013	Time period: Trend 2011–2013
Number of all-cause acute care hospital admissions, 2011–2013 trend	Medicare claims	0.15		X
Number of outpatient emergency department visits and observation stays, 2013	Medicare claims	0.25	X	
Number of outpatient emergency department visits and observation stays, 2011–2013 trend	Medicare claims	0.15		X
Outcomes related to quality of care				
30-day post-discharge unplanned readmission, 2013	Medicare claims	0.25	X	
30-day post-discharge unplanned readmission, 2011–2013 trend	Medicare claims	0.15		X
Timely follow-up after a discharge for acute exacerbations of chronic conditions, 2013	Medicare claims	0.25	X	
Timely follow-up after a discharge for acute exacerbations of chronic conditions, 2011–2013 trend	Medicare claims	0.15		X
Number of potentially preventable admissions, 2013	Medicare claims	0.25	X	
Number of potentially preventable admissions, 2011–2013 trend	Medicare claims	0.10		X
Patients' rating of their hospital care, 2013	Medicare claims	0.50	X	
Patients' rating of their hospital care, 2011–2013 trend	Medicare claims	0.15		X
Beneficiaries' rating of their primary care physician, 2013	Medicare claims	0.25	X	

Note: We conducted matching at the region (PUMA) level. When applicable, we aggregated data to the PUMA level before analyzing or matching. For example, claims- and survey-based variables measured at the beneficiary or respondent level (respectively) in the underlying data files were aggregated to the PUMA-year level for matching. Hospitals' characteristics were aggregated accounting for hospital sizes.

[a] The column "SD" refers to the maximum standardized differences we allow between Maryland and the comparison group. In our reweighting algorithm we can set tolerances for individual variables to be more (lower SD) or less (higher SD) similar between Maryland and the control group (see the section on reweighting method below for more details). We aimed for a standard of 0.25 SDs where possible, but some variables were too difficult to match on (required large tradeoffs in balance elsewhere or size of the comparison group) and thus were allowed to be more dissimilar on standardized differences (e.g., patients' rating of their hospital care in 2013).

[b] To reduce the chance that statistical noise will affect survey-based and hospital-level measures, we used three-year averages rather than data from a single year.

[c] Obesity prevalence is the 2012 BRFSS files that used smoothed average from years 2011-2013.

ACS = American Community Survey; AHRQ CHSP = Agency for Healthcare Research and Quality Comparative Health System Performance (CHSP) Initiative; BRFSS = Behavioral Risk Factor Surveillance System; ESRD = end-stage renal disease; FFS = fee for service; HCC = Hierarchical Condition Category; HRSA = Health Resources and Services Administration; IPPS = Inpatient Prospective Payment System; MD-PPAS = Medicare Provider Practice and Specialty; PCP = primary care physician; PUMA = Public Use Microdata Area; SD = standardized differences; TIN = Taxpayer Identification Number.

A.2.3. Reweighting comparison PUMAs to create the matched comparison group

To select our comparison group, we used a method called stable balancing weights developed by Zubizarreta (2015). This method belongs to a class of methods called *minimal dispersion approximately balancing weights*, or *minimal weights* for short, that reweight the comparison group units without explicitly modeling the propensity to receive the intervention (that is, propensity scores).[10] Instead of modeling propensity scores, these methods find the weights that directly optimize certain attributes of the weights, targeting covariate balance directly and simultaneously minimizing a measure of dispersion of the weights. In the case of stable balancing weights, the optimization finds the weights for comparison units that achieve preset criteria for balance on individual matching variables while minimizing the dispersion of weights across the comparison units.

Reweighting and matching methods for constructing a comparison group are closely related conceptually (Stuart 2010). The methods have similar objectives and are based on similar principles. The main difference is that matching selects a subset of potential comparison regions to form the comparison group (and thus does not use all the available data), and weighting methods use all comparison regions and give different regions more or less weight (thus using all the available data, though some PUMAs can receive zero weight).[11]

The stable balancing weights method offered two main advantages over traditional matching techniques in the MD TCOC evaluation:

1. *It allows matching on the many variables identified in Table A.1 as priorities.* Traditional matching methods based on propensity scores would likely not be able to match on so many variables because the propensity score model would risk overfitting with so many explanatory variables for only 44 intervention PUMAs.
2. *It allows for tailored balance criteria for each matching variable*. This tailoring enabled us to identify and make precise tradeoffs between balance on select variables versus the size and distribution of the comparison group.

Using an optimization-based approach, theoretically, any number of criteria can be set as constraints. As we add constraints (or tighten or require greater similarity between treated and comparison groups), however, the optimization problem becomes more difficult. The tradeoff to higher degrees of similarity across many different criteria is often the size of the comparison group represented. In other words, the algorithm will start to drop (that is, assign zero weight to) units that are too different from its target when there are no other options.

[10] Chattopadhyay et al. (2020) provided an overview of the *minimal weights* methods and contrasted them with more traditional inverse probability weighting approaches. Wang and Zubizarreta (2019) provided theoretical results.
[11] One way to think about it is that reweighting creates a matched comparison group (that is, a comparison group similar to the intervention group). Another is that matching is a form of reweighting (in which the weights for a region could be as simple as a 0 or 1).

A.2.4. Assessing the quality of the matched comparison group in terms of size, geographic spread, balance, and statistical power

In selecting the comparison group, we aimed for a group that:

1. Was large and spread across much of the country, both to improve statistical power to detect effects and to avoid the possibility that idiosyncratic health shocks in any one area would drive the results.
2. Had sufficient balance on all variables listed as priorities for matching.
3. Had sufficient statistical power to detect policy-relevant impacts.

Conditions (1) and (2) generally trade off with one another—with more precise balance coming at the expense of a smaller and less geographically disperse comparison group. We explored several alternative comparison groups with CMS and selected the one that represented the best tradeoffs across these three dimensions.

Size and geographic spread of the selected comparison group

The selected comparison group is large and covers much of the country. Table A.2 shows several statistics that give a sense of the matched comparison group on a national scale. For example, the comparison group includes 37 states with a positive weight, and about two-thirds of the weight concentrates in the top 10 states (Table A.3 shows weights of top 10 states). We also see that, in total, about 25 percent of the nation's Medicare FFS population has a positive weight in our comparison group (553 PUMAs), with 338 individual PUMAs accounting for about 90 percent of the total weight (Table A.4 shows the weight of the top 10 PUMAs in our comparison group). Finally, we also display the effective sample ratio, which is an estimate of the ratio of treatment to comparison units that accounts for the sum of the weights and the dispersion of those weights (Table A.2). Effective sample ratios of greater than 3:1 are generally considered to maximize the statistical power to detect effects for any given intervention group size. Higher ratios (for example, 10:1) only modestly increase statistical precision and can come at the cost of substantially worse balance on matching variables. Figure A.1 shows our final comparison group visually on a map of individual PUMAs. The more populous areas of the country have PUMAs with relatively small areas in the map, so populous PUMAs that received substantial weight might be hard to discern in this nationwide map.

Table A.2. Matched comparison group diagnostics

Statistic	Value
Number of states with positive weight	37
Total weight of top 10 states	66.7%
Percentage of Medicare FFS population outside Maryland with a positive weight	24.9%
Total weight of top 50 PUMAs	29.4%
Number of PUMAs accounting for 50 percent of total weight	111
Number of PUMAs accounting for 90 percent of total weight	338
Effective sample ratio treatment:comparison	7.05

FFS = fee for service; PUMA = Public Use Microdata Area.

Table A.3. Percentage of the selected comparison group in the top 10 most highly weighted states

State	PUMA FFS population (as % of total Medicare FFS population)	Comparison group FFS population (as % of Medicare FFS beneficiaries in the comparison group)
Illinois	5.0	12.4
Virginia	3.0	11.9
New Jersey	3.3	10.6
Georgia	2.9	6.3
New York	5.7	6.6
North Carolina	3.8	4.8
Pennsylvania	4.0	3.7
Florida	6.8	3.7
Connecticut	1.3	3.3
Texas	6.8	3.2

Note: As an example, 5 percent of the nation's Medicare FFS beneficiaries live in Illinois. In contrast, 12 percent of the Medicare FFS beneficiaries in the comparison group live in Illinois.

FFS = fee for service; PUMA = Public Use Microdata Area.

Table A.4. Comparison group statistics and top 10 PUMAs by final analysis weight

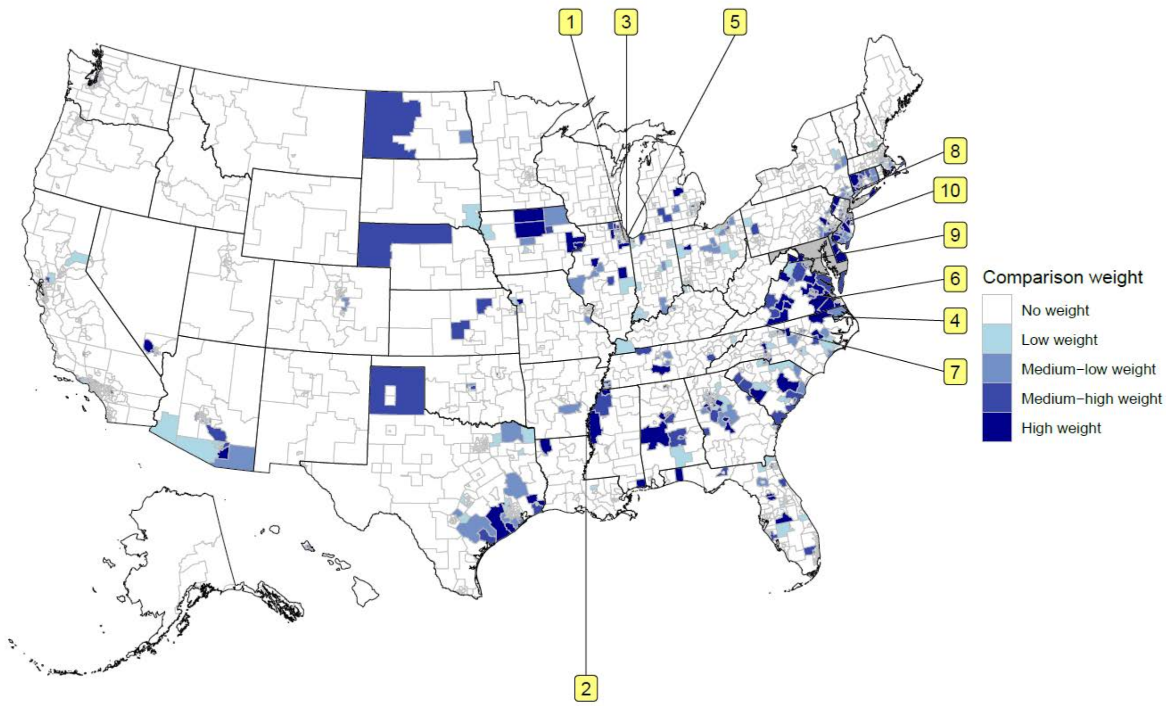
PUMA	PUMA FFS population[a]	Selected comparison group[b]
Illinois—Cook County (South) —Bloom and Rich Townships	21,426	1.3%
Mississippi—South Delta Region	19,821	0.9%
Illinois—Chicago City (South)--Auburn Gresham, Roseland, Chatham, Avalon Park and Burnside	16,842	0.9%
North Carolina—Halifax, Hertford, Northampton and Warren (East) Counties	20,711	0.9%
Illinois—Cook County (Southeast)--Thornton Township	18,735	0.9%
Virginia—Crater Planning District Commission	24,582	0.8%
Virginia—West Piedmont Planning District Commission	35,061	0.8%
Connecticut—Danbury, Ridgefield, Bethel, Brookfield, New Fairfield, Redding and Sherman Towns	18,560	0.8%
Virginia—Northern Shenandoah Valley Regional Commission (North)	16,017	0.8%
New Jersey—Ocean County (Northwest)	25,331	0.8%

[a] Number of Medicare FFS beneficiaries living in this PUMA in 2013.

[b] Final weight that the PUMA receives, which is a combination of the final matching weight and the PUMA size (FFS Medicare population).

FFS = fee for service; PUMA = Public Use Microdata Area.

Figure A.1. Map showing which PUMAs received positive weights in the selected comparison group



Note: Some PUMAs with very small areas are difficult to see on this map (for example, areas around Los Angeles, Chicago, and other major cities). As such, it might be difficult to see some PUMAs getting significant weight from this map alone. Yellow markers [1-10] indicate the top 10 PUMAs by weight in the comparison group. No weight = 0 weight; Low weight = weights in the 1st (0-25% quartile); Medium-low weight = weights in 2nd quartile; Medium-high weight – weights in the 3rd quartile; high weight – weights in the 4th quartile.

PUMA = Public Use Microdata Area.

Balance on matching variables

Overall, we achieved good balance in our selected comparison group, including on outcomes, with most measures no more than 0.25 standardized differences apart from Maryland (Table A.5).[12] The stable balancing weights method ensures all measures included in the algorithm meet the selected criteria (or the algorithm would fail). That is, we achieved balance that was *no worse* than the balance criteria specified in Table A.5 and, in many cases, significantly better. In general, the balance criteria can be used to assess the relative importance we assigned to an individual variable in our matching algorithm. Smaller standardized differences represent tighter balance. We also included in Table A.5 several variables we chose not to match on explicitly (for example, COVID-19-related variables) but that we were interested in checking balance on.

[12] Throughout our matching process, we intentionally calculated standard deviations used in constructing standardized differences at the PUMA level, rather than at the beneficiary level, as is often seen in final balance tables for beneficiary-level regressions. PUMA-level standard deviations are much smaller than beneficiary-level standard deviations, especially for measures such as HCC scores or beneficiary outcomes. This choice results in much stricter requirements on the standardized differences scale. We took this approach to be conservative, and because our comparison group is constructed at the PUMA level—a higher level of aggregation—we included several matching variables that are measured at the PUMA level. Because of these matching criteria, we achieved good balance for beneficiary-level measures in our final regressions.

Table A.5. Balance between Maryland and selected comparison group on key characteristics and outcomes

Variable description	Maryland pre-weighted mean	National pre-weighted mean	Standardized difference pre-weighting	Balance criteria (standardized differences)	Maryland post-weighted mean	Selected comparison group (National post-weighted mean)	Standardized difference post-weighting
Percentage with rural residence	16.35	25.78	-0.36	0.25	16.35	15.15	0.05
Percentage with rural residence, category [0,1]	0.41	0.27	0.32	0.15	0.41	0.41	0.00
Percentage with rural residence, category [1,25]	0.25	0.30	-0.12	0.15	0.25	0.32	-0.15
Percentage with rural residence, category [25,50]	0.26	0.20	0.15	0.15	0.26	0.20	0.15
Percentage with rural residence, category [50,75]	0.08	0.18	-0.27	0.15	0.08	0.07	0.03
Percentage with rural residence, category [75,100]	0.00	0.05	-0.23	0.15	0.00	0.00	-0.01
Average age	71.47	70.58	0.43	0.25	71.47	71.23	0.11
Percentage female	57.46	55.49	0.87	0.25	57.46	56.89	0.25
Percentage non-Hispanic White	70.58	80.21	-0.49	0.25	70.58	71.29	-0.04
Percentage Black	22.74	9.66	0.88	0.15	22.74	20.52	0.15
Percentage Black, category [0, 5]	0.20	0.57	-0.74	0.15	0.20	0.28	-0.15
Percentage Black, category [5, 15]	0.36	0.24	0.29	0.15	0.36	0.34	0.04
Percentage Black, category [15, 25]	0.18	0.09	0.33	0.15	0.18	0.14	0.15
Percentage Black, category [25, 50]	0.08	0.08	-0.02	0.15	0.08	0.09	-0.04
Percentage Black, category [50, 100]	0.18	0.03	0.89	0.15	0.18	0.16	0.15
Percentage Hispanic	7.30	13.47	-0.39	0.25	7.30	11.23	-0.25

Variable description	Maryland pre-weighted mean	National pre-weighted mean	Standardized difference pre-weighting	Balance criteria (standardized differences)	Maryland post-weighted mean	Selected comparison group (National post-weighted mean)	Standardized difference post-weighting
Percentage with original reason for Medicare entitlement: disability, ESRD	20.07	25.59	-0.61	0.25	20.07	22.33	-0.25
Average HCC risk score	1.11	1.11	0.03	0.25	1.11	1.13	-0.12
Percentage of adults with diabetes	9.33	9.19	0.08	0.25	9.33	9.19	0.08
Percentage with diabetes	30.12	28.33	0.32	0.25	30.12	29.54	0.10
Percentage below federal poverty level (adjusted for cost of living)	14.07	16.75	-0.36	0.25	14.07	15.92	-0.25
Percentage of residents in Medicare	15.03	17.85	-0.38	0.25	15.03	16.43	-0.19
PUMA has one or more acute care hospitals	0.71	0.87	-0.47	0.25	0.71	0.79	-0.25
Number of hospital beds	245.92	339.89	-0.27	0.25	245.92	333.92	-0.25
Percentage of all discharges from a major teaching hospital	14.35	15.14	-0.04	0.25	14.35	19.36	-0.25
Herfindahl-Hirschman Index	0.29	0.25	0.30	0.25	0.29	0.27	0.18
Percentage of PCPs in practices (TINs) that are small (1 NPI in TIN)	28.11	21.97	0.37	0.25	28.11	26.35	0.11
Percentage of PCPs in practices (TINs) that are large (6+ NPIs in TIN)	46.84	49.74	-0.12	0.25	46.84	46.49	0.01

Variable description	Maryland pre-weighted mean	National pre-weighted mean	Standardized difference pre-weighting	Balance criteria (standardized differences)	Maryland post-weighted mean	Selected comparison group (National post-weighted mean)	Standardized difference post-weighting
Number of primary care providers per 1,000 beneficiaries	6.14	5.05	0.22	0.25	6.14	5.06	0.21
Average health professionals shortage area index score (for primary care providers)	3.36	4.54	-0.22	0.25	3.36	3.62	-0.05
Percentage living in multi-unit structure, mobile home, or group quarters	19.04	25.60	-0.48	0.33	19.04	23.55	-0.33
Percentage older than age 64	13.86	14.97	-0.26	0.33	13.86	14.34	-0.11
Percentage younger than age 18	22.52	22.84	-0.09	0.33	22.52	23.21	-0.20
Percentage with high school degree (or equivalent)	92.06	89.65	0.44	0.33	92.06	91.16	0.17
Percentage speaks English well	97.57	96.43	0.25	0.33	97.57	96.88	0.15
Percentage living in crowded home	6.45	9.00	-0.47	0.33	6.45	8.04	-0.29
Percentage without a vehicle	7.03	6.34	0.09	0.33	7.03	7.36	-0.04
Non-hospital spending, 2013	5,299.52	5,185.46	0.09	0.25	5,299.52	5,609.06	-0.25
Non-hospital spending, 2011–2013 trend	-40.08	-71.05	0.30	0.15	-40.08	-55.60	0.15
Medicare Part A post-acute care spending, 2013	1,115.93	1,075.02	0.12	0.25	1,115.93	1,160.69	-0.14
Medicare Part A post-acute care spending, 2011–2013 trend	9.23	8.66	0.01	0.10	9.23	12.65	-0.05

Variable description	Maryland pre-weighted mean	National pre-weighted mean	Standardized difference pre-weighting	Balance criteria (standardized differences)	Maryland post-weighted mean	Selected comparison group (National post-weighted mean)	Standardized difference post-weighting
Standardized hospital spending, 2013	4,593.70	4,562.98	0.04	0.25	4,593.70	4,647.01	-0.07
Standardized hospital spending, 2011–2013 trend	98.87	69.36	0.28	0.15	98.87	83.05	0.15
Number of all-cause acute care hospital admissions, 2013	320.49	304.03	0.30	0.25	320.49	314.65	0.11
Number of all-cause acute care hospital admissions, 2011–2013 trend	-20.15	-14.71	-0.58	0.15	-20.15	-18.74	-0.15
Number of outpatient ED visits and observation stays, 2013	461.94	498.73	-0.29	0.25	461.94	454.57	0.06
Number of outpatient ED visits and observation stays, 2011–2013 trend	11.88	9.13	0.19	0.15	11.88	11.55	0.02
30-day post-discharge unplanned readmission, 2013	17.85	16.34	0.62	0.25	17.85	17.24	0.25
30-day post-discharge unplanned readmission, 2011–2013 trend	-0.76	-0.45	-0.50	0.15	-0.76	-0.66	-0.15
Timely follow-up after a discharge for acute exacerbations of chronic conditions, 2013	67.36	67.92	-0.10	0.25	67.36	67.76	-0.07
Timely follow-up after a discharge for acute exacerbations of chronic conditions, 2011–2013 trend	0.80	0.79	0.00	0.15	0.80	1.00	-0.15

Variable description	Maryland pre-weighted mean	National pre-weighted mean	Standardized difference pre-weighting	Balance criteria (standardized differences)	Maryland post-weighted mean	Selected comparison group (National post-weighted mean)	Standardized difference post-weighting
Number of potentially preventable admissions, 2013	56.96	54.88	0.13	0.25	56.96	56.04	0.06
Number of potentially preventable admissions, 2011–2013 trend	-4.22	-3.53	-0.23	0.10	-4.22	-4.23	0.01
Patients' rating of their hospital care, 2013	66.51	70.41	-0.93	0.50	66.51	68.61	-0.50
Patients' rating of their hospital care, 2011–2013 trend	0.35	1.01	-0.64	0.15	0.35	0.50	-0.15
Beneficiaries' rating of their primary care physician, 2013	89.88	89.99	-0.04	0.25	89.88	90.23	-0.13
Obesity prevalence, 2012	28.24	27.95	0.06	0.25	28.24	27.03	0.25
Number of COVID-19 outpatient ED visits and observation stays, 2020	6.40	9.07	-0.51	Not included in matching	6.40	7.56	-0.22
Number of COVID-19 outpatient ED visits and observations stays, 2021	10.30	12.30	NA	Not included in matching	10.42	10.56	NA
Excess number of all-cause acute care hospital admissions (2020 minus 2019)	-39.69	-39.95	0.02	Not included in matching	-39.69	-40.49	0.06
Excess number of outpatient ED visits and observation stays (2020 minus 2019)	-121.10	-114.80	-0.19	Not included in matching	-121.10	-117.19	-0.12

Variable description	Maryland pre-weighted mean	National pre-weighted mean	Standardized difference pre-weighting	Balance criteria (standardized differences)	Maryland post-weighted mean	Selected comparison group (National post-weighted mean)	Standardized difference post-weighting
Excess number of ED visits and observation stays ending in an inpatient stay (2020 minus 2019)	-27.08	-24.16	-0.23	Not included in matching	-27.08	-25.36	-0.14
Number of surgical hospitalizations (2020 minus 2019)	-14.75	-14.10	-0.14	Not included in matching	-14.75	-14.22	-0.12
Number of elective hospitalizations (2020 minus 2019)	-11.46	-11.73	0.05	Not included in matching	-11.46	-11.31	-0.03
Average HCC score, Black beneficiaries	1.34	1.42	-0.52	0.33	1.34	1.39	-0.33
Average HCC score, non-Hispanic White beneficiaries	1.05	1.06	-0.09	0.33	1.05	1.06	-0.04
Number of potentially preventable admissions, Black beneficiaries, 2013	80.05	77.54	0.13	Not included in matching	80.05	76.92	0.16
Number of potentially preventable admissions, 2011–2013 trend, Black beneficiaries	-4.63	-4.63	0.00	Not included in matching	-4.63	-5.51	0.11
Number of potentially preventable admissions, non-Hispanic White beneficiaries, 2013	51.90	52.99	-0.07	Not included in matching	51.90	51.64	0.02

Variable description	Maryland pre-weighted mean	National pre-weighted mean	Standardized difference pre-weighting	Balance criteria (standardized differences)	Maryland post-weighted mean	Selected comparison group (National post-weighted mean)	Standardized difference post-weighting
Number of potentially preventable admissions, 2011–2013 trend, non-Hispanic White beneficiaries	-4.27	-3.41	-0.28	Not included in matching	-4.27	-3.85	-0.14
30-day post-discharge unplanned readmission, Black beneficiaries, 2013	21.87	21.48	0.12	Not included in matching	21.87	21.98	-0.03
30-day post-discharge unplanned readmission, 2011–2013 trend, Black beneficiaries	-0.89	-0.49	-0.22	Not included in matching	-0.89	-0.66	-0.13
30-day post-discharge unplanned readmission, non-Hispanic White beneficiaries, 2013	16.71	15.77	0.46	Not included in matching	16.71	15.97	0.36
30-day post-discharge unplanned readmission, 2011–2013 trend, non-Hispanic White beneficiaries	-0.80	-0.45	-0.53	Not included in matching	-0.80	-0.66	-0.21

Notes: The pre-weighted means are the raw PUMA-level means (weighted only for FFS beneficiary count). Post-weighted means are weighted by the final matching weights. Standardized differences are a measured using the PUMA-level standard deviations.

ED = emergency department; ESRD = end-stage renal disease; FFS = fee for service; HCC = Hierarchical Condition Category; NPI = National Provider Identifier; PCP = primary care physician; PUMA = Public Use Microdata Area; TIN = Taxpayer Identification Number.

Statistical power

Based on the size of the selected comparison group, and assumptions about the variation in outcomes (and degree of clustering within PUMAs), we estimated that the evaluation would have sufficient statistical power. For example, we estimated that the model would be able to reliably detect an impact on hospital admissions of 2.5 percent or larger.[13] The strong statistical power stems, in part, from the large size of the comparison group, as indicated by the effective sample size ratio of comparison to intervention group beneficiaries of 7:1.

The impact estimates shown in Section 4 confirm that the estimates have good statistical power, with the model finding impacts on total spending as small as 1.0 percent being statistically different from zero.

A.3. Unadjusted mean outcomes over time, for Maryland and the comparison group

To help interpret what drives the difference-in-differences impact estimates, we include the size of the intervention and comparison groups over time (Table A.6) and the trends in unadjusted (but comparison group weighted) means for study outcomes since 2011 (Figures A.2, A.3, and A.4) for these populations. The figures are especially helpful for identifying the time trends in the intervention and comparison groups that underlie that differences-in-differences impact estimates—for example, that all-cause admissions have been falling steadily in the intervention and comparison groups but more so in the intervention group (Figure A.3, Panel B). For most outcomes, the trends extend through 2021. For standardized spending, total spending including non-claims payments, and patients' ratings of their hospitals, the trends run through 2020 instead because of lags in data availability. For patients' rating of their personal doctor, trends extend only through 2019 because the survey was not collected in 2020 because of the COVID-19 pandemic and data were not yet available for 2021. For the means figures, beneficiaries in Maryland are weighted by their observability in the year, and beneficiaries in the comparison group are weighted by their observability and their matching weights (see A.2.3). Because episodes-level outcomes (30-day unplanned readmission and timely follow-up after acute exacerbation of a chronic condition) are not annualized, episodes in Maryland receive a weight of 1, and episodes in the comparison group are weighted by their matching weights.

For the beneficiary and episode analyses, the ratio of comparison group to Maryland beneficiaries decreases slowly by year. This pattern occurs largely because more Medicare FFS beneficiaries enter Medicare Advantage (and exit the study population) over time in the comparison group than in Maryland, where rates of Medicare Advantage enrollment are low. Section A.6 discusses the how differential enrollment into Medicare Advantage might bias our impact estimates and the methods we use to limit that bias risk. In contrast, the ratio for the patients' ratings of their personal doctor stays relatively constant over time because that analysis includes both FFS and Medicare Advantage beneficiaries.

The unit of analysis for patients' rating of their hospital is the hospital—not the beneficiary—which is why the study population is so much smaller for that outcome. To be included in the analysis, the hospitals also needed to meet several inclusion criteria (see B.2.3). For example, they needed to have Hospital CAHPS survey results in at least one year. CMS only reports Hospital CAHPS data for hospitals with inpatient beds (so free-standing emergency rooms could not be included) and meet a minimum threshold for number of respondents. These filters are the reason the number of hospitals in the study

[13] By reliably detect, we mean that the regressions would have 80 percent power to detect a difference of at least the size indicated (using a two-tailed test and a $p < 0.10$ cutoff for statistical significance).

population for Maryland for this analysis range from 41 to 44, smaller than the 52 hospitals currently (in 2022) in the Maryland state agreement.[14]

Table A.6. Size of the Maryland and comparison groups over time (weighted)

Year	Weighted Maryland count	Weighted comparison count	Comparison: Maryland weighted ratio
Beneficiary-level analysis counts (number of Medicare FFS beneficiaries)			
2011	626,217	7,504,397	12.0
2012	645,971	7,614,399	11.8
2013	669,371	7,715,748	11.5
2014	691,471	7,743,808	11.2
2015	714,401	7,777,619	10.9
2016	723,420	7,900,084	10.9
2017	727,020	7,817,757	10.8
2018	735,154	7,778,144	10.6
2019	751,160	7,650,372	10.2
2020	759,065	7,478,593	9.9
2021	736,739	7,163,272	9.7
Episode analysis: 30-day post-discharge unplanned readmission index admission counts (number of index admissions)			
2011	209,338	2,473,462	11.8
2012	202,245	2,378,216	11.8
2013	198,208	2,262,289	11.4
2014	194,090	2,184,900	11.3
2015	193,968	2,181,226	11.2
2016	190,317	2,162,350	11.4
2017	186,110	2,138,622	11.5
2018	181,055	2,095,941	11.6
2019	177,093	2,055,665	11.6
2020	147,469	1,687,983	11.4
2021	147,818	1,609,875	10.9
Episode analysis: Timely follow-up after acute exacerbation of chronic conditions denominator counts (number of eligible index admissions and ED visits)			
2011	72,418	847,914	11.7
2012	74,493	848,486	11.4
2013	74,250	821,043	11.1
2014	75,463	812,035	10.8
2015	77,771	814,155	10.5
2016	72,256	776,753	10.8
2017	74,687	784,623	10.5
2018	73,096	755,827	10.3
2019	74,080	732,303	9.9
2020	55,677	527,390	9.5
2021	53,486	497,300	9.3

[14] The one exception was a data anomaly in 2016 when several Maryland hospitals did not report Hospital CAHPS data.

Year	Weighted Maryland count	Weighted comparison count	Comparison: Maryland weighted ratio
Beneficiary perspective analysis: Patients' rating of their personal doctor counts (weighted number of surveys)[a]			
2011	492,529	6,917,302	14.0
2012	555,809	7,376,531	13.3
2013	590,427	7,644,740	12.9
2014	657,352	8,830,296	13.4
2015	647,763	8,346,302	12.9
2016	650,101	8,488,112	13.1
2017	608,771	7,765,859	12.8
2018	675,987	9,105,769	13.5
2019	620,783	8,437,406	13.6
Beneficiary perspective analysis: Patients' rating of their hospital counts (weighted number of hospitals)[b]			
2011	41	703	17.1
2012	41	705	17.2
2013	42	703	16.7
2014	43	706	16.3
2015	44	708	16.1
2016[c]	n.a.	n.a.	n.a.
2017	44	710	16.2
2018	44	705	16.1
2019	43	696	16.0
2020	43	688	16.0

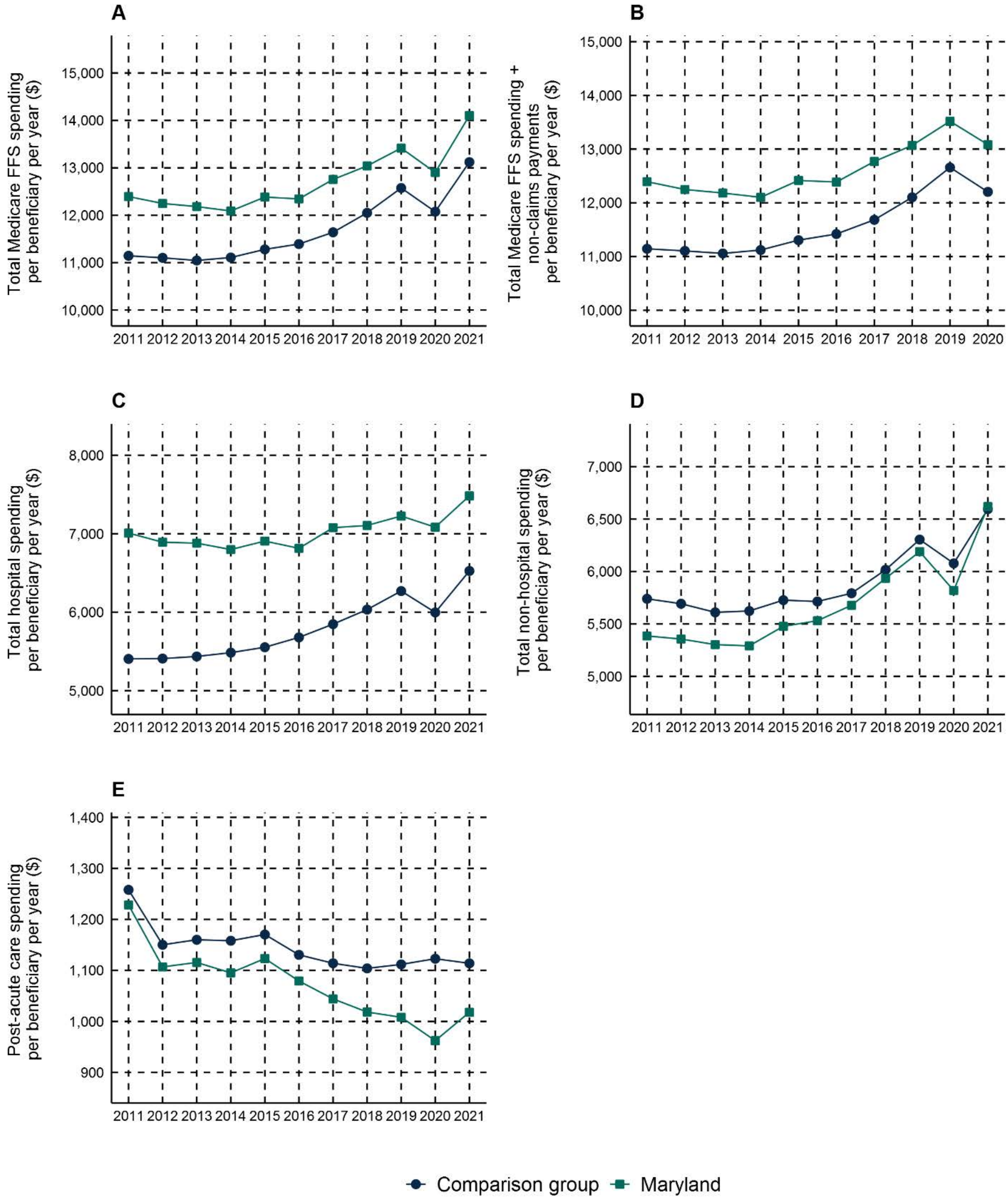
[a] The counts for patients' rating of their personal doctor are weighted by our PUMA matching weight (normalized to mean 1) multiplied by the CAHPS survey weights. CAHPS weights are designed to inflate back to approximately the total number of FFS and Medicare Advantage beneficiaries in the state, not the actual number of surveys completed. Numbers are lower than the total number of FFS beneficiaries in our beneficiary-level sample because not all people who take the survey respond to this question – only those with a primary doctor and who have received care in the last six months. Actual survey response rates are declining during this period from about 50 percent in 2011 to less than 33 percent among FFS beneficiaries in 2019.

[b] The counts for patients' rating of their hospital are weighted by our PUMA matching weight (normalized to mean 1) multiplied by the size of the hospital based on the number of discharges in 2013 or the year after the first year the hospital appears in our data (normalized to mean 1).

[c] Calendar year 2016 was excluded from analyses because several large hospitals in Maryland did not report scores in that year, potentially skewing results.

CAHPS = Consumer Assessment of Healthcare Providers and Systems; ED = emergency department; FFS = fee for service; n.a. = not applicable; PUMA = Public Use Microdata Areas.

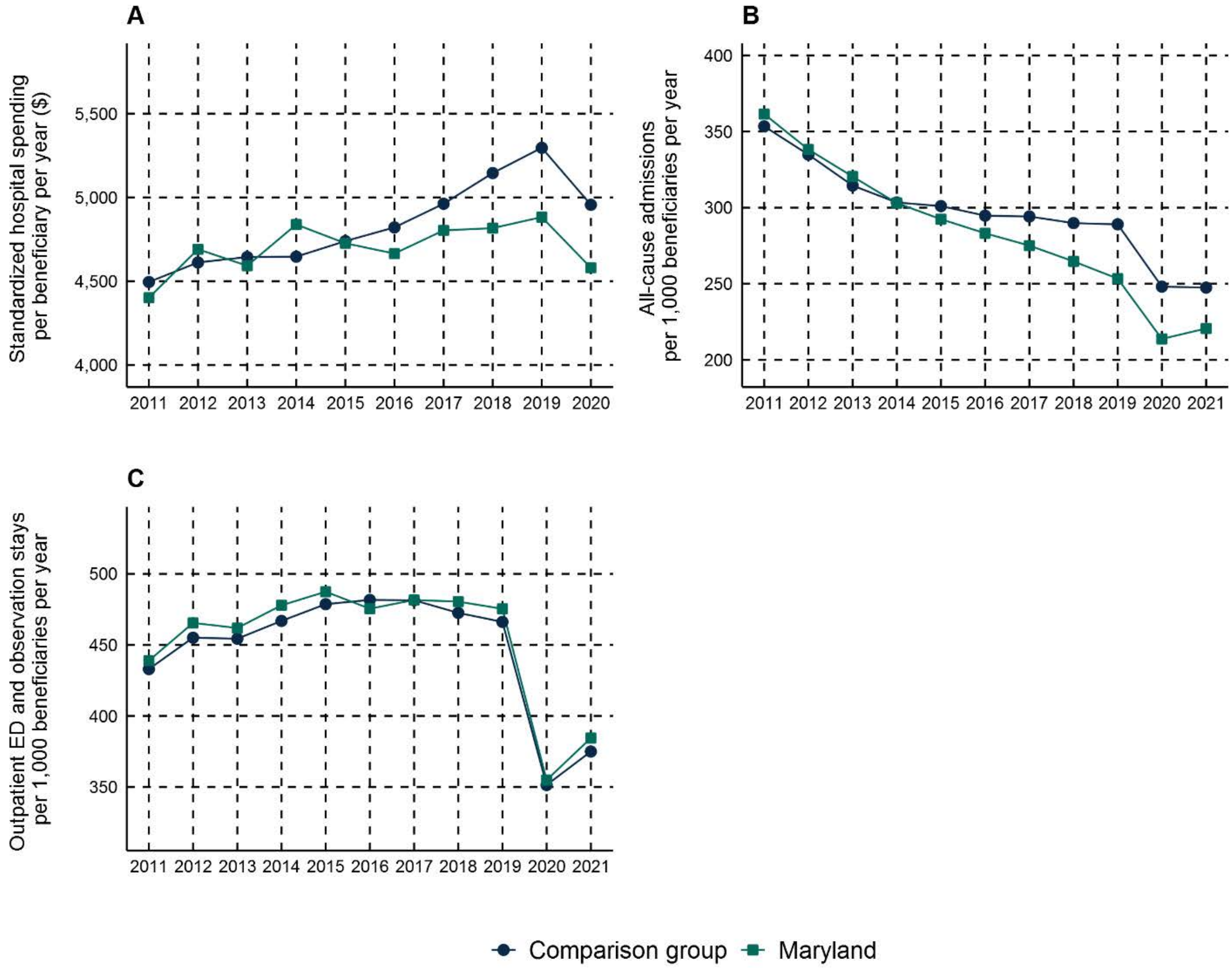
Figure A.2. Unadjusted spending per beneficiary per year after matching



Note: Maryland mean is weighted for observability in Medicare FFS claims. Comparison group mean is weighted for matching and observability.

FFS = fee for service

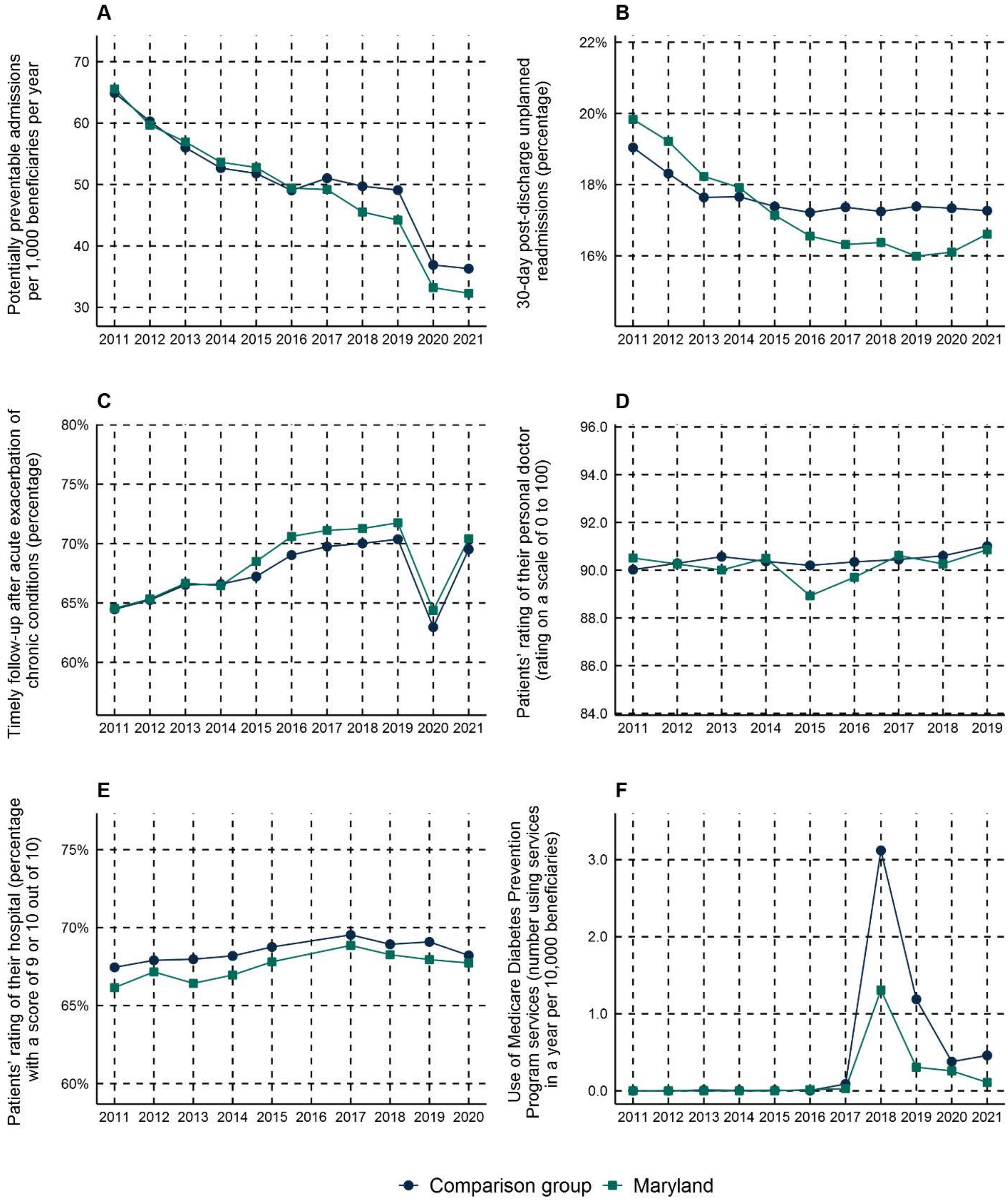
Figure A.3. Unadjusted utilization after matching



Note: Maryland mean is weighted for observability (except for 30-day unplanned readmissions and follow-up after acute exacerbation which are episode level). Comparison group mean is weighted for matching and observability.

ED = emergency department.

Figure A.4. Unadjusted quality outcomes and population health after matching



Note: Maryland mean is weighted for observability (except for 30-day unplanned readmissions and follow-up after acute exacerbation which are episode level). Comparison group mean is weighted for matching and observability. For hospital rating, 2016 was excluded from analyses because several large hospitals in Maryland did not report scores in that year, potentially skewing results.

A.4. Regression model specifications

A.4.1. Regression specifications and statistical testing for beneficiary-year and episode-year Medicare FFS claims-based analyses

We used linear regression models to implement the difference-in-differences impact analyses. We measured impacts separately for each year and separately for the MDAPM and MD TCOC periods. The findings in this report included three units of analysis: (1) analyses of observations for each Medicare FFS beneficiary in Maryland and the matched comparison regions for each year (beneficiary-year analyses, including patients' rating of their personal doctor), (2) analyses of episode outcomes with observations for each episode for each year (episode-year analyses), and (3) analyses of hospital ratings with observations for each hospital for each year they appear in the data (detailed in A.4.2). The beneficiary-year and episode-year models accounted for the clustering of beneficiaries within PUMAs through cluster-robust standard errors, controlled for time-invariant effects of unobserved confounders and common shocks through the use of fixed effects, and they included baseline and time-varying covariates as independent variables.

Impact estimates

The difference-in-differences regression models for the beneficiary-year analyses with claims-based outcome measures used Medicare FFS data with one observation per beneficiary for each year (2011 to 2021). The regression models for the episode-year analysis took the same form, but with the unit of analysis as the episode rather than the beneficiary. The regression model to estimate the yearly impact for beneficiary- and episode-level estimates took the following form:

$$(1) \qquad y_{it} = \sum\nolimits_{\tau=2014}^{2021} T_\tau M_r \delta_{t,\tau} + X_{it}\beta + \gamma_t + \mu_r + \varepsilon_{it}$$

In this model, y_{it} represents the outcome for beneficiary *i* (or episode *i*) in year *t* in region (PUMA) r, τ indexes years (with $\tau = 2011$ corresponding to the first year),[15] M_r equals 1 for Maryland beneficiaries (or episodes) and 0 for beneficiaries (or episodes) from the comparison regions, and T_τ is a dummy variable that equals 1 for observations in year τ and equals zero otherwise. X_{it} is a set of independent covariates whose relationship with the outcome we allow to change with time using an interaction term. The covariates are available in Table A.7. γ_t represents a set of year fixed effects and μ_r represents a set of PUMA-level fixed effects for beneficiary-year outcomes and hospital fixed effects for episode-year outcomes.

Beneficiaries in Maryland generally receive a weight of 1 in the regression models. But in cases in which a beneficiary is unobservable (that is, not alive and enrolled in Medicare Parts A and B with Medicare as their primary payer) the whole year, we annualized their beneficiary-year outcomes and constructed observability weights that reflect the amount of time that the beneficiary is observable in the year. For the comparison group beneficiaries, we applied the matching weights (detailed in Section A.2) to account for the PUMA-level reweighting along with the observability weights; the two weights were multiplied together to produce a final, beneficiary-level weight. For episode analyses, we applied the matching

[15] All time trends are relative to the last year of the baseline period (2013), which is the reference year in the regression models.

weights to comparison group beneficiaries, and Maryland beneficiaries received a weight of 1 because episode analyses were not annualized.

The impact estimates are the δ's—the change in mean outcomes in the intervention group each year after accounting for the changes in the comparison group in the respective year (the γ_t's). Separate estimates for each year (that is, one δ per year) allowed for nonlinearity in the effects (for example, effects might not occur immediately or could level off or decline over time).

In addition to the yearly impact estimates, we also estimated the combined effect during the MD TCOC period. The regression model to estimate the combined 2019–2021 impact estimates took the following form:

$$(2) \qquad y_{it} = \sum_{\tau=2014}^{2018} T_\tau M_r \delta_{t,\tau} + T_{2019-2021} M_r \delta_\gamma + X_{it}\beta + \gamma_t + \mu_r + \varepsilon_{it}$$

In this model, y_{it} represents the outcome for beneficiary *i* in year *t* in region (PUMA) r, τ indexes years (with $\tau = 2011$ corresponding to the first year), M_r equals 1 for Maryland beneficiaries and 0 for beneficiaries from the comparison regions, T_τ is a dummy variable that equals 1 for observations in year τ and equals zero otherwise, and $T_{2019-2021}$ is a dummy variable that equals 1 for observations in years 2019 to 2021. X_{it} is a set of independent covariates whose relationship with the outcome we allow to change with time using an interaction term. The covariates are listed below in Table A.7. γ_t represents a set of year fixed effects and μ_r represents a set of PUMA-level fixed effects for beneficiary-year outcomes and hospital fixed effects for episode-year outcomes. δ_γ represents the impact estimates during the MD TCOC period.

Finally, we also estimated models that produce an estimate of the *difference* between impacts during the MD TCOC period and the end of the MDAPM period. Models that estimate this difference took one of the two following forms:

$$(3) \qquad y_{it} = \sum_{\tau=2014}^{2016} T_\tau M_r \delta_{t,\tau} + T_{2017-2021} M_r \omega_\gamma + \sum_{\tau=2019}^{2021} T_\tau M_r \delta_{t,\tau} + X_{it}\beta + \gamma_t + \mu_r + \varepsilon_{it}$$

$$(4) \qquad y_{it} = \sum_{\tau=2014}^{2016} T_\tau M_r \delta_{t,\tau} + T_{2017-2021} M_r \omega_\gamma + T_{2019-2021} M_r \delta_\gamma + X_{it}\beta + \gamma_t + \mu_r + \varepsilon_{it}$$

Model (3) estimates the difference, between impacts during the MD TCOC period and the end of the MDAPM period *for each year of the MD TCOC period*. Model (4) estimates this same difference, as an average across the full MD TCOC period to date. The key new term in models (3) and (4) is $T_{2017-2021} M_r$. This term is 1 for Maryland observations in any year from 2017 to 2021 and 0 otherwise. Adding this term and including terms during the MD TCOC period alone, allows us to interpret the δ_t impact estimates during the MD TCOC period as *net* of effects during the last two years of MDAPM (2017-2018). We estimate models this way (instead of simply combining and subtracting estimates from models (1) or (2) above to generate the difference) to accurately generate confidence intervals and *p*-values for the difference in effects between the MD TCOC period (or individual years in the MD TCOC period) and the effects at the end of MDAPM (2017-2018). All other terms in models (3) and (4) that are shared with models (1) and (2) are interpreted the same.

Covariates

The covariates in Equations 1 to 4 are included to account for trends in the intervention and comparison groups, improve the precision of the impact estimates, and net out effects of any observed residual differences in characteristics between the intervention and comparison groups. A full list of the covariates included in the claims-based beneficiary-year analyses and for the episode analyses for readmission and timely follow-up is available in Table A.7. We control for beneficiaries' demographic characteristics (age, race, ethnicity, and sex) and Medicare enrollment characteristics (original reason for entitlement, and whether a new Medicare beneficiary in each year) and a measure of the beneficiary's PUMA Social Vulnerability Index ranking in the regression models. By incorporating beneficiary characteristics X_{it} from claims and other data sources (including the characteristics of the region in which the beneficiary lives), we control for shifts in beneficiary characteristics over time unrelated to the model that, if unaccounted for, might lead to spurious conclusions. The vector of coefficients, β, control for these types of effects. Each of the characteristics in X_{it} are interacted with year to allow their relationship with the outcome to vary over time.

Some of the beneficiary characteristics we included in the list of covariates were indicators of a beneficiary's health status each year. We identified health status based on the presence (or absence) of 36 condition categories and 1 indicator for greater or equal to three Chronic Condition Data Warehouse (CCW) conditions. We developed this list of 36 conditions as those that (1) were included in CMS Chronic Condition Data Warehouse 27 chronic conditions active from 2011–2020 or 40 Other Chronic Health, Mental Health, and Potentially Disabling Conditions; (2) had a prevalence large enough to reliably estimate its association with outcomes in individual years (3) were not conceptually endogenous, (that is they were not conditions that the Maryland Model explicitly aims to reduce).[16,17]

For the unplanned readmissions, we controlled for the index admission category and the beneficiary's health and chronic conditions covariates used in the beneficiary-level regressions.[18] For the timely follow-up outcome after acute exacerbations of chronic conditions, we controlled for the specific chronic conditions used to define the measure and the beneficiary's health and chronic conditions covariates used in the beneficiary-level regressions.

[16] We excluded sickle cell disease, pressure ulcers and chronic ulcers, spinal cord injury, spina bifida and other congenital anomalies of the nervous system, muscular dystrophy, traumatic brain injury and nonpsychotic mental disorders due to brain damage, cerebral palsy, and learning disabilities due to very small prevalence in FFS Medicare claims (<0.05 percent)

[17] Endogenous conditions are those whose prevalence might be changed by the Maryland Model. If we adjusted for changes in these conditions over time, we might adjust away impacts of the Maryland Model. We flagged CCW conditions related to diabetes and behavioral health as endogenous, particularly because of the focus on reducing body mass index and drug overdose deaths under MD TCOC (Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation, 2020).

[18] The index admission categories include surgical or cardio respiratory or cardiovascular or neurology or medicine based on the procedure codes and principal diagnosis, per CMS/Yale technical specifications.

The regional fixed effects, μ_r, in beneficiary-year models net out the effects of any time-invariant differences between the regions in Maryland and the comparison regions.[19] Controlling for PUMA fixed effects implicitly accounts for all PUMA-level baseline measures we used in constructing the matched comparison group, including characteristics of the Medicare beneficiaries in the region, the region and its population, the region's health care system and insurance market, hospitals in the region, practices and providers in the region, and primary care providers in the region. Therefore, we do not include any additional PUMA-level variables as control variables in the regressions. Hospital fixed effects for the episode analyses account for all time-invariant differences between Maryland and the comparison hospitals (including the types of services they provide) and changes in hospitals' market shares over time. Collectively, these terms improve the precision of the impact estimates (the δ's) by reducing the amount of unexplained variation in the outcome (ε_{it}).

Table A.7. Covariates for the impact analyses, by type of regression model or outcome

Domain and measure	Claims-based beneficiary-year-level outcomes	Readmission rates (episode analyses)	Timely follow-up (episode analyses)	Patients' rating of their doctor	Patients' rating of their hospital
Age	X	X	X	X	
Gender	X	X	X	X	
Race and ethnicity	X	X	X	X	
Social Vulnerability Index	X	X	X	X	
Original reason for Medicare entitlement[a]	X	X	X	X	
New Medicare beneficiary	X	X	X	X	
Rural residence	X	X	X	X	
Region (PUMA) in which the beneficiary resides	X			X	
Health conditions					
Health condition flags (based on CCWs)[b]	X		X		
Has three or more CCW conditions	X		X		
Self-reported health				X	
Hospital		X	X		
5 clinical cohorts for unplanned readmission measure case mix risk adjustment[c]		X			
6 chronic conditions used to define the timely follow-up after acute exacerbations of chronic conditions measure[d]			X		
Case mix index					X

[a] ESRD is measured in Medicare enrollment date in addition to claims. The ESRD category includes all beneficiaries with ESRD, and the Disability insurance benefits category does not.

[b] See text in section 2.2.2. for a full list of CCW conditions included

[c] Cohorts are surgical, cardiorespiratory, cardiovascular, neurology, or medicine. See the specifications for 30-day unplanned readmission developed by the Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (2020) for details on index admission cohort assignment.

[19] The size of the data from our analytic files (more than 90 million observations when stacked across years) means we must use SAS to implement regressions on the Virtual Research Data Center. SAS has limited options for absorbing many dummy variables in the regression (such as the > 2,300 PUMA fixed effects). For computational feasibility, we run regressions by "de-meaning" the outcomes and all covariates at the PUMA level. That is, for each variable in an observation (including the outcome and all covariates), we replace the variable's value with the observed value minus the PUMA-specific mean (across all years) for that variable. This method is mathematically equivalent (in linear models) to adding PUMA fixed effects but considerably faster because it does not need to estimate the PUMA fixed effects explicitly in the regressions.

[d] The categories are asthma or hypertension or coronary artery disease or heart failure or chronic obstructive pulmonary disease or diabetes. See the IMPAQ Health (2018) specifications for more details on the chronic condition category assignment.

ESRD = end-stage renal disease; FFS = fee for service; CCW = Chronic Condition Data Warehouse; PUMA = Public Use Microdata Area.

A.4.2. Regression specifications and statistical testing for beneficiary-year and hospital-year patients' ratings analyses

In addition to the claims-based outcomes described above, we estimated impacts for two measures of patient experience from national surveys; one on patients' ratings of their personal doctor and one on patients' rating of their hospital care.

Patients' rating of their personal doctor

Patients' rating of their personal doctor comes from the FFS and Medicare Advantage CAHPS surveys administered by CMS. These data contain a beneficiary ID that links directly to Medicare claims (see Appendix B for details). As such, the regression models estimating difference-in-differences impacts were similar to the beneficiary-year regression models describes in Section A.4.1.

Although similar, the regression models for this outcome differ from the models described in A.4.1. in the following ways:

- **Time period**. The COVID-19 pandemic caused a suspension of the surveys in 2020, and 2021 measures were not made available in time for this report. This means we can estimate only a single year of MD TCOC period effects (2019). Because we do not need to combine across years of the MD TCOC period for this outcome, we estimated only models described in equations (1) and (4) above (through to 2019).
- **Weighting.** In our main regression models, Maryland beneficiaries are weighted by their observability weights, and comparison group beneficiaries are weighted by their observability weight times the matching weight (see Section A.2). For patients' rating of their personal doctor, we weighted Maryland beneficiaries by the CAHPS survey weight, which is designed to correct for survey response bias and returns the weighted population counts to approximate the total FFS and MA population in the state (see Section A.3 for counts).[20] We weighted the comparison beneficiaries by the product of the CAHPS survey weight and the matching weight. Observability weight are not applicable in the survey analysis because respondents cannot be partially observed.
- **Covariates.** In the analysis for patients' rating of their personal doctors, we included beneficiaries' demographic characteristics, Medicare enrollment characteristics, and a measure of the beneficiary's PUMA Social Vulnerability Index ranking in the same way we did for other beneficiary- and episode-level outcomes. We also included a measure of the beneficiary's self-reported education because this information was available as part of the CAHPS surveys. We did *not* include, however, time-varying health condition controls as measured by CCW chronic conditions. The reason is that the sample of beneficiaries in this analysis includes FFS and Medicare Advantage beneficiaries. The primary reason for including time-varying health condition controls in our FFS sample was to correct for differential changes in the population because of beneficiaries leaving for Medicare Advantage (see Section A.6). Because we do not have beneficiaries leaving our sample, we do not need to correct for health status differences that are attributable to the changing sample, so we exclude health condition controls from

[20] The study population for the doctor rating outcome in Maryland is smaller than the total Medicare population (FFS and Medicare Advantage) in the state. This difference occurs largely because the survey only asks beneficiaries to rate their personal doctor if they say that they have a personal doctor who they have seen in the past six months.

our regressions. Another reason we do not include these controls is that they are missing for all Medicare Advantage beneficiaries because the CCW conditions require at least of year of FFS claims lookback to identify the conditions using diagnosis codes. We do, however, include a measure of self-reported health in our regressions, available for all beneficiaries who completed the survey. We include self-reported health to help further correct for survey response bias because the correction made by the CAHPS weights is done at the state level, and our matched comparison group is defined at the PUMA level. Importantly, our results do not materially change with and without the inclusion of self-reported health as a covariate.

Patients' rating of their hospital

Patients' rating of their hospital care comes from the Hospital CAHPS survey, which is administered by individual hospitals (or third-party contractors) to randomly selected patients recently discharged from the hospital (regardless of payer). The data from Hospital CAHPS is publicly available from CMS's website and stored at the hospital-year level (that is, an average set of responses for that hospital in the reporting period). The hospital rating measure that we used is defined as the percentage of survey respondents who rated their hospital overall 9 or 10 out of 10 (see Section B.1 for more details on the measure). Though the core of the difference-in-differences model we use is similar to the equations in Section A.4.1, the regression specifications for patients' rating of their hospital care has several important differences.

First, because data are available only at the hospital-year level, we estimate regressions using hospital-year observations. Because hospitals are different sizes and might contribute differently to our estimate of impacts, this has implications for how we weight observations in our sample. We continue to use our PUMA-level matching weight to ensure we use the same comparison group as for other outcomes, but we then multiply the matching weight by the normalized number of discharges observed in 2013 FFS Medicare claims (to avoid impacts on hospital admissions and discharges from affecting the regression weights). [21,22] Giving larger hospitals more weight reflects that fact that larger hospitals will influence experience with hospital care for more beneficiaries in the state. Weighting all hospitals equally may not accurately represent the average beneficiary hospital rating in the area if ratings differ for larger versus smaller hospitals.[23]

Next, we observed a potentially problematic data anomaly in the year 2016. Several Maryland hospitals, including two of its largest (hospitals associated with the University of Maryland system and the Johns Hopkins Hospital system), did not report Hospital CAHPS scores in 2016. We do observe scores for these hospitals in all other years from 2011 to 2020. Because the missing hospitals represent a significant amount of weight in the analysis, to avoid anomalous results in 2016, we removed that year of data from our analysis for all hospitals. In addition, similar to FFS Medicare Advantage CAHPS scores, data from 2021 were not available for this report, though we do have Hospital CAHPS scores based on reporting from the second half of 2020. As such, our regression models include estimated impacts through 2020. Notably, the impacts in 2020 were consistent with impacts in earlier years, suggesting that neither the partial year of data nor the COVID-19 pandemic appeared to have a large influence on results.

[21] We placed hospitals into PUMAs by geo-coding addresses using GIS software to generate X,Y coordinates for every hospital in the Hospital Compare database.
[22] If a hospital did not have claims in 2013, we used the year after the hospital appeared in our data. For example. If a hospital newly opened in 2016 we would use as their discharge weight the number of discharges in 2017.
[23] One limitation of this weighting approach is that our hospital size weights represent the average number of FFS beneficiaries (since they are based on number of discharges in the claims), but hospital ratings are for all patients.

Finally, an important key difference for estimating impacts on patients' rating of their hospital care relative to claims-based beneficiary-level and episode-level analyses is how we defined the list of covariates included in the regressions. To control for differences in case mix over time that could be the result of shifting care out of the hospital in Maryland, we included an index measuring hospital case mix based on hospital diagnosis-related groups (DRGs) in all models (see Section B.2.3.). We also considered other hospital-level controls, such as the hospital's wage index, measure of disproportionate share, resident-to-bed ratio, the percentage of the total population residing in a rural zip code for the PUMA the hospital is located in, and the average SVI of the PUMA, in addition to hospital fixed effects. Our final models chose not to control for any of these additional measures (other than case mix index) or for hospital fixed effects because doing so might control for differences in hospitals that could be the result of hospitals closing and leaving our sample. To the extent that the Maryland Model supports hospitals financially in a way that would avoid closures (leading to a more stable set of hospitals over time relative to the comparison group), we aimed to capture those effects as part of our impact estimates. Ultimately, this decision did not affect our conclusions because including hospital fixed effects and the hospital controls listed above did not materially change our results.

A.4.3. Regression-adjusted means and percentage impact

To help interpret the estimated difference-in-differences impact estimates, and to help understand the magnitude of effects across outcomes that are on different scales, we calculated regression-adjusted means and percentage impact for each of our estimated outcomes.

Regression-adjusted means

Regression-adjusted means help the reader decompose the difference-in-differences impact estimate into its component parts: the mean in Maryland and the mean in the comparison group, before and after the intervention. In all periods, including baseline (2011–2013), MDAPM (2014–2018), and MD TCOC (2019–2021), and their individual years, the regression-adjusted mean for Maryland is simply the mean of the outcome in Maryland during that period or year (weighted for observability in claims-based beneficiary-year analyses).

For the comparison group, in the baseline period, we calculated the regression-adjusted mean as the mean of the outcome in the comparison group weighted by the PUMA matching weights (times observability in claims-based beneficiary-year analyses). In all post-baseline years (2014–2021), we calculated the regression-adjusted mean in the comparison group as the Maryland mean in that period or year minus the difference-in-differences impact estimate associated with that period or year, minus the difference between Maryland and the comparison group in the baseline period. For example, Maryland averaged 340 all-cause admissions per 1,000 beneficiaries during our baseline, compared with 334 all-cause admissions in our weighted comparison group during that time, for a difference of 6 admissions per 1,000 beneficiaries. In 2021, Maryland's admissions had fallen to 221 per 1,000 beneficiaries. To calculate the regression-adjusted comparison group mean, we took 221, minus the estimated difference-in-differences impact of -37 admissions, minus the difference of 6 admissions from the baseline to get 252 admissions per 1,000 beneficiaries. This approach ensures that the difference between Maryland and the comparison group at baseline (first difference) minus the difference between Maryland and the comparison group in 2021 (second difference) equals the estimated impact in 2021.

Percentage impact

Percentage impacts help describe the magnitude of impact estimates on a scale common to all outcomes. In all post-baseline years (2014–2021) we calculated the percentage impact as the ratio of the impact estimate in any given year to the estimated counterfactual, multiplied by 100. The estimated counterfactual is the difference between the actual Maryland mean and the estimated impact. Using the same example as above, in 2021, we calculated the percentage impact on all-cause admissions as the impact estimate of -37 divided by the difference between the Maryland mean of 221 minus the impact estimate of -37, which equals 16.1 percent. The percentage impact for estimates of whether the Maryland Model changed outcomes more during the MD TCOC period than it did at the end of the MDAPM period are calculated slightly differently. For these, we simply subtracted the two percentage impacts; the MD TCOC period minus the end of the MDAPM period for a percentage point difference. For example, -16.1 percent (MD TCOC period percentage impact) minus -10.0 percent (MDAPM period percentage impact) equals 6.1 percentage points.

A.5. Tables of impact estimates and regression adjusted means by year

In this section, we present, in tables, regression-adjusted means as well as impact estimates of the Maryland Model by year. Using all-cause admissions as an example (Table A.8), the following is a description of how readers can interpret the tables in this section:

- The regression-adjusted means during the baseline period show little difference between the intervention and comparison groups in the admission rate (340 versus 334 admissions per 1,000 beneficiaries per year), as expected (and required through matching).
- From baseline (2011–2013) to the first year of MD TCOC (2019), admissions declined faster in Maryland than for the comparison group (87 [=340 – 253] versus 40 [=334 – 294] per 1,000 beneficiaries, respectively). Thus, the difference-in-differences estimate for the Maryland Model during the first year of the MD TCOC period was -47 admissions per 1,000 beneficiaries (=-87- (-40)). This is a 15.7 percentage point reduction (=-47 / (253 – (-47)) with a 90% CI of -55 to -39. As reflected in the 90% CI, this estimate is statistically different from zero ($p < 0.01$).[24] We calculated the impacts in 2020 and 2021 the same way.
- Combining the three estimates from 2019, 2020, and 2021, we reach a similar effect during the three years of the MD TCOC period of -44 admissions per 1,000 beneficiaries, which is statistically significant.
- We calculated the difference in estimates during the MD TCOC period and later MDAPM period in the same way, but we used the combined later MDAPM period estimates as the baseline. Using 2021 as an example, a decline in Maryland of 49 per 1,000 beneficiaries (from 270 to 221) compared with a decline in the comparison group of 42 per 1,000 beneficiaries (from 294 to 252) represents a difference-in-differences estimate of -7 admissions per 1,000 beneficiaries, which is statistically significant ($p < 0.05$) with a 90% CI of -11 to -2.
- To calculate the change in impact from the end of the MDAPM period to the MD TCOC period, we subtracted the percent impact during the end of the MDAPM period from the estimate in the MD TCOC period. Continuing the example from above, the impact estimate for admissions is 4.3 percentage points larger in 2021 (14.3%) than the estimate during the later MDAPM period (10.0%).

[24] The percentage equals the impact estimate divided by the estimated counterfactual (which equals the Maryland mean minus the impact estimate).

See Appendix A.4 for regression model specification details that produce the different impact estimates and their confidence intervals.

A.5.1. Impacts on health care utilization

Table A.8. Impacts of the Maryland Model on health care utilization

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
All-cause admissions (number per 1,000 beneficiaries per year)					
Baseline period (2011–2013)	340	334	6		
Early MDAPM period					
2014	303	304	-1	-7** (-12; -2)	-2.3%
2015	292	301	-9	-15*** (-21; -10)	-4.9%
2016	283	295	-12	-18*** (-24; -12)	-6.0%
Later MDAPM period					
2017	275	296	-21	-27*** (-34; -20)	-8.9%
2018	265	293	-28	-34*** (-41; -27)	-11.4%
Combined (2017-2018)	270	294	-24	-30*** (-37; -23)	-10.0%
MD TCOC period					
2019	253	294	-41	-47*** (-55; -39)	-15.7%
2020	214	254	-40	-46*** (-55; -38)	-17.7%
2021	221	252	-31	-37*** (-46; -29)	-14.3%
Combined (2019-2021)	229	267	-38	-44*** (-52; -35)	-16.1%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-17*** (-20; -13)	-5.7pp
2020				-16*** (-21; -12)	-7.7pp
2021				-7** (-11; -2)	-4.3pp
Combined (2019-2021)				-13*** (-17; -9)	-6.1pp
Outpatient emergency department and observation stays (number per 1,000 beneficiaries per year)					
Baseline period (2011-2013)	456	448	8		
Early MDAPM period					
2014	478	471	7	-1 (-7; 5)	-0.2%
2015	487	484	3	-5 (-13; 3)	-1.0%
2016	475	486	-11	-19*** (-27; -11)	-3.8%
Later MDAPM period					
2017	482	487	-5	-13*** (-21; -5)	-2.6%
2018	480	481	-1	-9 (-18; 0)	-1.8%
Combined (2017–2018)	481	484	-3	-11** (-19; -3)	-2.2%

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
MD TCOC period					
2019	475	480	-5	-13** (-23; -4)	-2.7%
2020	355	367	-12	-20*** (-29; -12)	-5.3%
2021	384	391	-7	-15*** (-25; -6)	-3.8%
Combined (2019–2021)	405	413	-8	-16*** (-25; -8)	-3.8%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-2 (-7; 2)	-0.5pp
2020				-9** (-16; -3)	-3.1pp
2021				-4 (-12; 3)	-1.6pp
Combined (2019–2021)				-5* (-11; -0)	-1.6pp
Intensity of hospital care (measured by standardized hospital spending) ($ per beneficiary per year)[b]					
Baseline period (2011–2013)	$4,564	$4,585	-$21		
Early MDAPM period					
2014	$4,839	$4,648	$191	$212*** ($139; $285)	4.6%
2015	$4,727	$4,743	-$16	$5 (-$43; $53)	0.1%
2016	$4,665	$4,808	-$143	-$122*** (-$171; -$73)	-2.5%
Later MDAPM period					
2017	$4,804	$4,956	-$152	-$131*** (-$191; -$71)	-2.7%
2018	$4,817	$5,155	-$338	-$317*** (-$382; -$252)	-6.2%
Combined (2017–2018)	$4,810	$5,056	-$246	-$225*** (-$283; -$166)	-4.5%
MD TCOC period					
2019	$4,883	$5,336	-$453	-$432*** (-$507; -$358)	-8.1%
2020	$4,581	$4,997	-$416	-$395*** (-$481; -$309)	-7.9%
Combined (2019–2020)	$4,731	$5,166	-$435	-$414*** (-$491; -$337)	-8.0%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-$207*** (-$251; -$164)	-3.6pp
2020				-$171*** (-$231; -$110)	-3.4pp
Combined (2019–2020)				-$189*** (-$236; -$141)	-3.5pp

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$

[a] The percentage is calculated as the impact estimate for the year divided by the estimate of the counterfactual for the year. We estimated the counterfactual as the mean outcome observed that year in Maryland minus the difference-in-differences impact estimate for the year.

[b] Standardized spending is included under “utilization” because it is an aggregate measure of intensity of hospital services, inpatient and outpatient. It removes differences in hospital spending between Maryland and the comparison group because of HSCRC rate setting and other adjustments. Standardized hospital spending in 2021 was not available for this report.

CI = confidence interval; pp= percentage point

A.5.2. Impacts on Medicare FFS spending

Table A.9. Impacts of the Maryland Model on Medicare FFS spending, dollars per beneficiary per year

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
Total Medicare FFS spending ($ per beneficiary per year)					
Baseline period (2011–2013)	$12,272	$11,097	$1,175		
Early MDAPM period					
2014	$12,089	$11,100	$989	-$186*** (-$278; -$94)	-1.5%
2015	$12,383	$11,270	$1,113	-$62 (-$152; $28)	-0.5%
2016	$12,344	$11,348	$996	-$179*** (-$266; -$92)	-1.4%
Later MDAPM period					
2017	$12,755	$11,631	$1,124	-$51 (-$181; $79)	-0.4%
2018	$13,040	$12,081	$959	-$216*** (-$347; -$85)	-1.6%
Combined (2017–2018)	$12,898	$11,857	$1,041	-$134* (-$258; -$9)	-1.0%
MD TCOC period					
2019	$13,415	$12,690	$725	-$450*** (-$595; -$304)	-3.2%
2020	$12,902	$12,153	$749	-$426*** (-$592; -$261)	-3.2%
2021	$14,102	$13,089	$1,013	-$162 (-$362; $38)	-1.1%
Combined (2019-2021)	$13,467	$12,640	$827	-$348*** (-$504; -$192)	-2.5%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-$316*** (-$394; -$238)	-2.2pp
2020				-$292*** (-$404; -$180)	-2.2pp
2021				-$28 (-$179; $122)	-0.1pp
Combined (2019–2021)				-$214*** (-$306; -$121)	-1.5pp
Hospital spending ($ per beneficiary per year)					
Baseline period (2011–2013)	$6,926	$5,417	$1,509		
Early MDAPM period					
2014	$6,799	$5,487	$1,312	-$197*** (-$268; -$125)	-2.8%
2015	$6,907	$5,559	$1,348	-$161*** (-$230; -$92)	-2.3%
2016	$6,814	$5,654	$1,160	-$349*** (-$417; -$280)	-4.9%
Later MDAPM period					
2017	$7,077	$5,824	$1,253	-$256*** (-$355; -$158)	-3.5%
2018	$7,106	$6,021	$1,085	-$424*** (-$521; -$328)	-5.6%
Combined (2017-2018)	$7,092	$5,924	$1,168	-$341*** (-$434; -$248)	-4.6%

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
MD TCOC period					
2019	$7,226	$6,297	$929	-$580*** (-$693; -$467)	-7.4%
2020	$7,083	$6,017	$1,066	-$443*** (-$577; -$309)	-5.9%
2021	$7,483	$6,481	$1,002	-$507*** (-$660; -$354)	-6.3%
Combined (2019-2021)	$7,262	$6,263	$999	-$510*** (-$633; -$387)	-6.6%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-$240*** (-$299; -$180)	-2.8pp
2020				-$102* (-$192; -$13)	-1.3pp
2021				-$166** (-$278; -$54)	-1.7pp
Combined (2019–2021)				-$169*** (-$241; -$97)	-2.0pp
Non-hospital spending ($ per beneficiary per year)					
Baseline period (2011–2013)	$5,347	$5,680	-$333		
Early MDAPM period					
2014	$5,291	$5,613	-$322	$11 (-$27; $48)	0.2%
2015	$5,477	$5,711	-$234	$99*** ($58; $140)	1.8%
2016	$5,530	$5,693	-$163	$170*** ($129; $211)	3.2%
Later MDAPM period					
2017	$5,677	$5,804	-$127	$206*** ($148; $264)	3.8%
2018	$5,935	$6,060	-$125	$208*** ($146; $270)	3.6%
Combined (2017–2018)	$5,807	$5,933	-$126	$207*** ($148; $265)	3.7%
MD TCOC period					
2019	$6,189	$6,391	-$202	$131*** ($61; $200)	2.2%
2020	$5,819	$6,135	-$316	$17 (-$55; $88)	0.3%
2021	$6,620	$6,608	$12	$345*** ($262; $428)	5.5%
Combined (2019–2021)	$6,205	$6,376	-$171	$162*** ($95; $229)	2.7%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-$76*** (-$115; -$37)	-1.5pp
2020				-$190*** (-$237; -$143)	-3.4pp
2021				$138*** ($72; $204)	1.8pp
Combined (2019–2021)				-$45* (-$84; -$5)	-1.0pp
Total Medicare spending + non-claims payments ($ per beneficiary per year)[b]					
Baseline period (2011–2013)	$12,273	$11,102	$1,171		
Early MDAPM period					
2014	$12,102	$11,114	$988	-$183*** (-$275; -$91)	-1.5%
2015	$12,417	$11,296	$1,121	-$50 (-$141; $41)	-0.4%
2016	$12,387	$11,376	$1,011	-$160*** (-$246; -$74)	-1.3%

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
Later MDAPM period					
2017	$12,770	$11,675	$1,095	-$76 (-$205; $53)	-0.6%
2018	$13,069	$12,135	$934	-$237*** (-$366; -$108)	-1.8%
Combined (2017–2018)	$12,920	$11,906	$1,014	-$157** (-$280; -$34)	-1.2%
MD TCOC period					
2019	$13,518	$12,776	$742	-$429*** (-$573; -$285)	-3.1%
2020	$13,084	$12,284	$800	-$371*** (-$537; -$204)	-2.8%
Combined (2019–2020)	$13,300	$12,529	$771	-$400*** (-$547; -$252)	-2.9%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-$272*** (-$349; -$194)	-1.9pp
2020				-$214*** (-$327; -$100)	-1.6pp
Combined (2019–2020)				-$243*** (-$326; -$159)	-1.7pp
Post-acute care spending ($ per beneficiary per year)					
Baseline period (2011–2013)	$1,149	$1,189	-$40		
Early MDAPM period					
2014	$1,095	$1,149	-$54	-$14 (-$29; $2)	-1.3%
2015	$1,123	$1,159	-$36	$4 (-$15; $22)	0.4%
2016	$1,079	$1,111	-$32	$8 (-$11; $27)	0.7%
Later MDAPM period					
2017	$1,044	$1,104	-$60	-$20 (-$43; $3)	-1.9%
2018	$1,018	$1,099	-$81	-$41** (-$68; -$13)	-3.9%
Combined (2017–2018)	$1,031	$1,101	-$70	-$30** (-$55; -$6)	-2.8%
MD TCOC period					
2019	$1,008	$1,123	-$115	-$75*** (-$106; -$44)	-6.9%
2020	$962	$1,119	-$157	-$117*** (-$153; -$80)	-10.8%
2021	$1,018	$1,093	-$75	-$35 (-$73; $4)	-3.3%
Combined (2019–2021)	$996	$1,112	-$116	-$76*** (-$108; -$44)	-7.1%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-$45*** (-$60; -$29)	-4.1pp
2020				-$86*** (-$113; -$60)	-8.0pp
2021				-$5 (-$34; $25)	-0.5pp
Combined (2019–2021)				-$46*** (-$66; -$25)	-4.3pp

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$

[a] The percentage is calculated as the impact estimate for the year divided by the estimate of the counterfactual for the year. We estimated the counterfactual as the mean outcome observed that year in Maryland minus the difference-in-differences impact estimate for the year.

[b] Total Medicare spending + non-claims payments in 2021 was not available for this report.

CI = confidence interval. pp = percentage points

A.5.3. Impacts on quality of care and population health

Table A.10. Impacts of the Maryland Model on quality of care and population health

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
Potentially preventable admissions (number per 1,000 beneficiaries per year)					
Baseline period (2011-2013)	60.6	60.4	1.0		
Early MDAPM period					
2014	53.6	52.5	1.1	0.1 (-1.1; 1.3)	0.2%
2015	52.8	51.7	1.1	0.1 (-1.3; 1.5)	0.2%
2016	49.4	49.1	0.3	-0.7 (-2.2; 0.9)	-1.4%
Later MDAPM period					
2017	49.2	51.7	-2.5	-3.5*** (-5.3; -1.7)	-6.6%
2018	45.5	50.6	-5.1	-6.1*** (-8.7; -3.6)	-11.8%
Combined (2017-2018)	47.3	51.1	-3.8	-4.8*** (-6.9; -2.7)	-9.2%
MD TCOC period					
2019	44.2	50.6	-6.4	-7.4*** (-9.8; -5.0)	-14.3%
2020	33.2	38.9	-5.7	-6.7*** (-9.0; -4.3)	-16.8%
2021	32.3	38.1	-5.8	-6.8*** (-9.4; -4.3)	-17.4%
Combined (2019-2021)	36.6	42.6	-6.0	-7.0*** (-9.3; -4.6)	-16.1%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-2.6*** (-3.5; -1.7)	-5.1pp
2020				-1.8** (-3.1; -0.6)	-7.6pp
2021				-2.0*** (-3.2; -0.8)	-8.2pp
Combined (2019-2021)				-2.1*** (-3.1; -1.2)	-6.9pp
30-day post-discharge unplanned readmissions (percentage of discharges)					
Baseline period (2011-2013)	19.1%	18.4%	0.8pp		
Early MDAPM period					
2014	17.9%	17.4%	0.5pp	-0.2pp (-0.5pp; 0.0pp)	-1.1%
2015	17.1%	17.1%	0.1pp	-0.7pp*** (-0.9pp; -0.4pp)	-3.9%
2016	16.6%	16.8%	-0.3pp	-1.0pp*** (-1.3pp; -0.8pp)	-5.7%
Later MDAPM period					
2017	16.3%	17.0%	-0.7pp	-1.5pp*** (-1.8pp; -1.1pp)	-8.4%
2018	16.4%	16.9%	-0.6pp	-1.3pp*** (-1.6pp; -1.0pp)	-7.3%
Combined (2017-2018)	16.3%	17.0%	-0.6pp	-1.4pp*** (-1.7pp; -1.1pp)	-7.9%
MD TCOC period					
2019	16.0%	17.1%	-1.2pp	-1.9pp*** (-2.3pp; -1.6pp)	-10.6%
2020	16.1%	17.2%	-1.1pp	-1.8pp*** (-2.2pp; -1.5pp)	-10.1%
2021	16.6%	17.2%	-0.6pp	-1.3pp*** (-1.6pp; -1.0pp)	-7.3%
Combined (2019-2021)	16.2%	17.2%	-0.9pp	-1.7pp*** (-2.0pp; -1.4pp)	-9.5%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-0.5pp*** (-0.8pp; -0.3pp)	-2.7pp
2020				-0.4pp** (-0.7pp; -0.1pp)	-2.2pp

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
2021				0.1pp (-0.2pp; 0.3pp)	0.6pp
Combined (2019-2021)				-0.3pp** (-0.5pp; -0.1pp)	-1.6pp
Timely follow-up after acute exacerbation of chronic conditions (percentage of discharges)					
Baseline period (2011-2013)	65.5%	65.4%	0.1pp		
Early MDAPM period					
2014	66.5%	66.6%	-0.2pp	-0.3pp (-0.8pp; 0.2pp)	-0.4%
2015	68.5%	67.2%	1.3pp	1.2pp*** (0.7pp; 1.7pp)	1.8%
2016	70.6%	68.9%	1.7pp	1.6pp*** (1.0pp; 2.2pp)	2.3%
Later MDAPM period					
2017	71.1%	69.5%	1.6pp	1.5pp*** (0.9pp; 2.1pp)	2.2%
2018	71.3%	69.8%	1.4pp	1.3pp*** (0.6pp; 2.0pp)	1.9%
Combined (2017-2018)	71.2%	69.7%	1.5pp	1.4pp*** (0.8pp; 2.0pp)	2.0%
MD TCOC period					
2019	71.7%	69.9%	1.8pp	1.7pp*** (1.0pp; 2.4pp)	2.4%
2020	64.4%	62.4%	1.9pp	1.8pp*** (1.0pp; 2.6pp)	2.9%
2021	70.4%	68.7%	1.8pp	1.6pp*** (0.9pp; 2.3pp)	2.3%
Combined (2019-2021)	69.1%	67.3%	1.8pp	1.7pp*** (1.1pp; 2.3pp)	2.5%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				0.3pp (-0.2pp; 0.7pp)	0.4pp
2020				0.4pp (-0.3pp; 1.1pp)	0.9pp
2021				0.2pp (-0.2pp; 0.7pp)	0.3pp
Combined (2019-2021)				0.3pp (-0.1pp; 0.7pp)	0.5pp
Patients' rating of their personal doctor (mean rating on a scale of 0 to 100)[b]					
Baseline period (2011-2013)	90.2	90.3	-0.1		
Early MDAPM period					
2014	90.5	90.0	0.5	0.6 (-0.2; 1.3)	0.7%
2015	88.9	90.0	-1.1	-1.0 (-2.1; 0.0)	-1.1%
2016	89.7	90.4	-0.7	-0.7 (-1.5; 0.2)	-0.8%
Later MDAPM period					
2017	90.6	90.4	0.2	0.2 (-0.7; 1.1)	0.2%
2018	90.3	90.4	-0.1	0.0 (-0.8; 0.7)	0.0%
Combined (2017-2018)	90.4	90.4	0.0	0.1 (-0.5; 0.7)	0.1%
MD TCOC period					
2019	90.9	90.6	0.3	0.4 (-0.3; 1.0)	0.4%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				0.3 (-0.5; 1.1)	0.3pp

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
Patients' rating of their hospital (percent of respondents within each hospital who rated the hospital 9 or 10 out of 10)[c]					
Baseline period (2011-2013)	66.6%	67.8%	-1.2pp		
Early MDAPM period					
2014	67.0%	68.3%	-1.4pp	-0.2pp (-1.3pp; 1.0pp)	-0.3%
2015	67.8%	68.7%	-0.9pp	0.3pp (-1.0pp; 1.6pp)	0.4%
2016[d]	n.a.	n.a.	n.a.	n.a.	n.a.
Later MDAPM period					
2017	68.9%	69.5%	-0.7pp	0.5pp (-0.7pp; 1.8pp)	0.7%
2018	68.3%	68.8%	-0.5pp	0.7pp (-0.8pp; 2.2pp)	1.0%
Combined (2017-2018)	68.6%	69.2%	-0.6pp	0.6pp (-0.7pp; 1.9pp)	0.9%
MD TCOC period					
2019	68.0%	69.0%	-1.1pp	0.1pp (-1.3pp; 1.6pp)	0.1%
2020	67.7%	68.1%	-0.3pp	0.9pp (-0.8pp; 2.6pp)	1.3%
Combined (2019-2020)	67.8%	68.5%	-0.7pp	0.5pp (-1.0pp; 2.0pp)	0.7%
Difference in estimates during MD TCOC period and later MDAPM period					
2019				-0.5pp (-1.2pp; 0.3pp)	-0.8pp
2020				0.3pp (-0.9pp; 1.4pp)	0.4pp
Combined (2019-2020)				-0.1pp (-0.9pp; 0.6pp)	-0.2pp
Use of Medicare Diabetes Prevention Program services (number using services in a year per 10,000 beneficiaries)[d]					
Baseline period (2011-2013)	0.0	0.0	0.0		
Early MDAPM period					
2014	0.0	0.0	0.0	n.a.	n.a.
2015	0.0	0.0	0.0	n.a.	n.a.
2016	0.0	0.0	0.0	n.a.	n.a.
Later MDAPM period					
2017	0.0	0.1	-0.1	-0.1 (-0.1; 0.0)	-66.7%[e]
2018	1.3	3.1	-1.8	-1.8*** (-2.7; -0.9)	-57.5%[e]
Combined (2017-2018)	0.7	1.6	-0.9	-0.9*** (-1.4; -0.5)	-57.9%[e]
MD TCOC period					
2019	0.3	1.1	-0.8	-0.8* (-1.5; 0.0)	-71.0%[e]
2020	0.3	0.4	-0.1	-0.1 (-0.4; 0.1)	-31.6%[e]
2021	0.1	0.4	-0.3	-0.3*** (-0.5; -0.2)	-75.6%[e]
Combined (2019-2021)	0.2	0.6	-0.4	-0.4** (-0.7; -0.1)	-63.5%[e]

	Regression-adjusted mean			Difference-in-differences impact estimate, by year	
	Maryland	Comparison group	Difference	Estimate (90% CI)	% Impact[a]
Difference in estimates during MD TCOC period and later MDAPM period					
2019				0.2 (-0.5; 0.8)	-13.1pp
2020				0.8*** (0.3; 1.3)	26.3pp
2021				0.6** (0.1; 1.0)	-17.7pp
Combined (2019-2021)				0.5* (0.1; 1.0)	-5.6pp

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$

[a] The percentage is calculated as the impact estimate for the year divided by the estimate of the counterfactual for the year. We estimated the counterfactual as the mean outcome observed that year in Maryland minus the difference-in-differences impact estimate for the year.

[b] Complete information on patients' rating of their personal doctor in 2020 and 2021 were not available for this report.

[c] Data on hospital ratings in 2016 were missing for several important Maryland hospitals. The analysis omits 2016 data for all hospitals to avoid spurious findings in that year.

[d] Impact estimates from 2014-2016 on use of Diabetes Prevention Program services are effectively zero because use of these services from 2014-2016 was nearly zero, before the program was rolled out and reimbursed nationally. We do not report impacts during these years. For the same reason, baseline estimates of the mean are effectively zero, but we show means in all years in the table for completeness.

[e] Percentage impacts for use of the Diabetes Prevention Program services are large even though impacts are small because the denominator used in calculating the percentage impacts (Maryland mean) was very small as well, making this number unstable, and possibly misleading. For this reason, we suppressed the calculation of the percentage impact in the main tables of this report.

CI = confidence interval; n.a. = not applicable.

A.6. Controlling for health conditions measured in Medicare claims

Our main impact estimates include controls for time-varying health conditions in the regressions to limit the potential for the higher rates of Medicare Advantage enrollment in the comparison group versus Maryland to bias estimates of model impacts. In this section, we report sensitivity results for key outcomes that remove controls for health conditions, detail our rationale for including these controls, and provide additional empirical support for our hypothesis that Medicare Advantage enrollment—if not accounted for—could bias impact estimates.

A.6.1. Rationale for including health condition controls

Estimating impacts of a policy or intervention requires researchers to consider all the ways the intervention might affect outcomes. Typically, when evaluating payment reform models such as the Maryland Model, we might consider changes in health status (as measured by individual health conditions) to be one of the mechanisms through which the model could improve outcomes. For example, if the model prompted primary care providers to better identify and treat early heart disease, the model could prevent some hospitalizations because of more serious heart conditions. In that case, we would not want to control for the time-varying prevalence of serious heart conditions after the intervention began because those controls could inadvertently remove some of the effects of the intervention itself. In Maryland, however, Medicare Advantage enrollment is lower, leading to a more consistent FFS Medicare population over time relative to a nationally drawn comparison group. These differences in the analytic sample we use to estimate impacts could lead to bias if, for example, healthier beneficiaries were more likely to leave for Medicare Advantage. This greater exit of healthier beneficiaries to Medicare Advantage in the comparison group would make the remaining beneficiaries in the comparison group look sicker relative to those in Maryland. This difference could, artificially, make it look like the Maryland Model is making beneficiaries healthier, when really the difference in health status is just because of who is exiting to Medicare Advantage. In that case, we *would* want to control for time-varying prevalence of conditions to account for changes in the population that are unrelated to the effects of the Maryland Model itself.

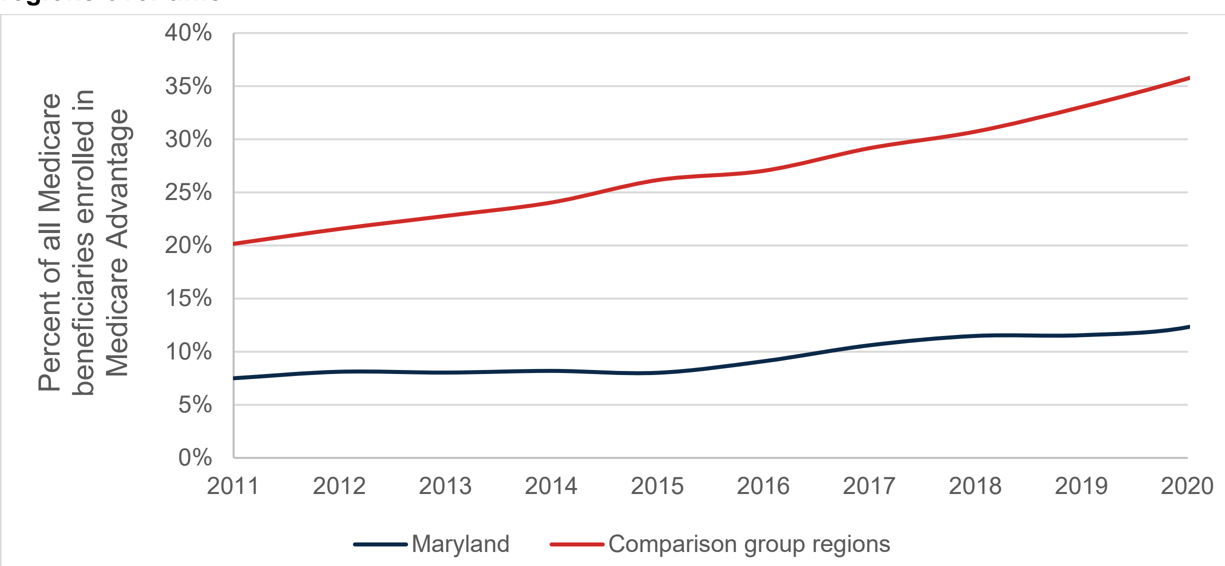
To estimate impacts of the Maryland Model, we are balancing the concern of over-controlling for model impacts with the threat of bias that stems from higher rates of Medicare Advantage enrollment in the intervention group than in the comparison group.

Medicare Advantage enrollment in Maryland and the comparison group

Medicare Advantage enrollment has been growing nationwide since the early 2000s, but Medicare Advantage enrollment in Maryland is much lower than in the rest of the nation (Figure A.5), potentially because of, in part, Maryland's all-payer rate-setting system that began in the 1970s (HSCRC 2020b). For most of the baseline (2011–2013) and MDAPM periods (2014–2018), Medicare Advantage enrollment in Maryland was less than 10 percent and relatively stable over time. In contrast, in the comparison group regions, Medicare Advantage enrollment grew steadily from about 20 percent in 2011 to more than 30 percent by the start of the MD TCOC period in 2019.

Fundamentally, because most of the outcomes we plan to estimate impacts on for this evaluation are measured in the FFS Medicare population, this exit to Medicare Advantage means that our analytic sample is changing over time, and it is changing differently for Maryland and the comparison group. Traditionally, Medicare Advantage beneficiaries have been healthier than FFS beneficiaries (Shimada 2009), and we see this in our analysis as well. As healthier beneficiaries exit the FFS analytic sample more quickly in the comparison group than in the intervention group, the beneficiaries remaining in the FFS analytic sample could look sicker in the comparison group than in Maryland. This difference alone, if not accounted for, could make it look the Maryland Model improved outcomes in Maryland (by, for example, lowering spending and reducing utilization), biasing the impact estimates.

Figure A.5. Trends in Medicare Advantage enrollment in Maryland and the comparison group regions over time



Understanding whether differences in Medicare Advantage can explain differences in condition prevalence

Beneficiaries are moving in and out of our comparison group and intervention group every year of our analysis (because we define those groups each year as all Medicare FFS beneficiaries living in Maryland or the comparison group PUMAs that year). The difference-in-differences strategy we use does not require that the *same* people be present in our analysis year to year, but it *does* assume (without additional controls) that the characteristics of the populations in Maryland and the comparison group change in similar ways over time—or, if they change differentially, that those differences are because of the impact of the Maryland Model and not some external factor such as Medicare Advantage enrollment. We know that (1) entry into Medicare Advantage (and therefore exit from the analytic sample) is higher in the comparison group than in the intervention group (Figure A.5) and, (2) *conceptually,* this difference could lead the sample in the comparison group to become healthier than the intervention group on measured conditions for reasons unrelated to model impacts. The *empirical* question we tried to answer below is whether the movement in and out of our analytic sample for Medicare Advantage can explain differences in the prevalence of chronic conditions between the intervention and comparison groups large enough to matter to the evaluation on its own.

We considered the Medicare Advantage influence to be large enough to matter if that influence alone could lead to divergences between the intervention and comparison groups in the prevalence of key chronic conditions that are similar in size to the divergences we're observing in the actual intervention and comparison groups. We focused on three key chronic conditions: anemia, chronic kidney disease, and congestive heart failure to demonstrate the effects of movement in and out of the analytic population because early testing of regression models demonstrated that including or not including these conditions could influence impact results.

To understand movement in and out of our analytic sample in each year, we constructed mutually exclusive groups that represent whether each beneficiary stayed in our sample from the prior year, newly entered our sample that year, exited our sample that year, or both entered and exited that year (Table A.11). All beneficiaries in our sample each year fall into one of the categories in Table A.11. We did this separately for Maryland and the comparison group to see how entry and exit differed between the groups. Our key observations include the following:

- In Maryland, the total analytic sample is growing. This is largely because of a high rate of new entrants to Medicare and a low rate of people exiting to Medicare Advantage. In the comparison group, the total analytic sample grows until about 2016 and then starts to decline, largely because of an increase in the rate of Medicare Advantage exit. This mirrors what we know about FFS and Medicare Advantage enrollment nationally.
- As expected, particularly in earlier years, many fewer beneficiaries exit to Medicare Advantage in Maryland than exit to Medicare Advantage in the comparison group. But Medicare Advantage isn't just less common in Maryland; the *growth* in Medicare Advantage enrollment is slower in Maryland. Except in 2020, the proportion of beneficiaries leaving for Medicare Advantage in Maryland is largely similar across years or just marginally growing (1.1 percent in 2013 versus 1.5 percent in 2019), whereas in the comparison group, the proportion of beneficiaries leaving for Medicare Advantage is higher and growing faster (2.9 percent in 2013 versus 4.7 percent in 2019). Rates of exit to Medicare Advantage were higher in 2020 in Maryland and the comparison group, but the rates were still higher in the comparison group overall (5.5 percent versus 3.9 percent).

- There are reasons other than Medicare Advantage enrollment that beneficiaries enter or exit our analytic sample. New entrants to Medicare and exits because of death are the most common and have somewhat similar levels and trends in Maryland and the comparison group. Entry into the intervention group from beneficiaries moving into the relevant region is higher for the intervention group (that is, people moving into Maryland) than the comparison group (that is, people moving into the comparison regions).

Table A.11. Entry and exit from the analytic sample in Maryland and the comparison group over time

Year	Analytic sample	Net change	Percentage of analytic sample: Stayer[a]	Enter (Medicare Advantage)	Enter (treatment or comparison region)	Enter (new to Medicare)	Exit (death)	Exit (Medicare Advantage)	Exit (region, unknown or unobserved)	Enter and exit (death)	Enter and exit (not death)
Maryland											
2012	693,349	2.9%	83.2%	0.4%	0.8%	8.3%	4.0%	0.9%	0.3%	0.2%	2.9%
2013	718,477	3.5%	82.8%	0.7%	0.9%	7.9%	4.0%	1.1%	0.3%	0.2%	3.5%
2014	740,351	3.0%	83.1%	0.7%	0.9%	7.7%	3.9%	1.0%	0.3%	0.2%	3.0%
2015	764,546	3.2%	82.4%	1.0%	0.9%	7.3%	3.9%	1.6%	0.3%	0.2%	3.2%
2016	773,351	1.1%	82.8%	0.5%	0.9%	7.0%	3.9%	2.3%	0.4%	0.2%	1.1%
2017	777,643	0.6%	83.4%	0.5%	0.9%	7.0%	3.9%	1.6%	0.4%	0.2%	0.6%
2018	785,906	1.1%	83.9%	0.5%	0.9%	6.8%	3.8%	1.2%	0.4%	0.2%	1.1%
2019	802,985	2.1%	83.1%	0.9%	1.0%	6.7%	3.8%	1.5%	0.4%	0.2%	2.1%
2020	812,597	1.2%	81.0%	0.5%	0.9%	6.5%	4.3%	3.9%	0.3%	0.2%	1.2%
Comparison group											
2012	8,223,837	1.5%	83.1%	0.7%	0.1%	7.7%	4.2%	2.6%	0.3%	0.1%	1.5%
2013	8,345,950	1.5%	82.8%	0.9%	0.1%	7.3%	4.1%	2.9%	0.3%	0.2%	1.5%
2014	8,459,879	1.3%	81.4%	1.2%	0.1%	7.2%	4.0%	4.0%	0.3%	0.2%	1.3%
2015	8,433,778	-0.3%	82.8%	1.2%	0.1%	7.1%	4.0%	2.7%	0.3%	0.2%	-0.3%
2016	8,549,484	1.4%	81.3%	1.3%	0.1%	6.9%	4.0%	4.3%	0.4%	0.2%	1.4%
2017	8,473,158	-0.9%	82.2%	1.1%	0.1%	6.9%	4.0%	3.7%	0.4%	0.2%	-0.9%
2018	8,464,451	-0.1%	81.2%	1.1%	0.1%	6.7%	3.9%	4.5%	0.4%	0.2%	-0.1%
2019	8,338,410	-1.5%	81.3%	0.8%	0.1%	6.5%	3.9%	4.7%	0.4%	0.2%	-1.5%
2020	8,167,111	-2.1%	80.1%	0.6%	0.1%	6.4%	4.5%	5.5%	0.3%	0.2%	-2.1%

Note: The table shows the number of Medicare fee-for-service beneficiaries in our analytic sample by year, and how the sample moves across years. In column headers, parentheses after Enter denote where the beneficiary came from, and parentheses after Exit denote where the beneficiary went.

[a] Stayer means that the beneficiary was in our sample in that year as well as the prior year.

In Table A.11, we observed that the rate of Medicare Advantage exit was higher and growing faster in our comparison group than in Maryland. Next, we investigated whether beneficiaries leaving for Medicare Advantage have better or worse underlying health that could lead to differences in health of the FFS population remaining in the analytic population over time. To do this, we calculated the prevalence of individual CCW conditions (specifically, those that we identified above as having a meaningful influence on our impact results) for Maryland and the comparison group for each of the mutually exclusive categories defined in Table A.11. From this, we saw that beneficiaries who exit to Medicare Advantage have modestly lower rates of anemia than those in the full sample, particularly for the comparison group (Table A.12). For example, in 2019, 20.9 percent of comparison-group beneficiaries exiting to Medicare Advantage had anemia, whereas 23.5 percent of the full analytic sample that year had anemia. We focus on anemia because adding or removing this specific control variable materially affects the size of the impact estimates. We saw similar patterns for other conditions, such as chronic kidney disease and congestive heart failure, whose inclusion or exclusion from the regressions affect the impact results. That is, those exiting to Medicare Advantage generally had modestly lower rates of kidney disease or congestive heart failure than the full set of beneficiaries in the analytic sample.

Table A.12. Prevalence of anemia by entry/exit group in Maryland and comparison group over time

Year	Analytic sample	Stayer	Enter (Medicare Advantage)	Enter (treatment or comparison region)	Enter (new to Medicare)	Exit (death)	Exit (Medicare Advantage)	Exit (region, unknown or unobserved)	Enter and Exit (death)	Enter and exit (not death)
Maryland										
2012	27.3%	26.4%	n.a.	17.8%	13.7%	53.6%	26.3%	7.5%	50.0%	13.1%
2013	27.0%	26.0%	n.a.	18.1%	15.4%	53.9%	26.8%	7.9%	43.3%	13.0%
2014	26.2%	25.3%	n.a.	16.8%	14.4%	52.8%	26.1%	6.9%	51.8%	14.3%
2015	25.2%	24.3%	n.a.	16.0%	15.0%	52.4%	23.7%	8.2%	49.0%	14.1%
2016	24.3%	23.4%	n.a.	15.8%	13.6%	50.9%	23.9%	8.0%	47.8%	12.8%
2017	23.7%	22.8%	n.a.	14.6%	14.1%	50.5%	22.4%	8.2%	42.1%	12.5%
2018	24.2%	23.3%	n.a.	15.2%	13.4%	50.8%	24.3%	9.0%	43.2%	13.7%
2019	24.4%	23.4%	n.a.	15.2%	15.2%	50.7%	25.6%	8.2%	43.4%	14.0%
2020	24.5%	23.3%	n.a.	14.9%	15.1%	50.8%	24.5%	9.3%	43.5%	12.5%
Comparison group										
2012	25.7%	25.0%	n.a.	20.6%	13.3%	51.9%	22.0%	7.7%	41.1%	9.2%
2013	25.4%	24.6%	n.a.	18.6%	13.2%	51.5%	22.2%	7.6%	45.1%	11.7%
2014	24.8%	24.1%	n.a.	16.6%	13.2%	50.8%	21.9%	7.6%	43.4%	13.3%
2015	24.1%	23.3%	n.a.	15.9%	12.7%	50.4%	21.6%	7.1%	45.1%	11.3%
2016	23.6%	22.7%	n.a.	15.3%	13.3%	50.2%	23.0%	8.4%	43.2%	11.4%
2017	23.2%	22.3%	n.a.	14.1%	13.0%	49.8%	21.2%	8.2%	44.6%	11.2%
2018	23.4%	22.6%	n.a.	16.1%	12.5%	50.2%	22.1%	8.0%	43.9%	11.5%
2019	23.5%	22.7%	n.a.	15.9%	13.2%	50.4%	20.9%	8.5%	40.6%	12.1%
2020	23.6%	22.6%	n.a.	15.6%	13.0%	50.6%	22.3%	8.8%	43.0%	11.6%

Note: The table shows, for all Medicare FFS beneficiaries in our analytic sample by year, the percentage diagnosed with CCW anemia. Stayer means that the beneficiaries was in our sample in that year as well as the prior year. CCW conditions (including anemia) require at least a year of FFS claims history for identification. Beneficiaries entering from Medicare Advantage do not have a year of FFS claims history and therefore do not have CCWs identified. In column headers, parentheses after Enter denote where the beneficiary came from, and parentheses after Exit denote where the beneficiary went.

CCW = Chronic Conditions Warehouse; FFS = fee for service; n.a. = not applicable.

The evidence from Table A.12 is important because, together with Table A.11, it establishes, *in our data*, what we hypothesized: that Medicare Advantage enrollment rates are higher in Maryland than in the comparison group *and* that the beneficiaries who are leaving have lower prevalence of conditions known to be associated with spending outcomes.

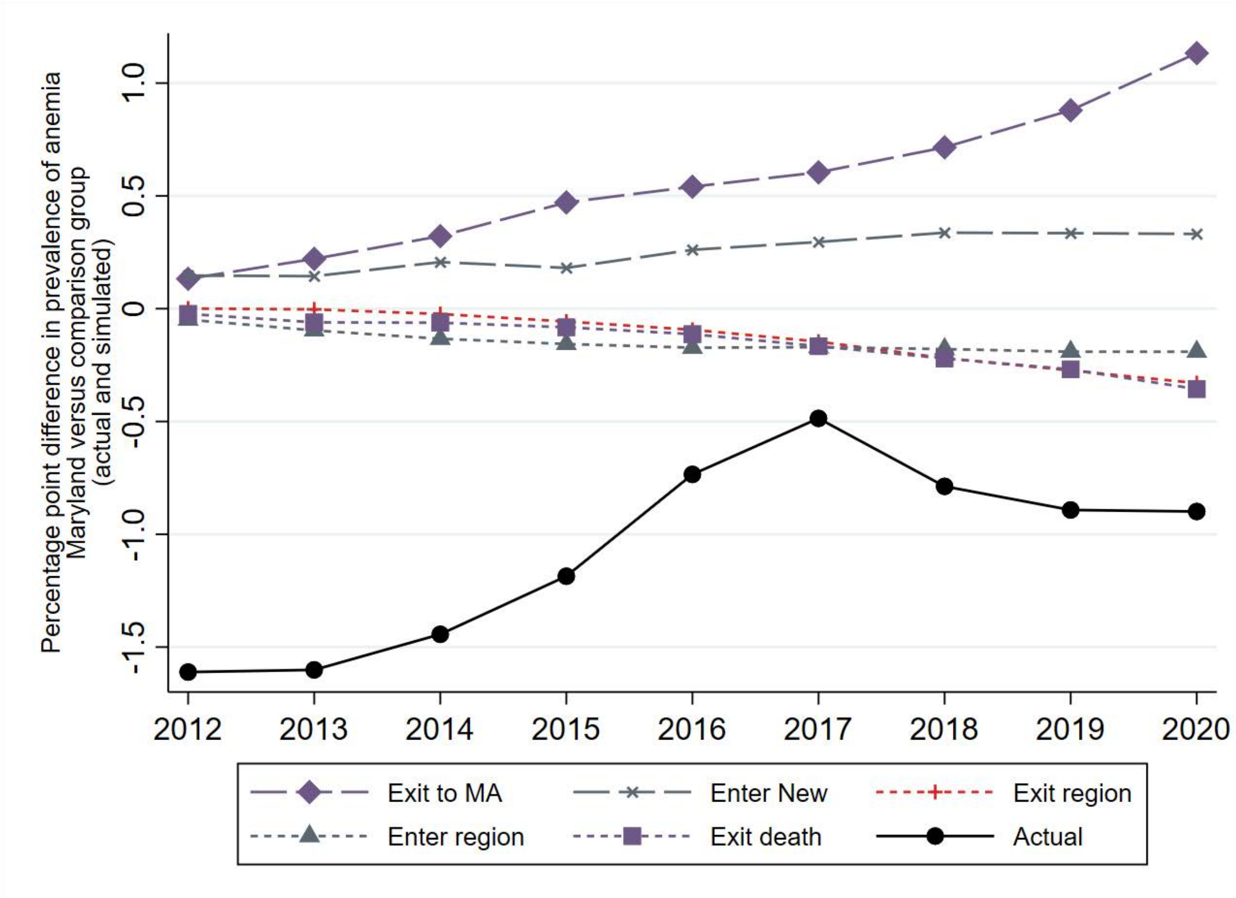
But the evidence in Table A.12 does not, by itself, tell us the final impact of this change in our sample or specifically whether the differences we observe in Medicare Advantage enrollment could be mostly or entirely driving observed differences in population health status. To do this, we set up a small, simulated population exercise designed to hold constant all but one type of movement in and out of the simulated population at a time. Isolating individual entry and exit groups helps us to understand the impact that each specific type of movement has on health conditions alone. Here, we use an example to illustrate how we produced these results.

Simulated population exercise, exiting to Medicare Advantage

- First, we started with the unweighted analytic sample population in 2011 (672,898 in Maryland).
- We then moved the population forward one year to 2012 and assumed no entry or exit *except* for the exiting to Medicare Advantage. To do that, we multiplied the rate of exit to Medicare Advantage in Maryland in 2012 (0.9 percent) by the 2011 population (672,898*0.09 = 5,899)[25] and subtracted those beneficiaries from the 2011 population (672,898 - 5,899 = 666,999) to get the 2012 population.
- Then, we calculated the expected 2012 prevalence of anemia in this new simulated population. To do that, we took what we know about the prevalence of anemia in the full population in 2011, as well as the prevalence of anemia in the people who exited to Medicare Advantage in 2012 (Table A.12), to calculate the average anemia prevalence among those *not* exiting to Medicare Advantage (the remaining population). This value became the expected prevalence in 2012 of anemia, if the *only* movement was from those exiting to Medicare Advantage.
- We then repeated the process from the prior two steps through to 2020 (using information in each year that is based on the prior year's simulated population, *not* the full analytic population, which is subject to all inflows and outflows) and separately for Maryland and the comparison group.
- Finally, we compared the expected prevalence of anemia in each year between Maryland and the comparison group and took the difference.[26] That difference can be interpreted as the cumulative difference in expected prevalence of anemia between Maryland and our comparison group that comes exclusively from movement out of our sample into Medicare Advantage (purple line with diamond markers). We plot the difference, alongside the *actual* difference in prevalence of anemia between Maryland and the comparison group in Figure A.6 (solid black line).

[25] Rounding from the rate of Medicare Advantage means this calculation is close but not exact.
[26] To help with interpretability, we also remove from this difference in each year the baseline difference in prevalence of anemia between Maryland and the comparison group. That helps ensure the simulated population starts from zero difference and is easier to read the magnitude of the differences.

Figure A.6. Actual and projected differences in anemia prevalence between Maryland and the comparison group because of simulated entry to, and exit from, the analytic sample



Note: For ease of viewing, the figure plots only those inflows and outflows that represent a sizable portion of the total analytic sample and represent deviations from zero on this plot. Enter New means new entrants to Medicare. Enter/exit region means enter or exit treatment or comparison region. Exit to MA means exit to Medicare Advantage.

MA = Medicare Advantage.

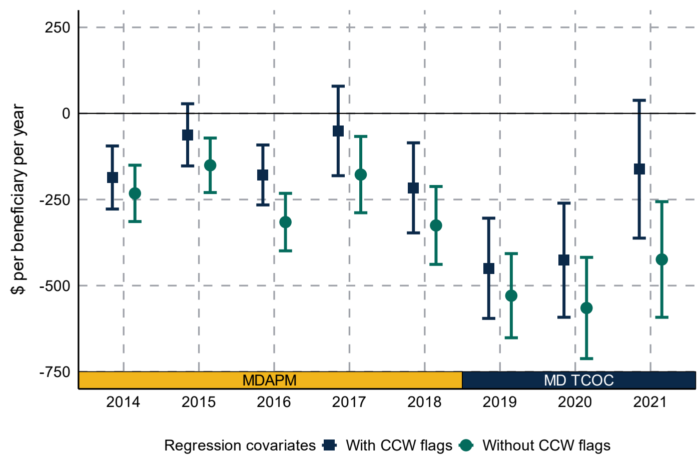
In Figure A.6, we see from the Exit to MA line that we might expect a cumulative difference of more than 1 percentage point in the prevalence of anemia just from the movement out of our sample and into Medicare Advantage alone. Against the scale of the actual differences in prevalence between Maryland and the comparison group (solid black line, ranging from -1.5 percent to -0.5 percent), we think this represents a meaningful projected difference in prevalence that, if not corrected for, could lead to bias in our results. In general, for other conditions such as chronic kidney disease or congestive heart failure, we see patterns that reaffirm the conclusion that changes because of movement in or out of the sample alone are of notable magnitude relative to the prevalence of those conditions. This leads us to prefer to control for these differences directly using time-varying health conditions to avoid introducing bias.

The exercises from the sections above show that Medicare Advantage enrollment alone can lead to sizeable and increasing differences in prevalence in Maryland and the comparison group for conditions that we know can matter for regression adjustment. Further, these differences are of similar magnitude and direction of the observed differences—though other factors influence trends as well. Those other forces can include other reasons for entering and exiting the sample, which are also not because of model impacts and can be partly corrected for with time-varying health condition controls. But they could also include some true model effects on conditions—we can't ever rule these out entirely—meaning that presenting sensitivity results, such as those in A.6.2, still offer important additional context and interpretation of results. Ultimately, given the large differences in Medicare Advantage enrollment between Maryland and the comparison group, and the evidence that those differences alone can drive prevalence differences of the size observed, we believe that controlling for health conditions remains the best path to mitigate bias risk in the evaluation, although we acknowledge it could introduce some small risk of bias through overcontrolling for impacts and so it is important to present results that bound that risk.

A.6.2. Results of the sensitivity test that removes health condition controls

- In general, we find that impacts on key outcomes, particularly spending outcomes, were qualitatively consistent but moderately larger when we removed health condition controls (Figures A.7 and A.8).
- The difference between models that did and did not control for health conditions was largest in 2020 and 2021. In general, the difference is growing over time, which is consistent with the idea that the population in Maryland and the comparison group continue to diverge on health status because of differential enrollment in Medicare Advantage.
- Impacts on utilization and quality of care outcomes such as all-cause, acute-care hospitalizations and preventable hospitalizations were only minorly affected by the decision to include or not include chronic health conditions (Figure A.7)

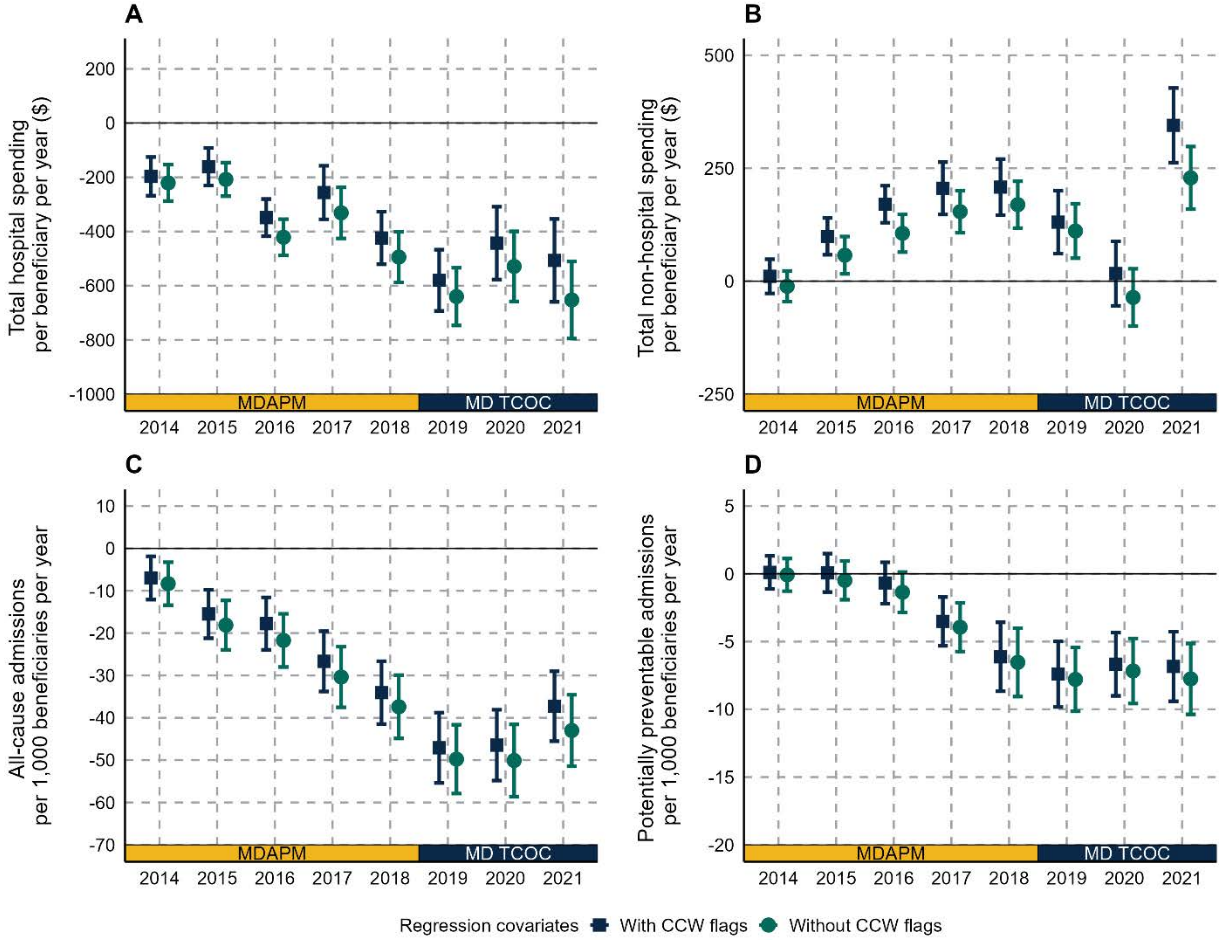
Figure A.7. Estimated impact of the Maryland Model on total spending with and without time-varying health condition controls, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CCW = Chronic Conditions Data Warehouse; CI = confidence interval.

Figure A.8. Impact of the Maryland Model on key measures with and without time-varying health condition controls, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval.

Appendix B: Measures, definitions and file construction

This appendix describes how we constructed claims-based outcomes measures and survey-based patient experience measures, PUMA-level matching variables, and regression covariates for the impact analyses in this report. We first describe in detail how we defined the outcomes measures, starting with the claims-based measures and then the survey-based patient experience measures. For the claims-based measures, we organized this appendix by whether they are measured at the beneficiary-year level or the discharge level (Section B.1). We then describe how we rolled up the claims-based beneficiary- and discharge-level outcomes measures along with the survey measures and other claims- and non-claims-based measures—including beneficiaries' demographic and enrollment characteristics, health status measures, and geographic characteristics—to develop PUMA-level matching variables (Section B.2.1). Finally, we describe the analysis files used for beneficiary- and discharge-level impact models, including definitions of covariates constructed from claims, enrollment, area-level, and patient survey data (Section B.2.2) as well as the files constructed at the hospital-level for analyses of hospital-based patient experience (Section B.2.3).

We constructed annual files with outcomes, matching variables, and regression covariates –for beneficiary-level, discharge-level (episodes), and hospital-level outcomes. The annual claims-based beneficiary file contains one observation per beneficiary per year for all beneficiaries who were observable for at least one month in Medicare FFS claims data during the year (that is, they were alive, enrolled in Medicare Parts A and B FFS, and had Medicare as primary payer). Beneficiaries can be in the file in all years of our analytic period or only one or a limited number of years, depending on their observability status. The annual discharge file contains discharges paid for by FFS Medicare that met denominator inclusion criteria for 30-day unplanned readmissions (Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation, 2020) or follow-up after acute exacerbations of chronic conditions (IMPAQ Health, 2018). For analyses of patients' ratings of their personal doctors, the annual file contains survey responses from Consumer Assessment of Healthcare Providers and Systems (CAHPS) from Medicare FFS and Medicare Advantage respondents in each year along with demographic and enrollment-related characteristics from Medicare administrative files. For analyses of hospital ratings, the analysis file contains hospital-year-level average annual ratings along with covariates that measure characteristics of the hospitals.

B.1. Measures and definitions

Claims-based outcomes measures constructed at the beneficiary-year level

To construct claims-based outcomes at the beneficiary-year level, we relied on the Medicare FFS Research Identifiable Files (RIFs) claims data from the Virtual Research Data Center. These files provide data on all services funded by Medicare FFS. We used claims data with at least 90 days of runout at the time we pulled the data, the standard for evaluation purposes. We used all claims to measure outcomes, regardless of geography. For example, we included all Medicare claims for a Maryland resident, regardless of whether the beneficiary received the covered services from providers in Maryland or elsewhere. We supplemented these data, as described later, with data from the Medicare Geographic Variation Data Base (GVDB) to measure standardized hospital spending.

B.1.1. Medicare spending measures

Our measures of Medicare spending include Medicare payments recorded in Parts A and B RIF claims data. For all spending measures, we started by assigning the amount Medicare paid for each service to a year based on the end date (or through date) on the claim. The one exception was for post-acute care claims, for which the services provided can often span many months even if paid in only a single month. In those cases, we apportioned the spending or service use recorded on the claim according to the number of post-acute care days falling in the respective years.

We then summed Part A and Part B payments for the months that a beneficiary was observable in FFS claims that year (that is, the beneficiary was alive, enrolled in Medicare Parts A and B FFS, and had Medicare as primary payer) and annualized the payments to account for the number of months the beneficiary was observable in FFS claims. For example, if a beneficiary was observable for 10 of 12 months of the year, and we observed $10,000 in Medicare Parts A and B payments for this beneficiary over 10 months, then their annualized spending would be ($12,000). These amounts exclude the amounts that third parties and beneficiaries paid for deductibles, coinsurance, and copayments. They also exclude Medicare payments for Part D prescription drugs and any Medicare payment amounts on home health interim RAP (request for anticipated payment) claims. We set negative Medicare payments to zero.

Total Medicare FFS spending (dollars per beneficiary per year)

This outcome measures Medicare spending, in dollars per beneficiary per year, for Parts A and B covered services during the year among beneficiaries who were observable for at least one month during the year. It is the sum of Medicare payments across inpatient, outpatient, skilled nursing facility (SNF), home health, hospice, carrier (or Part B), and durable medical equipment claims. This variable excludes nonclaims payments (that is, payments from the CMS to providers that were made separately from claims).

Medicare FFS spending, hospital and non-hospital spending

We also measured Medicare FFS spending for Parts A and B covered services during the year stratified by type of service: hospital and non-hospital spending. Specifically, we constructed the following categories:

1. **Hospital spending** includes spending for Part A inpatient and Part B outpatient claims at short-stay acute care hospitals, critical access hospitals, children's hospitals, inpatient rehabilitation hospitals, long-term care hospitals, and psychiatric hospitals.
2. **Non-hospital spending** measures the sum of all Parts A and B spending that was not classified as hospital spending according to the earlier definition. Specifically, non-hospital spending is the sum of the following measures:
   - 2.1. **Post-acute care spending** measures the sum of Part A spending for SNF and home health services, defined as follows:
     - 2.1.1. **SNF spending** measures all spending for service use recorded in the SNF claims file. It includes spending for SNF services provided in swing beds in short-term acute care hospitals.
     - 2.1.2. **Home health visit Medicare Part A spending** measures Medicare Part A spending for service use recorded in the home health agency claims file. Medicare Part B also covers home health care, but Part A provides coverage following a qualifying

inpatient hospital stay. This measure aims to capture post-acute care home health spending, so we limited spending to home health care claims covered by Part A, including (a small number of) claims covered by Medicare Parts A and B.

2.2. Ambulatory care visit with primary care providers and specialist physicians spending is the sum of the two ambulatory care visit spending measures below.

2.2.1. Ambulatory care visit with primary care provider spending measures Medicare Part B professional (carrier claim) spending for ambulatory visits with primary care practitioners, nurse practitioners (NPs), physician assistants (PAs), and other advanced practice nurses (APNs). It also includes Part B outpatient spending for ambulatory visits at clinics (Federally Qualified Health Centers and Rural Health Clinics).[27]

2.2.2. Ambulatory care visit with specialist physicians spending measures Medicare Part B professional (carrier claim) spending for ambulatory visits with specialist physicians.[27]

2.3. Non-hospital Part B drug spending measures spending for drugs covered by Medicare Part B that is not classified earlier as hospital spending. Specifically, we identified Medicare spending for claims lines in the non-hospital outpatient claims, carrier claims, and durable medical equipment claims files in which the procedure (Healthcare Common Procedural Coding System, or HCPCS) code was for a drug paid for under the average sales price payment system.

2.4. Ambulatory surgical center facility spending measures facility charges for services at ambulatory surgical centers. Ambulatory surgical center claims were identified by the claim type of service code (“F”). Spending on Part B drugs was excluded (because this spending was captured in the measure described before).

2.5. Imaging and testing professional spending measures spending for professional services associated with imaging and testing. Specifically, it includes spending for claim lines in the carrier claims file in which the procedure code was classified as imaging or testing according to the Berenson Eggers Type of Service (BETOS) or Restructured BETOS Classification System (RBCS) algorithm (CMS 2022a) (we applied the RBCS to all claims, but because the RBCS includes Medicare-covered procedure codes starting in 2014 only, we back-filled the imaging and testing variable in the 2011 through 2013 carrier files with any codes that the RBCS did not classify, but that the BETOS algorithm classified as imaging and testing) . Professional spending excludes any outpatient facility charges for imaging and testing conducted in settings for which outpatient facility claims are also submitted.

2.6. Other non-hospital spending measures the sum of all Parts A and B spending not captured by any of the measures described before. This measure includes Medicare Part A spending on non-hospital inpatient services[28] and hospice; Part B spending on home health care and ambulatory care visits with behavioral health providers, and Part B spending for non-hospital outpatient, professional (carrier) services, and durable medical equipment not otherwise

[27] Some primary care providers and specialists participating in Comprehensive Primary Care Plus, Global and Professional Direct Contracting, and MDPCP have their Medicare payment amounts on ambulatory claims adjusted downwards because these visits are otherwise covered partially or wholly under capitated arrangements with CMS. We removed these adjustments to obtain the amount Medicare would have paid under FFS (in the absence of capitation) for these visits.
[28] This category includes claims from facilities that are excluded from our definition of hospital spending, such as religious non-medical health care institutions.

captured in the measures before (for example, not previously categorized as spending on Part B drugs).

3. **Total Medicare spending plus non-claims payments** measures total spending, plus payments made in support of alternative payment models. Specifically, it includes, in Maryland and the comparison group, when applicable, payments for the following programs: Pioneer ACO, ACO Shared Savings Program (MSSP), Next Generation ACO, Comprehensive Primary Care Initiative (CPCI), Comprehensive Primary Care Plus (CPC+), MDPCP, and payments to providers who participated in advanced alternative payment models under the Quality Payment Program (QPP). For MDPCP, CPCI and CPC+, payments include all payments, including care management fees, performance-based incentive payments (PBIP), and comprehensive primary care payments (CPCP), where applicable.

B.1.2. Service use measures

Intensity of hospital care (measured by standardized hospital spending)

We computed measures of annualized standardized hospital spending using the Medicare GVDB, produced by the CMS Office of Information Products and Data Analytics. The database includes claim-level standardized payment amounts for Part A claims (inpatient, SNF, hospice, and home health) and Part B institutional (outpatient) claims. We merged the standardized payment amounts onto the RIF files (at the claim level for Part A claims and Part B institutional claims). Then we calculated standardized hospital payments across the same set of claims in the hospital spending category described above with the standardized payment amounts from the GVDB in place of actual hospital payment amounts. Standardized spending removes differences in spending across claims because of difference in the prices paid to different providers (for example, those from wage indices in different parts of the country or HSCRC rate setting), so it measures intensity of service use in aggregate.

All-cause acute care hospital admissions (number of admissions per beneficiary per year)

This measure is the annualized number of hospitalizations for short-stay acute hospitals, critical access hospitals, and children's hospital admissions reported in the RIF inpatient claims file for the beneficiary during the year. Multiple claims for acute admissions that involved transfers between hospitals were combined into a single record, as were multiple claims for the same beneficiary at the same facility with overlapping dates, so these count as one admission. We excluded hospitalizations for psychiatric care, inpatient rehabilitation stays, and long-term hospital stays.

Outpatient ED visits and observation stays (number of visits per beneficiary per year)

This measure is the annualized number of outpatient ED visits and observation stays for the beneficiary during the year that do not lead to a hospitalization. Visits that do not lead to a hospitalization are identified in the outpatient department RIF hospital claims file using revenue center line items equal to 045X or 0981 (emergency room care), 0762 (treatment or observation room), or 0760 (treatment or observation room—general classification). We counted a visit as an observation stay if it was longer than eight hours and had a corresponding HCPCS code of G0378 (hospital observation services per hour). We then capped the number of either type of visit (observation stays and ED visits) to one per day.

B.1.3. Quality of care measures

Potentially preventable admissions (number of admissions per beneficiary per year)

This measure is the annualized number of hospitalizations for short-stay acute hospital, critical access hospital, and children's hospital admissions reported in the inpatient claims file for the beneficiary during the year in which the admission met the criteria for the Prevention Quality Indicators (PQI) overall composite measure (PQI #90). To construct this measure, we applied the Agency for Healthcare Research and Quality's 2020 Quality Indicators Software to all inpatient hospital claims for acute stays (defined earlier) and then counted the number of hospital admissions for the beneficiary each year that the software flagged as being admissions for one of the following PQIs: diabetes short-term complications (PQI #01), diabetes long-term complications (PQI #03), chronic obstructive pulmonary disease or asthma in older adults (PQI #05), hypertension (PQI #07), heart failure (PQI #08), community-acquired pneumonia (PQI #11), urinary tract infection (PQI #12), uncontrolled diabetes (PQI #14), asthma in younger adults (PQI #15), or lower-extremity amputation among patients with diabetes (PQI #16) (AHRQ n.d.).

B.1.4. Population health measures

Use of Medicare Diabetes Prevention Program Services (yes or no for the beneficiary during the year)

This measures whether the beneficiary received any Medicare Diabetes Prevention Program (MDPP) services during the year (yes or no). A beneficiary was considered to have received Medicare Diabetes Prevention Program services if they had at least one outpatient or carrier claim with procedure code 0403T, 0488T, G9873, G9874, G9875, G9876, G9877, G9878, G9879, G9882, G9883, G9884, G9885, G9880, G9881, G9890, or G9891. Medicare started funding Medicare Diabetes Prevention Program services in 2018. Therefore, this outcome will have a value of 0 for all beneficiaries from 2011 to 2017.

B.1.5 Quality of care outcomes measured at the discharge-year level

30-day post-discharge unplanned readmission (yes or no for the event)

We used Medicare FFS RIF inpatient claims and enrollment data for this measure. The analytic file has one observation for each inpatient discharge. Beneficiaries can be included in the file once, more than once, or not at all depending on how many discharges they had. Multiple claims for acute admissions that involved transfers between hospitals were combined into a single record, as were multiple claims for the same beneficiary at the same facility with overlapping dates, so these count as one discharge.

The all-cause 30-day post-discharge unplanned readmission measure indicates whether the discharge (the index admission) was followed by an unplanned hospital admission within 30 days. An unplanned readmission is defined as any hospitalization that does not follow an established plan of care (examples of planned admissions include those for chemotherapy and planned admission for transplant surgery). The measure equals 1 if there was an unplanned readmission within 30 days of discharge to any hospital, regardless of whether the readmission occurred at the same hospital or a different hospital. The measure equals 0 if there was no unplanned readmission within 30 days.

Our definition of this measure is based on the Yale readmission measure developed by the Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (2020) used in the Hospital Readmission Reduction Program under Section 3025 of the Affordable Care Act. An admission that counts as a readmission because it fell within 30 days of an earlier index stay can also count as an index stay for a potential subsequent readmission as long as it meets the index admission inclusion criteria. We count an index admission in a year if the discharge date is in that year. We then look for an unplanned readmission within 30-days of that index admission (the readmission could occur in the following year)

Timely follow-up after acute exacerbations of chronic conditions (yes or no for the event)

This measures whether follow-up was received within the time frame recommended by clinical practice guidelines in a non-emergency outpatient setting following an ED visit or hospitalization for one of the following six chronic conditions: hypertension, asthma, heart failure, coronary artery disease, chronic obstructive pulmonary disease, or diabetes mellitus (Type I or Type II). IMPAQ Health (2018) developed the measure specifications. HSCRC has included improvement on this measure as one of its quality goals in Statewide Integrated Health Improvement Strategy.

To develop this measure, we first identified hospital admissions and outpatient emergency visits and observation stays that met the denominator criteria for one of the six chronic conditions. Unlike the readmission measure defined before, this measure is not strictly at the inpatient discharge level; the denominator includes outpatient ED visits and observation stays as well as inpatient discharges. Nonetheless, we group the measure with other discharge-level outcome measures because we analyzed the outcome with the same methods. We then applied the measure's additional denominator inclusion criteria with just one minor modification (that is, we included index events in December because we had claims data for the subsequent year). We then flagged qualifying events with timely follow-up—an outpatient or carrier claim for the same patient after the index event for a non-emergency outpatient visit that constitutes appropriate follow-up (for example, a general office visit or telehealth). The follow-up visit must occur within the condition-specific time frame to be considered timely: within 7 days of the date of discharge for hypertension; within 14 days for asthma, heart failure, and coronary artery disease; and within 30 days for chronic obstructive pulmonary disease and diabetes.

B.1.6 Quality of care measures from patient experience surveys

Patients' rating of their personal doctor (score from 0 to 100)

We used the FFS CAHPS and Medicare Advantage CAHPS RIFs from the Virtual Research Data Center to construct survey respondent-level files for Maryland and the comparison group. The FFS and MA CAHPS files were linked to the Medicare beneficiary analytic files with the annual claims-based outcomes using each beneficiary's unique beneficiary identifier. We limited the CAHPS data to respondents who received a non-zero or non-missing survey weight. The file has one observation per respondent, grouped by year.

This CAHPS questionnaire asks respondents to rate their personal doctor. The rating question states: "Using any number from 0 to 10, where 0 is the worst personal doctor possible and 10 is the best personal doctor possible, what number would you use to rate your personal doctor?" Therefore, the measure that rates beneficiaries' personal doctor includes all responses to this question, and the measure of beneficiaries' primary care provider is restricted to those who answer that their personal doctor is not a specialist. We then multiplied this measure by 10 to put in on a scale of 0 to 100. Responses to this question are heavily top-coded, with means in Maryland and the comparison group above 90 in most years (Appendix A.3).

Although we limited our claims-based analyses to FFS enrollees because of data availability, we included Medicare Advantage enrollees in these analyses for several reasons: (1) having Medicare Advantage enrollees in the analysis sample reduces concerns that impact estimates could be biased because of differential enrollment in Medicare Advantage over time among beneficiaries with different health care needs and expected spending (see section A.6); (2) doctors in Maryland participating in MDPCP (one key mechanism for improving patient experience scores) are scored and incentivized (via value-based payments) based on scores from FFS and Medicare Advantage CAHPS; (3) we improve the reliability and power of these survey-based analyses by including more survey respondents in these analyses.

In processing the data, we noticed a data anomaly in the years 2015 and 2016. Specifically, survey response rates and mean ratings dropped considerably in those years for Maryland and the comparison group. We also observed that the CAHPS survey weights accounted for this drop, and when we applied the survey weights, we did not see large drops in the number of survey respondents (Appendix A.3). In testing our regressions, we did not see a material difference in the results with or without including 2015 and 2016. As such, we decided to continue to use these years in our primary regression models.

Patients' rating of their hospital

To assess patients' rating of their hospital, we used information from the Hospital CAHPS survey contained as part of the publicly available Hospital Compare database in each year from 2011 to 2020 (data only available through 2020 at the time of this report) (CMS, 2021). Reporting for the Hospital CAHPS survey was suspended in the first half of 2020 because of COVID-19. Values from 2020 files are based on surveys from the second half of 2020 only.

The public-use files are based on survey responses from patients who had an inpatient hospitalization during the year, administered to patients between 48 hours and 6 weeks after discharge from the hospital. Importantly, scores contained in the public-use Hospital CAHPS files are averages based on Medicare and non-Medicare patients. Specifically, the survey asks, "Using any number from 0 to 10, where 0 is the worst hospital possible and 10 is the best hospital possible, what number would you use to rate this hospital during your stay?" Survey responses are then averaged for each individual hospital in the year and then reported in the public-use data files at the hospital-year level. Before being publicly reported, data are adjusted for the effects of patient-mix and mode of survey administration (HCAHPS, 2022). Because the mean score was not reported in all years of our analysis, we used the percentage of patients who rated their hospital a 9 or 10 out of 10 (which is available in all years) instead.

Several major hospital systems in Maryland did not report Hospital CAHPS scores in 2016. Because of the influence these hospitals have on the Maryland mean, we chose to exclude the year 2016 from all of our analyses of this measure.

B.2. Matching and analytic files construction

B.2.1. PUMA-year-level file with variables for developing the matched comparison group

To develop the PUMA-year-level matching file, we first assigned each beneficiary to the PUMA associated with the beneficiary's mailing address zip code in each year. We then rolled-up the beneficiary-year-level demographic and enrollment file and claims-based outcomes file to the PUMA and year level—that is, one observation per PUMA per year—and calculated the mean value of each variable over all Medicare FFS beneficiaries who resided in that PUMA in that year, weighted by the number of months that each beneficiary was observable in Medicare claims in that year. In addition, we linked the CAHPS patient experience data to beneficiaries in the beneficiary-year-level files and rolled up the survey data to the PUMA-year level based on beneficiaries' assigned PUMAs. We also rolled up the discharge-year file to the PUMA and year level based on the beneficiary's home PUMA (even if the beneficiary was hospitalized outside the PUMA) and calculated the mean value of discharge-related outcomes over all discharges among beneficiaries in each PUMA. We similarly calculated matching variables or variables for checking balance between Maryland and comparison PUMAs from beneficiary-level and discharge-level claims measures, such as COVID-19-related hospitalizations and outpatient ED visits and observation stays and a PUMA-level measure of hospital market concentration and rolled these up to the PUMA-year-level.

As part of the process for constructing the matching files, we also rolled up data from other sources that we merged to the beneficiary-year-level file, including (1) American Community Survey (ACS) data for characteristics of the beneficiaries' zip codes; (2) Health Resources Services Administration data for the primary care shortage area score of each zip code; and (3) HCC scores and individual condition categories from beneficiary-year-level tables on the Virtual Research Data Center. Similarly, we also merged data from Hospital Compare and the IPPS Historical Impact Files to the discharge-year-level file by hospital that were then rolled-up to the PUMA level. Finally, we obtained survey data from the Centers for Disease Control and Prevention's Diabetes Atlas, which is derived from respondent-level data from the Behavioral Risk Factor Surveillance System. The Diabetes Atlas data are based on the Behavioral Risk Factor Surveillance System data and provide annual estimates of county-level age-adjusted obesity prevalence, diabetes incidence, and diabetes prevalence for adults older than 20 in the county. We mapped counties to their corresponding PUMAs and, for PUMAs with more than one county, we derived PUMA-level estimates using a weighted average based on county population size using the ACS data.

Table B.1 describes the variables included in the matching algorithm or in balance checks. All demographic, enrollment, and geographic variables reflect the characteristics of the PUMAs in 2013. The claims-based outcomes measures and some survey measures include variables for both levels of the outcomes in 2013 and trends in the mean yearly rate of change over the full baseline period, 2011 to 2013.

Table B.1. PUMA-level matching variables

Variables	Data source	Definition
Medicare FFS spending		
Medicare FFS Part A post-acute care spending: baseline levels	Medicare FFS claims	Mean Part A post-acute care spending (that is, for SNF and home health care covered under Part A) per beneficiary per year, calculated over all Medicare FFS beneficiaries in the PUMA in 2013
Medicare FFS Part A post-acute care spending: baseline trends	Medicare FFS claims	Mean change in Part A post-acute care spending per beneficiary per year, 2012 to 2013
Medicare FFS non-hospital spending: baseline levels	Medicare FFS claims	Mean non-hospital spending (that is, all Medicare spending for services provided outside of acute care hospitals, excluding Part D drugs) per beneficiary per year, calculated over all Medicare FFS beneficiaries in the PUMA in 2013.
Medicare FFS non-hospital spending: baseline trends	Medicare FFS claims	Mean change in non-hospital spending per beneficiary per year, 2011 to 2013
Standardized Medicare FFS spending		
Standardized hospital spending: baseline levels	Medicare FFS claims (GVDB files)	Mean standardized hospital spending per beneficiary per year, calculated over all FFS beneficiaries in the PUMA in 2013
Standardized hospital spending: baseline trends	Medicare FFS claims (GVDB files)	Mean change in standardized hospital spending per beneficiary per year, 2011 to 2013
Service use		
All-cause acute care hospitalizations: baseline levels	Medicare FFS claims	Mean number of all-cause acute care hospitalizations per 1,000 beneficiaries per year, calculated over all Medicare FFS beneficiaries in the PUMA in 2013
All-cause acute care hospitalizations: baseline trends	Medicare FFS claims	Mean change in all-cause acute care hospitalizations per 1,000 beneficiaries per year, 2011 to 2013
Outpatient ED visit and observation stays: baseline levels	Medicare FFS claims	Mean number of outpatient ED visits and observation stays per 1,000 beneficiaries per year, calculated over all Medicare FFS beneficiaries in the PUMA in 2013
Outpatient ED visits and observation stays: baseline trends	Medicare FFS claims	Mean change in outpatient ED visits and observation stays per 1,000 beneficiaries per year, 2011 to 2013
Quality		
Potentially preventable hospitalizations: baseline levels	Medicare FFS claims	Mean number of potentially preventable hospitalizations per 1,000 beneficiaries per year, calculated over all Medicare FFS beneficiaries in the PUMA in 2013
Potentially preventable hospitalizations: baseline trends	Medicare FFS claims	Mean change in the number of potentially preventable hospitalizations per 1,000 beneficiaries per year, 2011 to 2013

Variables	Data source	Definition
30-day post-discharge unplanned readmissions: baseline levels	Medicare FFS claims	Percentage of hospitalizations that met the criteria for an index stay and were followed by a 30-day all-cause unplanned readmission, calculated over all index stays for Medicare FFS beneficiaries in the PUMA in 2013
30-day post-discharge unplanned readmissions: baseline trends	Medicare FFS claims	Mean change in the percentage of index hospitalizations followed by a 30-day all-cause unplanned readmission, 2011 to 2013
Timely follow-up after a discharge for acute exacerbations of chronic conditions: baseline levels	Medicare FFS claims	Percentage of all hospitalizations or outpatient ED visits or observation stays for any of six chronic conditions (hypertension, asthma, heart failure, coronary artery disease, chronic obstructive pulmonary disease, or diabetes mellitus) that had a non-emergency outpatient follow-up visit within the relevant time frame calculated over all hospitalizations for Medicare FFS beneficiaries in the PUMA in 2013
Timely follow-up after a discharge for acute exacerbations of chronic conditions: baseline trends	Medicare FFS claims	Mean change in the percentage of hospitalizations or outpatient ED visits or observation stays for the six chronic conditions that had a non-emergency outpatient follow-up visit within the relevant time frame, 2011 to 2013
Patients' ratings of their hospital care: baseline levels	Hospital Compare	Percentage of all Medicare FFS discharges from hospitals that had a rating of 9 or 10 out of 10 in Hospital Compare, calculated over all hospitalizations for Medicare FFS beneficiaries in the PUMA in 2013
Patients' ratings of their hospital care: baseline trends	Hospital Compare	Mean change in the percentage of all Medicare FFS discharges from hospitals that had a rating of 9 or 10 out of 10 in Hospital Compare, 2011 to 2013
Beneficiaries' ratings of their PCP: baseline levels	CAHPS	Mean provider rating among all Medicare FFS and MA beneficiaries in the PUMA whose personal doctor is not a specialist and who responded to the CAHPS survey, 2013
Population health		
Obesity prevalence	CDC's Diabetes Atlas[a]	Mean age-adjusted obesity rates for all residents ages 20 and older in the PUMA.
Medicare beneficiaries' characteristics in 2013		
Age	Medicare enrollment data	Mean age of all FFS beneficiaries in the PUMA
Sex	Medicare enrollment data	Percentage of all FFS beneficiaries in the PUMA who are female
Race and ethnicity		
Black	Medicare enrollment data	Percentage of FFS beneficiaries in the PUMA who are Black
Non-Hispanic White	Medicare enrollment data	Percentage of FFS beneficiaries in the PUMA who are non-Hispanic White

Variables	Data source	Definition
Rural residence, (3 variables)	Medicare enrollment data (zip code) and Census Urban and Rural classification by ZCTA (rural)	(1) Average percentage of the population living in a rural area in the PUMA; (2) categorical variable for quartiles of the average percentage of the population living in a rural area in the PUMA
Disabled or ESRD	Medicare enrollment data	Percentage of FFS beneficiaries in the PUMA with original reason for entitlement of disability and/or ESRD
HCC score	Medicare HCC files	Mean HCC score for all FFS beneficiaries in the PUMA
Diabetes for FFS beneficiaries	Medicare HCC files	Percentage of FFS beneficiaries in the PUMA with diabetes
Diabetes among all adults	BRFSS	Percentage of adult population in the PUMA with diabetes
Population characteristics of zip code (2011–2013)		
Percentage Hispanic	ACS	Mean percentage of Hispanic residents, calculated across each Medicare beneficiary's zip code for all beneficiaries in the PUMA
Percentage below federal poverty level, adjusted for cost of living	ACS Census supplemental poverty measure	Mean percentage of residents living below the federal poverty level, calculated across each Medicare beneficiary's zip code for all beneficiaries in the PUMA, and adjusted for cost of living using the Census supplemental poverty measure
Percentage living in multi-unit structure, mobile home or group quarters	ACS	Mean percentage of residents living in multi-unit structures, mobile homes, or group quarters, calculated across each Medicare beneficiary's zip code for all beneficiaries in the PUMA
Percentage older than 65	ACS	Mean percentage of residents ages 65 years and older, calculated across each Medicare beneficiary's zip code for all beneficiaries in the PUMA
Percent younger than 18	ACS	Mean percentage of residents ages birth to 18 years, calculated across each Medicare beneficiary's zip code for all beneficiaries in the PUMA
Percentage with a high school degree or equivalent	ACS	Mean percentage of residents with a high school degree or equivalent, calculated across every Medicare beneficiary's zip code for all beneficiaries in the PUMA
Percentage who speaks English well	ACS	Mean percentage of residents who speak English well, calculated across every Medicare beneficiary's zip code for all beneficiaries in the PUMA
Percentage living in crowded home	ACS	Mean percentage of residents living in a crowded home, calculated across every Medicare beneficiary's zip code for all beneficiaries in the PUMA
Percentage without a vehicle	ACS	Mean percentage of residents without a vehicle, calculated across every Medicare beneficiary's zip code for all beneficiaries in the PUMA

Variables	Data source	Definition
Characteristics of the health care system in the PUMA in 2013		
Medicare coverage	Medicare enrollment data and ACS	Percentage of PUMA residents enrolled in Medicare
At least one acute care hospital, yes or no	Medicare FFS claims	= 1 if the PUMA has one or more acute care hospitals = 0 if the PUMA has no acute care hospital
Number of hospital beds	IPPS	A count of the number of hospital beds in the PUMA
Discharges from a major teaching hospital	Medicare FFS claims and Hospital Compare	Percentage of all hospitalizations from major teaching or very major teaching academic medical center (defined as according to a resident to bed ratio of greater than 0.25), calculated over all hospitalizations for Medicare FFS beneficiaries in the PUMA
Herfindahl-Hirschman Index	Medicare FFS claims	This variable measures the relative amount of competition in the market.[b] For each hospital in each PUMA, we calculated its market share as the percentage of discharges in the PUMA from that hospital. We then squared the market share value of each hospital and summed the squared values across all hospitals in a PUMA. Markets with higher summed values have less competition (and more market concentration) relative to markets with lower summed values (and less market concentration).
PCPs in practices (TINs) that hare small	Medicare FFS claims	Percentage of PCPs who practice in small practices in the PUMA (small is defined as 1 NPI per TIN, or a solo practice)
PCPs in practices (TINs) that are large	Medicare FFS claims	Percentage of PCPs who practice in large practices in the PUMA (large is defined as 6 or more NPIs per TIN)
Number of PCPs per 1,000 Medicare beneficiaries	Medicare FFS claims and enrollment data	For each PUMA, this is the total number of PCPs practicing in the PUMA divided by the total number of Medicare beneficiaries (FFS and MA) and multiplied by 1,000
Health professional primary care shortage area score	HRSA	Mean value of the health professional primary care shortage area score (higher indicates greater shortage) associated with the zip code of all FFS beneficiaries in the PUMA
Subgroups		
Characteristics of Black beneficiaries, baseline		
HCC score, mean	Medicare HCC files	Mean HCC score among Black beneficiaries in the PUMA in 2013
Potentially preventable admissions: baseline levels	Medicare FFS claims	Mean number of potentially preventable hospitalizations per 1,000 beneficiaries per year, calculated over all Black Medicare FFS beneficiaries in the PUMA in 2013
Potentially preventable admissions: baseline trends	Medicare FFS claims	Mean change in the number of potentially preventable hospitalizations per 1,000 beneficiaries per year, calculated over all Black Medicare FFS beneficiaries in the PUMA, 2011 to 2013

Variables	Data source	Definition
30-day post-discharge unplanned readmissions: baseline levels	Medicare FFS claims	Percentage of hospitalizations that met the criteria for an index stay and were followed by a 30-day all-cause unplanned readmission, calculated over all index stays for Black Medicare FFS beneficiaries in the PUMA in 2013
30-day post-discharge unplanned readmissions: baseline trends	Medicare FFS claims	Mean change in the percentage of index hospitalizations followed by a 30-day all-cause unplanned readmission for Black Medicare FFS beneficiaries in the PUMA, 2011 to 2013
Characteristics of Non-Hispanic White beneficiaries, baseline		
HCC score, mean	Medicare HCC files	Mean HCC score among all non-Hispanic White beneficiaries in the PUMA in 2013
Potentially preventable admissions: baseline levels	Medicare FFS claims	Mean number of potentially preventable hospitalizations per 1,000 beneficiaries per year, calculated over all non-Hispanic White Medicare FFS beneficiaries in the PUMA in 2013
Potentially preventable admissions: baseline trends	Medicare FFS claims	Mean change in the number of potentially preventable hospitalizations per 1,000 beneficiaries per year, calculated over all non-Hispanic White Medicare FFS beneficiaries in the PUMA, 2011 to 2013
30-day post-discharge unplanned readmissions: baseline levels	Medicare FFS claims	Percentage of hospitalizations that met the criteria for an index stay and were followed by a 30-day all-cause unplanned readmission, calculated over all index stays for non-Hispanic White Medicare FFS beneficiaries in the PUMA in 2013
30-day post-discharge unplanned readmissions: baseline trends	Medicare FFS claims	Mean change in the percentage of index hospitalizations followed by a 30-day all-cause unplanned readmission for non-Hispanic White Medicare FFS beneficiaries in the PUMA, 2011 to 2013
Number of potentially preventable admissions, 2013	Medicare FFS claims	Mean number of potentially preventable hospitalizations per 1,000 beneficiaries per year, calculated over all non-Hispanic White Medicare FFS beneficiaries in the PUMA in 2013
COVID-19 checks, 2019–2021		
COVID-19 inpatient admissions	Medicare FFS claims	Mean number of COVID-19 inpatient visits per 1,000 beneficiaries per year, calculated across all Medicare FFS beneficiaries in the PUMA in 2020 and 2021
COVID-19 outpatient ED visits and observation stays	Medicare FFS claims	Mean number of COVID-19 outpatient ED visits or observation stays per 1,000 beneficiaries per year, calculated across all Medicare FFS beneficiaries in the PUMA in 2020 and 2021
Excess outpatient ED visits and observation stays (2020 minus 2019)	Medicare FFS claims	The difference between the mean number of outpatient ED visits and observation stays in 2020 minus the mean number in 2019 for each PUMA

Variables	Data source	Definition
Excess ED visits and observation stays ending in an inpatient stay (2020 minus 2019)	Medicare FFS claims	The difference between the mean number of ED visits and observation stays that ended in an inpatient stay in 2020 minus the mean number in 2019 for each PUMA
Excess all-cause acute care hospital admissions (2020 minus 2019)	Medicare FFS claims	The difference between the mean number of all-cause acute care hospitalizations in 2020 minus the mean number in 2019 for each PUMA
Excess surgical hospitalizations (2020 minus 2019)	Medicare FFS claims	The difference between the mean number of surgical hospitalizations in 2020 minus the mean number in 2019 for each PUMA
Excess elective hospitalizations (2020 minus 2019)	Medicare FFS claims	The difference between the mean number of elective hospitalizations in 2020 minus the mean number in 2019 for each PUMA

[a] See this page for CDC's Diabetes Atlas: https://gis.cdc.gov/grasp/diabetes/DiabetesAtlas.html.

[b] For an overview of the Herfindahl-Hirschman Index, see this page from the U.S. Department of Justice: https://www.justice.gov/atr/herfindahl-hirschman-index. Accessed September 29, 2021.

BRFSS = Behavioral Risk Factor Surveillance System; CAHPS = Consumer Assessment of Healthcare Providers and Systems; CDC = Centers for Disease Control and Prevention; ED = emergency department; ESRD = end-stage renal disease; FFS = fee for service; HCC = Hierarchical Condition Category; HRSA = Health Resources and Services Administration; IPPS = Inpatient Prospective Payment System; MA = Medicare Advantage; NPI = National Provider Identifier; PCP = primary care physician; PUMA = Public Use Microdata Area; SNF = skilled nursing facility; TIN = Taxpayer Identification Number; ZCTA = Zip Code Tabulation Area.

B.2.2. Beneficiary-year and discharge-year level files for impact analyses

We fit regression models at the beneficiary-year or discharge-year level, as relevant, for the claims-based outcomes and for the CAHPS-based patient experience measures to assess impacts of the Maryland Model on key outcomes. The regression models use the same beneficiary-year- and discharge-year-level files described in Section B.2.1 as inputs to the PUMA-year files. Briefly, the beneficiary-year-level analytic file contains one observation PBPY for all beneficiaries ever enrolled in Medicare. For the analyses of claims-based measures, we then limited the file to those who were observable for at least one month in Medicare FFS claims data during the year (that is, they were alive, enrolled in Medicare Parts A and B FFS, and had Medicare as primary payer). For the analyses of CAHPS data, we included all respondents to the survey who were enrolled in FFS and met the criteria described above or who were enrolled in Medicare Advantage with Medicare as primary payer (by definition, Medicare Advantage enrollees are enrolled in Parts A and B). Beneficiaries can be in the file in all years of our analytic period or only one or a limited number of years, depending on their observability status.

The construction of this file involved the following steps:

1. Pulling enrollment and demographic information for the full Medicare population (that is, all beneficiaries who were ever enrolled in Medicare) during each year from the Master Beneficiary Summary File (MBSF) and Enrollment Database (EDB)
2. Identifying the first FFS observable month, if any, among the full Medicare population (many beneficiaries are never observable during the year because, for example, they are enrolled in Medicare Advantage for the full year or have employer-sponsored insurance as primary payer or only have Part A or only Part B coverage during enrolled months) for the claims-based analyses
3. Constructing variables to reflect demographic and enrollment characteristics for the year, as described in Table B.2; characteristics that could change within any year, such as dual eligibility status and residence (based on zip code); and those that are characterized based on beneficiaries' data in the first observable month of the year
4. Developing claims-based measures for all observable months for all FFS beneficiaries and merging these measures to the beneficiary-level file by unique beneficiary identifier
5. Obtaining patient survey measures for all FFS and Medicare Advantage enrollees who responded to the CAHPS in each year
6. Annualizing the claims-based measures based on the number of months observable (except for binary variables, such as "any hospitalization")
7. Merging on data from external sources (this included merging HCC scores and Master Beneficiary Summary File chronic condition categories by unique beneficiary identifier and year) and merging on characteristics of beneficiaries' PUMA from the ACS by PUMA and year (we used the ACS five-year files, which combine data for each PUMA across five years, so each PUMA will have the same values of the ACS variables across the five-year period covered)
8. Applying a final set of exclusion criteria for each year's file to exclude beneficiaries from the analytic sample if we could not map them to a location in the United States (either because they lived outside the United States or had bad zip code data).

The discharge-year-level file contains one observation per Medicare FFS discharge per year—either a discharge from an acute inpatient hospital, regardless of the reason for the hospitalization, or an outpatient discharge from the ED or observation unit with a diagnosis code for any of the six chronic conditions included in the follow-up after acute exacerbations of chronic conditions measure. Each record in this file represents a single discharge, meaning any inpatient stays that involved more than one claim were collapsed into a single record (for these stays, we retained diagnosis and procedure codes from the first and last claims in the stay). We limited outpatient ED visits and observation stays to one per day; if any outpatient ED or observation claim on the same day contained relevant diagnoses for the six chronic conditions, we included it in the discharge-level file.

We then merged demographic and enrollment characteristics, and Medicare CCW condition categories to use as covariates in the regressions onto the files by unique beneficiary identifier and year. CCW conditions included are the following original conditions: acquired hypothyroidism; acute myocardial infarction; Alzheimer's disease and related disorders or senile dementia; anemia; asthma; atrial fibrillation; benign prostatic hyperplasia; cancer – breast; cancer – colorectal; cancer – endometrial; cancer – lung; cancer – prostate; cataract; chronic kidney disease; chronic obstructive pulmonary disease and bronchiectasis; glaucoma; heart failure; hip/pelvic fracture; hyperlipidemia; hypertension; ischemic heart disease; osteoporosis; rheumatoid arthritis/osteoarthritis; and stroke/transient ischemic attack. We also included the following other chronic and potentially disabling conditions: blindness and visual impairment; cystic fibrosis and other metabolic developmental disorders; epilepsy; fibromyalgia, chronic pain and fatigue; hearing impairment; human immunodeficiency virus and/or acquired immunodeficiency syndrome (HIV/AIDS); intellectual disabilities and related conditions; leukemias and lymphomas; migraine and chronic headache; mobility impairments; muscular dystrophy; and peripheral vascular disease. We excluded the original CCW condition category for diabetes as well as original and other chronic and potentially disabling conditions related to behavioral health or drug use conditions – specifically, alcohol use disorder; anxiety; bipolar disorder; depression; schizophrenia; opioid use disorder; and tobacco use - as they are related to current or future planned outcomes. Table B.2 defines the rest of the covariates used in the impact regressions.

Table B.2. Covariates for beneficiary- and discharge-level regression models

		Included in regressions at:			
Covariate	**Definition**	**Beneficiary level: Claims measures**	**Beneficiary level: CAHPS measures**	**Discharge level: 30-day unplanned readmission**	**Discharge level: Timely follow-up**
Demographics and enrollment characteristics					
Age category	**Calculated based on the first day observable in Medicare data for the year (that is, alive, enrolled in Parts A and B FFS Medicare, with Medicare as primary payer)**				
Age less than 65 years (omitted)[a]	= 1 if age < 65 years = 0 otherwise				
Ages 65 to 69 years (reference category)	= 1 if age >= 65 years & age <= 69 years = 0 otherwise				
Ages 70 to 74 years	= 1 if age >= 70 years & age <=74 years = 0 otherwise	X	X	X	X
Ages 75 to 79 years	= 1 if age >= 75 years & age <=79 years = 0 otherwise	X	X	X	X
Ages 80 to 84 years	= 1 if age >= 80 years & age<=84 years = 0 otherwise	X	X	X	X
Ages 85 years and older	= 1 if age >= 85 years = 0 otherwise	X	X	X	X
Sex					
Male (reference category)	= 1 if male or unknown sex = 0 if female				
Female	= 1 if female = 0 if male or unknown sex	X	X	X	X
Race and ethnicity[b]					
White (reference category)	= 1 if RTI race variable = 1 = 0 if RTI race variable not equal to 1				
Black	= 1 if RTI race variable = 2 = 0 if RTI race variable not equal to 2	X	X	X	X

Covariate	Definition	Included in regressions at: Beneficiary level: Claims measures	Beneficiary level: CAHPS measures	Discharge level: 30-day unplanned readmission	Discharge level: Timely follow-up
Hispanic	= 1 if RTI race variable = 5 = 0 if RTI race variable not equal to 5	X	X	X	X
Other minorities	= 1 if RTI race variable = {0,3,4,6} = 0 if RTI race variable = {1,2,5}	X	X	X	X
OREC and age interaction					
Not disabled or ESRD (reference category)	= 1 if OREC = disabled, ESRD, or disabled and ESRD = 0 otherwise				
Age < 65 and disabled or ESRD	= 1 if OREC = disabled, ESRD, or disabled and ESRD and age<65 years = 0 if age>65 years or age <65 and OREC = aged	X	X	X	X
Age >=65 and ESRD	= 1 if OREC = ESRD or disabled and ESRD and age >=65 years = 0 if OREC = aged or disabled or if age < 65 years	X	X	X	X
Ages 65 to 69 years and disabled	= 1 if OREC = disabled and age >= 65 and age < 70 = 0 if OREC = aged, ESRD, or disabled and ESRD or age < 65 or age >= 70	X	X	X	X
Ages 70 to 74 years and disabled	= 1 if OREC = disabled and age >= 70 and age < 75 = 0 if OREC = aged, ESRD, or disabled and ESRD or age < 70 or age >= 75	X	X	X	X
Age 75 to 79 years and disabled	= 1 if OREC = disabled and age >= 75 and age < 80 = 0 if OREC = aged, ESRD, or disabled and ESRD or age < 75 or age >= 80	X	X	X	X
Age 80 to 84 years and disabled	= 1 if OREC = disabled and age >= 80 and age < 85 = 0 if OREC = aged, ESRD, or disabled and ESRD or age < 80 or age >= 85	X	X	X	X
Age 85 years and older and disabled	= 1 if OREC = disabled and age >= 85 = 0 if OREC = aged, ESRD, or disabled and ESRD or age < 85	X	X	X	X
Social Vulnerability Index	CDC Social Vulnerability Index merged onto individual beneficiaries at the Census tract level. Overall ranking of Census tract from 15 social factors, including poverty, lack of vehicle access, and crowded housing	X	X	X	X

		Included in regressions at:			
Covariate	Definition	Beneficiary level: Claims measures	Beneficiary level: CAHPS measures	Discharge level: 30-day unplanned readmission	Discharge level: Timely follow-up
Rural residence	= 1 if more than 50 percent of residents in that zip code are living in rural areas, per Census Urban and Rural classification by ZCTA = 0 if 50 percent or fewer residents in that zip code are living in rural areas, per Census Urban and Rural classification by ZCTA	X	X		
Education	4-category variable for whether the beneficiary reported their highest level of educational attainment as: 1) less than high school; 2) high school (reference category); 3) some college; or 4) college degree or higher		X		
CCW-related variables[c]					
CCW Condition flags	= 1 if the beneficiary had claims-based evidence of the condition in the relevant look-back period = 0 if the beneficiary had no claims-based evidence of the condition in the relevant look-back period = missing if the beneficiary was not observable in the relevant look-back period	X		X	X
Observable for CCW one-year look-back period[d]	= 1 if the beneficiary was observable in the one-year look-back period used for multiple CCW condition categories = 0 if the beneficiary was not observable in the one-year look-back period used for multiple CCW condition categories	X		X	X
Observable for CCW two-year look-back period[e]	= 1 if the beneficiary was observable in the two-year look-back period used for multiple CCW condition categories = 0 if the beneficiary was not observable in the two-year look-back period used for multiple CCW condition categories	X		X	X

		Included in regressions at:			
Covariate	Definition	Beneficiary level: Claims measures	Beneficiary level: CAHPS measures	Discharge level: 30-day unplanned readmission	Discharge level: Timely follow-up
Observable for CCW three-year look-back period[f]	= 1 if the beneficiary was observable in the three-year look-back period used for the CCW Alzheimer's and related disorders and senile dementia condition category = 0 if the beneficiary was not observable in the three-year look-back period used for the CCW Alzheimer's and related disorders and senile dementia condition category	X		X	X
New enrollee flag	= 1 if the beneficiary was not observable in all months of the prior year = 0 if the beneficiary was observable in all months of the prior year	X		X	X
Three or more physical health conditions	=1 if the number of individual CCW condition categories related to physical health is greater than or equal to 3, excluding hypertension and hyperlipidemia =0 if the number of individual HCC condition categories related to physical health is 0, 1, or 2, excluding hypertension and hyperlipidemia = missing if the beneficiary was not observable in all months of the prior year	X		X	X

Covariate	Definition	Included in regressions at: Beneficiary level: Claims measures	Beneficiary level: CAHPS measures	Discharge level: 30-day unplanned readmission	Discharge level: Timely follow-up
Other measure of health status					
Self-reported health status	3-category variable for whether the beneficiary self-reported their health status as 1) poor/fair; 2) good (reference category) or 3) very good or excellent		X		
Episode of care-related variables					
Chronic condition category for acute exacerbation follow-up[g]	= Asthma or hypertension or coronary artery disease or heart failure or chronic obstructive pulmonary disease or diabetes if the principal diagnosis code is sufficient to diagnose the condition or if the principal diagnosis code is related to the condition and a secondary diagnosis on the claim is sufficient, per IMPAQ technical specifications				X
Index admission category for 30-day unplanned readmission[h]	=Surgical, cardiorespiratory, cardiovascular, neurology, or medicine based on the procedure codes and principal diagnosis, per CMS/Yale technical specifications			X	
COVID-19 variables[i]					
COVID-19 inpatient stay	Two variables, one measured for 2020 and another separately for 2021, but both defined as: =1 if beneficiary had at least one inpatient admission with a diagnosis of COVID-19 =0 if beneficiary had no inpatient admissions with a diagnosis of COVID-19	X		X	X
COVID-19 ED or observation visit	Two variables, one measured for 2020 and another separately for 2021, but both defined as: =1 if beneficiary had at least one ED visit or observation stay with a diagnosis of COVID-19 =0 if beneficiary had no ED visits or observation stays with a diagnosis of COVID-19	X		X	X

[a] Age less than 65 years was collinear with Age < 65 and OREC = disabled or ESRD.

[b] We combined other minorities into a single category for regression due to the small number of beneficiaries who meet this definition in Maryland.

[c] See text in Section B.2.2 for a list of conditions included.

[d] The following original CCW condition categories have a one-year look-back period: acquired hypothyroidism; acute myocardial infarction; anemia; asthma; atrial fibrillation; benign prostatic hyperplasia; cancer – breast; cancer – colorectal; cancer – endometrial; cancer – lung; cancer – prostate; cataract; chronic obstructive pulmonary disease and bronchiectasis; glaucoma; hip/pelvic fracture; hyperlipidemia; hypertension; osteoporosis; and stroke/transient ischemic attack.

[e] The following original CCW condition categories have a two-year look-back period: chronic kidney disease; heart failure; ischemic heart disease; rheumatoid arthritis/osteoarthritis. In addition, all of the other chronic and potentially disabling conditions have a two-year look-back period, including: blindness and visual impairment; cystic fibrosis and other metabolic developmental disorders; epilepsy; fibromyalgia, chronic pain and fatigue; hearing impairment; human immunodeficiency virus and/or acquired immunodeficiency syndrome (HIV/AIDS); intellectual disabilities and related conditions; leukemias and lymphomas; migraine and chronic headache; mobility impairments; muscular dystrophy; and peripheral vascular disease.

[f] The CCW condition category for Alzheimer's disease and related disorders or senile dementia has a three-year look-back period.

[g] See the IMPAQ Health (2018) specifications for details on the chronic condition category assignment

[h] See the specifications for 30-day unplanned readmission developed by the Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (2020) for details on index admission category assignment

[i] The COVID-19 variables were included in sensitivity analyses and not the main results. See Appendix C.

CCW = Chronic Conditions Warehouse; CAHPS = Consumer Assessment of Healthcare Providers and Systems; ED = emergency department; HCC = Hierarchical Condition Category; MSBF = Medicare Beneficiary Summary File; OREC = Original Reason for Entitlement Code; RTI = Research Triangle Institute; TIA = transient ischemic attack; VRDC = Virtual Research Data Center; ZCTA = Zip Code Tabulation Area.

B.2.3. Hospital-year level files for impact analyses of patients' rating of their hospital

We fit regression models at the hospital-year level for the HCAHPS-based patient experience measure to assess impacts of the Maryland Model on patients' overall ratings of their hospital stays (see Appendix A.4. for more details on regression specifications). The hospital-year-level analytic file is constructed from publicly available data obtained from CMS's Hospital Compare website; it contains one observation per hospital per year (2011–2020) for all Medicare-certified hospitals, identified by their CMS Certification number (CCN) (CMS 2021). To control for differences in case-mix over time between Maryland and comparison group hospitals, we also merged on hospital case mix index from CMS's IPPS public use files in each year by CCN (CMS n.d.; CMS 2022b). Hospital case mix index is calculated from the Medicare Severity Diagnosis Related Groups (MS-DRG) weight for each stay, which reflects the mean severity of all stays at the hospital during the year.

To ensure we had an appropriate analytic sample on which to estimate impacts, we made several restrictions to the full list of hospitals in Hospital Compare for our primary regression models, including the following:

- We dropped hospitals without CAHPS scores in a given year. Hospitals might appear in the data without a score because they are exempted (that is, not subject to the IPPS) or because they failed to reach the minimum number of survey responses to avoid suppression in the data.
- We dropped hospitals not located in a PUMA that is part of our comparison group (that is, received zero matching weight) or that did not have any address information (street or ZIP code). Though the data are at the hospital-year level, we continue to weight hospitals in the rest of the nation using our PUMA-level matching weight to ensure we are using the same comparison group as we are for other outcomes. This means we had to assign hospitals to PUMAs using address information available in Hospital Compare. We used GIS mapping software to identify addresses and place hospitals into PUMAs.
- We dropped hospitals that did not have any Medicare FFS discharges in 2013 or, if not in our data in 2013, the year after they first appear in Hospital Compare. Our final regression weight for patient hospital ratings multiplies the PUMA matching weight by a weight representing the size of the hospital, measured by total number of discharges in 2013. Hospitals that do not have any Medicare FFS discharges in claims could not be assigned a weight. We fixed the hospital size weight based on hospital discharges in 2013 (the year before the MDAPM period began) because the Maryland Model might impact the number of discharges in years after the intervention began. For hospitals not in our data in 2013 (for example, new hospitals in later years), we used the number of discharges in the year after they first appear as their size weight in all years.
- We dropped hospitals with missing information for case mix index in all years that appear in our analysis.[29] In our primary models, we want to control for patient case mix because it is possible that the Maryland Model and its incentives to move care out of the hospital could change the case mix in Maryland hospitals relative to hospitals in the comparison group. A consequence of this restriction is that all Critical Access Hospitals (CAHs) drop out of our primary sample because they are not subject to the IPPS and therefore do not have case mix index information in any year. There are no official CAHs in Maryland, but several hospitals, particularly in rural areas, function similarly to how CAHs

[29] A small number of hospitals had case-mix index information for some years but not others. We chose to impute the missing case mix index in these cases using the mean for that hospital in the years we had valid observations. We then included a missing indicator flag for missing case mix index in our regressions.

do in the rest of the nation. We tested models that retained CAHs in our regressions by dropping case mix index and found our results to be very consistent with our main findings.

- We dropped hospitals that only appeared in a single year in our data (after making the restrictions above). Our primary models chose not to include hospital-level fixed effects in an effort to avoid over-controlling for potential effects of the Maryland Model on hospital closure (see Section A.4. for more details). Still, hospitals that only appear once are likely data anomalies (for example, a hospital converting to a different type or newly merging with another hospital). We don't believe these single data-point hospitals meaningfully contribute to our analysis, and they could introduce bias if mismeasured differentially in Maryland and the comparison group.
- Finally, in Maryland, we restrict observations in our analysis to the list of hospitals that are part of the 2013 or 2019 State Agreements (CMS 2018). For this report, after applying the criteria above, this did not remove any additional hospitals from our analytic sample but may come into play in future years.

The final hospital analytic panel contains 753 unique hospitals over 10 years that meet the above criteria.

B.3. Comparison of measures in this report to those in the state agreement and SIHIS

The 14 outcomes in this report partially align with those in the legal agreement that established the MD TCOC model (Table B.3) and SIHIS (Table B.4). This report does not contain some of the state agreement or SIHS-related measures because of data limitations or because they were out of the scope of the evaluation. Further, although some of the state agreement and SIHIS measures focus on all Maryland residents, the measures in this report are—with one exception—for Medicare beneficiaries. The one exception is patients' ratings for hospitals, which is measured among all patients.

Although we aligned outcome measures in this report with the state agreement and SIHIS when feasible and appropriate, we did not aim to align methods for estimating effects for these measures. The state agreement and SIHIS set their own methods for assessing progress toward the stated goals, which typically do not rely on a matched comparison group. By contrast, all the impact estimates in this report use the same method for estimating impacts—a difference-in-differences model with a matched comparison group drawn from geographic areas (PUMAs) outside Maryland.

This report also includes seven outcome measures for Medicare FFS beneficiaries that are not explicit goals in either the state agreement or in SIHIS. We included these measures in this report because the model's incentives and supports could logically lead to improvements in them or because the model could have unintended consequences, worsening these outcomes. The outcome measures are the following:

1. All-cause acute care hospital admissions
2. Outpatient ED visits and observations of stays
3. Intensity of hospital care (measured by standardized hospital spending)
4. Non-hospital spending
5. Post-acute care spending
6. Patients' rating of their personal doctor
7. Patients' rating of their hospital

Table B.3. Alignment between outcome measures in the state agreement and this report

Category	Measure	Commitment in the state agreement	Similar measure in this report's estimates of model impacts
Spending	All-payer hospital spending per Maryland resident	Limit growth to no more than 3.58 percent per year (the long-term growth rate of the state economy)	Medicare FFS spending for hospital care per Maryland Medicare beneficiary per year
	Annual Medicare FFS savings	Meet specific annual savings targets—for example, $300 million in 2023 (assessed by comparing actual Medicare spending in Maryland with what spending would be if Maryland's 2013 spending grew at the national rate)	Total Medicare FFS spending per year (with and without non-claims payment)
Quality of care	Medicare readmissions rate	Must at least maintain improvements achieved during MDAPM	Medicare 30-day post-discharge unplanned readmission rate
	All-payer reductions in hospital-acquired conditions	Must at least maintain improvements achieved during MDAPM	None

Source: CMS 2018.

FFS = fee for service; MDAPM = Maryland All-Payer Model

Table B.4. Alignment between outcome measures and populations in SIHIS with the measures and populations in this report

SIHIS outcome measures, baseline performance, and targets			Impact estimates in this report	
Outcome measure (population)	**2018 baseline**	**2026 final target**	**Related outcome measure**	**Population for the impact evaluation**
SIHIS domain: Hospital quality				
Avoidable Admissions: Risk-Adjusted PQI-90 Rates (all payer)	1,335 admits per 100,000	25% improvement	Potentially preventable admissions (using the PQI-90 composite)	Medicare FFS
Readmission disparities (all payer)	Hospital-specific risk difference across levels of Patient Adversity Index	Half of eligible hospitals achieving 50% improvement in disparity	30-day unplanned readmission rates	Medicare FFS
SIHIS domain: Care transformation across the system[a]				
Timely Follow-up After Acute Exacerbations of Chronic Conditions (Medicare FFS)	71.36%	75.00% (5.1% improvement)	Timely follow-up after acute exacerbations of chronic conditions	Medicare FFS
SIHIS domain: Total population health				
Mean BMI in the population of adult Maryland residents (all residents)	28.13 kg/m2	Achieve more favorable change from baseline compared with control	Use of Diabetes Prevention Program services [an intermediate goal in SIHIS]	Medicare FFS
Overdose mortality (all residents)	Age-adjusted death rate of 37.2/100,000	Achieve more favorable change from baseline compared to control	None	n.a.
SMM (all payer)	243.1 SMM rate per 10,000 delivery hospitalizations	197.1 SMM rate per 10,000 delivery hospitalizations	None	n.a.
Asthma-related ED visit rate for ages 2-17 (all payer)	9.2 visits per 1,000	5.3 per 1,000	None	n.a.

Source: HSCRC 2022b.

[a] SIHIS has one other measure in this domain: "Increase the amount of Medicare Total Cost of Care or number of Medicare beneficiaries under Care Transformation Initiatives, Care Redesign Program, or successor payment model." This is a process measure to show how the model is being implemented over time—and the reach of alternative payment approaches within Maryland. But because it's not a quality or efficiency outcome for individual people, we are not estimating impacts on this measure.

BMI = body mass index; ED = emergency department; FFS = fee for service; n.a. = not applicable; PQI = Prevention Quality Indicators; SIHIS = Statewide Integrated Health Improvement Strategy; SMM = Severe Maternal Morbidity.

Appendix C: Accounting for COVID-19 in the impact estimates

Here, we describe how COVID-19 could bias our 2020 and 2021 impact estimates and the actions we took to help mitigate identified risks.

COVID-19 could introduce bias in our results if the pandemic affected outcomes in Maryland and our selected comparison group differently in ways not related to the Maryland Model. The bias could also occur if Medicare beneficiaries, including those who do not get COVID-19, respond differently to the pandemic (for example, if beneficiaries in Maryland are more or less likely to avoid hospital care because of COVID-19). On the other hand, it's possible that the Maryland Model had a true effect on our key outcomes by affecting COVID-19 related outcomes. For example, MDPCP might have helped practices learn about COVID-19 early, or the model might have allowed hospitals more flexibility and financial security under global budgets that improved access for Maryland beneficiaries (Perman et al. 2021). Indeed, the percent of the population with at least one dose of the COVID-19 vaccine as of 12/31/2021 was higher in Maryland (80.6%) than it was nationally (73.5%) according to CDC (CDC 2022). We do not want to adjust away any true effects the model might have on the rate of COVID-19 or COVID-19-related outcomes in Maryland. For this reason, our primary regression models *do not* include controls for COVID-19 hospitalizations and ED visits.

C.1. Mitigating COVID-19 bias risk

To mitigate the risk of bias in our 2020 and 2021 estimates from COVID-19, we took a multipronged approach, including accounting for social vulnerability in matching and through regression specifications and sensitivity analyses.

C.1.1 Accounting for social vulnerability in matching and checking balance on COVID-19 outcomes

We chose not to include COVID-19 variables in our matching to avoid matching on future outcomes the model might have the ability to affect. We did, however, include in our matching several of the individual components from the Centers for Disease Control and Prevention's Social Vulnerability Index, with the idea that we want Maryland and our comparison groups to have similar levels of vulnerability to disease outbreaks, including COVID-19. Specifically, we matched on the following variables (defined at the PUMA level): the percentage of the population living in multi-unit structures, mobile homes, or group quarters; the percentage older than 64; the percentage younger than 18; the percentage with a high school degree (or equivalent); the percentage that speaks English well; the percentage living in a crowded home; and the percentage without a vehicle. Together with other matching variables, we captured most components of the Social Vulnerability Index that enabled us to find a comparison group with a similar level of social vulnerability as Maryland (in 2011–2013).

We also checked balance (without including it directly in our matching algorithm) on 2020 and 2021 COVID-19 measures in Maryland and our selected comparison group (Table C.1). We found the following:

- Rates of ED and observation visits for COVID-19 in Maryland and the selected comparison group were broadly similar in 2020 and very similar in 2021, with a weighted difference of 0.14 fewer visits per 1,000 people in Maryland than in the comparison group in 2021.
- Rates of COVID-19 hospitalizations were similar between groups in 2021, with a weighted difference of 1.14 fewer stays per 1,000 people in Maryland than in the comparison group. In terms of

standardized differences, the number of COVID-19 hospitalizations in 2020 was lower in Maryland than in our selected comparison group. But the size of this difference was small relative to all hospital admissions (a difference of about five hospitalizations per 1,000 people was about 1.6 percent of total inpatient hospitalizations in Maryland in 2013).

- The declines from 2019 to 2020 in hospitalizations (all-cause, elective, and surgical) and outpatient ED visits were similar between Maryland and the selected comparison group. This indicates that the large declines in service use that occurred early in the COVID-19 pandemic occurred in similar amounts in Maryland and the comparison group.

Table C.1. Balance on COVID-19 and COVID-19-related variables

Variable description	Maryland pre-weighted mean	Comparison group pre-weighted mean	Difference pre-weighting	Standardized difference pre-weighting	Included in matching?	Maryland post-weighted mean	Comparison group post-weighted mean	Difference post-weighting	Standardized difference post-weighting
Variables defined during the matching period (2013)									
Percentage living in multi-unit structure, mobile home, or group quarters	19.04	25.60	-6.56	-0.48	Yes	19.04	23.55	-4.52	-0.33
Percentage older than age 64	13.86	14.97	-1.11	-0.26	Yes	13.86	14.34	-0.48	-0.11
Percentage younger than age 18	22.52	22.84	-0.32	-0.09	Yes	22.52	23.21	-0.69	-0.20
Percentage with high school degree (or equivalent)	92.06	89.65	2.41	0.44	Yes	92.06	91.16	0.90	0.17
Percentage that speaks English well	97.57	96.43	1.14	0.25	Yes	97.57	96.88	0.69	0.15
Percentage living in crowded home	6.45	9.00	-2.56	-0.47	Yes	6.45	8.04	-1.59	-0.29
Percentage without a vehicle	7.03	6.34	0.69	0.09	Yes	7.03	7.36	-0.33	-0.04
Variables defined after the matching period (2019–2021) and not used in matching algorithm									
Number of COVID-19 hospitalizations (2020)	13.66	16.50	-2.84	-0.34	No	13.66	18.68	-5.02	-0.59
Number of COVID-19 hospitalizations (2021)	13.85	15.63	-1.78	NA	No	12.29	13.43	-1.14	NA
Number of COVID-19 outpatient ED visits and observation stays (2020)	6.40	9.07	-2.67	-0.51	No	6.40	7.56	-1.16	-0.22
Number of COVID-19 outpatient ED visits and observation stays (2021)	10.30	12.30	-2.00	NA	No	10.42	10.56	0.14	NA
Excess number of all-cause acute care hospital admissions (2020 minus 2019)[a]	-39.69	-39.95	0.26	0.02	No	-39.69	-40.49	0.80	0.06
Excess number of outpatient ED visits and observation stays (2020 minus 2019)[a]	-121.10	-114.80	-6.31	-0.19	No	-121.10	-117.19	-3.91	-0.12
Excess number of ED visits and observation stays ending in inpatient stay (2020 minus 2019)[a]	-27.08	-24.16	-2.92	-0.23	No	-27.08	-25.36	-1.72	-0.14
Number of surgical hospitalizations (2020 minus 2019)[a]	-14.75	-14.10	-0.65	-0.14	No	-14.75	-14.22	-0.52	-0.12
Number of elective hospitalizations (2020 minus 2019)[a]	-11.46	-11.73	0.27	0.05	No	-11.46	-11.31	-0.15	-0.03

[a] Difference in the 2020 rate per 1,000 Medicare fee-for-service beneficiaries and the 2019 rate per 1,000 Medicare fee-for-service beneficiaries.

ED = emergency department; NA = not available.

C.1.2. Regression-based approaches to account for COVID-19

In addition to matching, we implemented a few regression-based mitigation strategies related to COVID-19 in our main regression specification and through sensitivity analyses.

- First, we designed our regression models to estimate the combined effect of the Maryland Model from 2019 through 2021, as well as the individual yearly effects separately. Doing so allows us to interpret the effect of the model separately in its first three years. If we see large differences between yearly estimates that we think are unlikely to be related to changes made to the model, we likely would interpret those differences as attributable, at least in part, to the direct or indirect effects of COVID-19. As shown in the tables in Section C.2, the impact estimates were similar in 2021, 2020, and 2019 for most outcomes.
- Second, each of our regression models explicitly control for the Social Vulnerability Index measure noted above (as defined in 2011–2013). The measure itself represents a percentile ranking of vulnerability (which is different from the individual components we included in matching) and is designed to further control for differences between Maryland and the comparison group on social vulnerability.
- Third, we conducted a sensitivity analysis in which we included a flag for COVID-19 hospitalizations and ED visits in our regression models. If we believe that COVID-19 is largely exogenous (that is, not influenced by the Maryland model), these estimates will control for differences between Maryland and the comparison group that we should otherwise not be attributing to the model.
- Finally, we conducted a second sensitivity test that adjusted health condition controls measured during COVID-19. We observed a decline in prevalence in Maryland and the comparison group for several conditions measured during the pandemic that are unlikely to be all due to true declines.[30] The CCW health conditions we are using indicate that a person has a condition if they have claims for those conditions, typically in the past 12 or 24 months depending on the specific conditions. Early in the COVID-19 pandemic (especially spring 2020), beneficiaries received substantially less care (inpatient and outpatient) than they did before the pandemic and in later pandemic periods. This lower use of care means that, all else equal, beneficiaries will have fewer claims in 2020 and therefore a lower likelihood that underlying health condition will be detected in claims. Our main results for 2021 use CCW flags based on claims from 2019 and 2020, but to test the sensitivity of the results to potential mismeasurement of conditions in 2021, we reestimated impacts, setting a beneficiary's conditions in 2021 to be based on those from 2020 (measured using claims in 2018 and 2019, avoiding the early pandemic period). Results from the two sensitivity analyses are in the section that follows.

[30] For example, in Maryland we observe the prevalence of chronic obstructive pulmonary disease in our sample to be about 9.5 percent in both 2019 and 2020 (based on claims from 2017–2019). But in 2021 (based on 2019–2020 claims), the prevalence of chronic obstructive pulmonary disease in Maryland drops to 8.1 percent.

C.2. Results from COVID-19 sensitivity analyses

C.2.1. Controlling for COVID-19 hospitalizations and ED visits

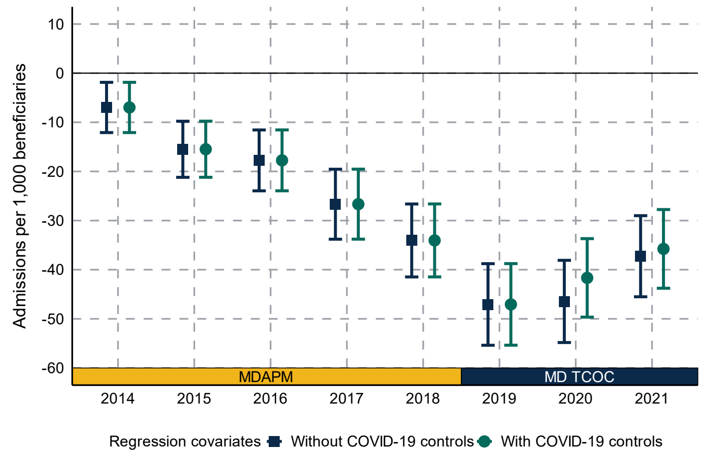
We ran sensitivity analyses for our each of our key outcomes that include as covariates COVID-19 hospitalizations and ED visits in 2020 and 2021 to control for differences in the rate of these outcomes between Maryland and our comparison group (selected outcomes, Figures C.1 to C.3, and Table C.2). Specifically, we add as a control variable for each yearly observation whether a beneficiary had a hospital visit (inpatient or outpatient ED) with a COVID-19 diagnosis that year. In general, controlling for COVID-19 outcomes led to impact estimates that were closer to zero in both 2020 and 2021, especially for spending outcomes. In all cases, though the impact estimates were smaller, qualitative conclusions (including statistically significant findings) did not change.

Several recent articles have argued (Haft et al. 2020; Peterson and Schumacher 2020) that the Maryland Model—including MDPCP and hospital global budgets—might have decreased the rates and severity of COVID-19 in the state and improved care for patients with COVID-19. These articles suggest that controlling for COVID-19 rates is inappropriate because it could control away effects of the program. Because of this, we believe the main regression specification—which does not control for COVID-19 hospital visits—is the most appropriate. Nonetheless, our main results and conclusions are not sensitive to adding the COVID-19 controls in 2020 and 2021.

Key results and selected figures comparing regression models with and without COVID-19 controls

- Overall, key findings on utilization were robust to sensitivity analyses that controlled for COVID-19 admissions and ED visits.
  - Impacts in 2021 were very similar with and without COVID-19 controls.
  - In 2020, effects on hospital admissions were attenuated modestly toward zero (-42 per 1,000 beneficiaries with COVID-19 controls versus -46 per 1,000 beneficiaries without) (Figure C.1 and Table C.2).
- Impacts on spending followed a similar pattern with results that were very similar with and without COVID-19, particularly in 2021.
  - Similar to admissions, reductions in hospital (-$334 PBPY with COVID controls versus -$443 PBPY without) and total spending were modestly attenuated toward zero in the first year of the COVID-19 pandemic.
- Each of the outcomes related to quality and population health showed consistent impacts in 2019 through 2021, and none were sensitive to the inclusion of COVID-19 hospital admissions and ED visits control variables (Table C.2, showing potentially preventable admissions as an example).

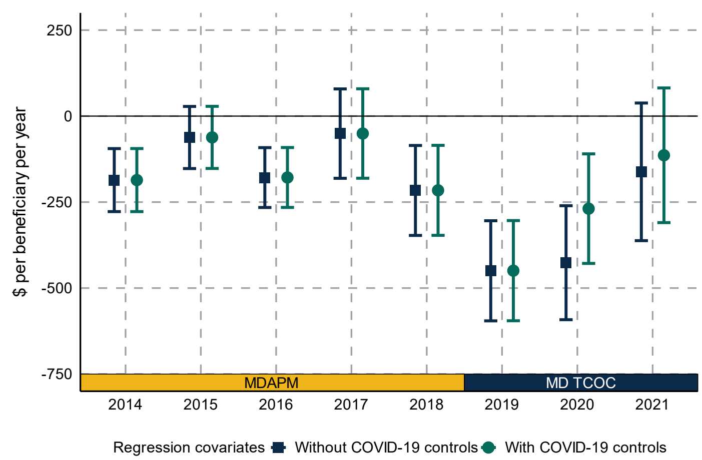
Figure C.1. Impact of the Maryland Model on all-cause acute care admissions with and without COVID-19 controls, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval.

Figure C.2. Impact of the Maryland Model on total Medicare FFS spending with and without COVID-19 controls, by year



Note: Errors bars are 90% CIs for the yearly impact estimates. Estimates in which the intervals do not span zero are statistically different from zero at a $p < 0.10$ threshold.

CI = confidence interval; FFS = fee for service.

Table C.2. Impacts of the Maryland Model on selected outcomes, controlling for COVID-19 hospitalizations and ED visits

	Difference-in-differences impact estimates: main estimate		Difference-in-differences impact estimates: controlling for COVID-19 hospitalizations and ED visits	
	Estimate (90% CI)	% Impact	Estimate (90% CI)	% Impact[a]
Total Medicare FFS spending ($ per beneficiary per year)				
Early MDAPM period				
2014	-$186*** (-$278; -$94)	-1.5%	-$186*** (-$278; -$94)	-1.5%
2015	-$62 (-$152; $28)	-0.5%	-$62 (-$152; $28)	-0.5%
2016	-$179*** (-$266; -$92)	-1.4%	-$178*** (-$266; -$91)	-1.4%
Later MDAPM period				
2017	-$51 (-$181; $79)	-0.4%	-$51 (-$181; $79)	-0.4%
2018	-$216*** (-$347; -$85)	-1.6%	-$216*** (-$347; -$85)	-1.6%
Combined (2017-2018)	-$134* (-$258; -$9)	-1.0%	-$134* (-$258; -$9)	-1.0%
MD TCOC period				
2019	-$450*** (-$595; -$304)	-3.2%	-$449*** (-$595; -$304)	-3.2%
2020	-$426*** (-$592; -$261)	-3.2%	-$269*** (-$428; -$110)	-2.0%
2021	-$162 (-$362; $38)	-1.1%	-$114 (-$310; $82)	-0.8%
Combined (2019-2021)	-$348*** (-$504; -$192)	-2.5%	-$279*** (-$431; -$127)	-2.0%
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-$316*** (-$394; -$238)	-2.2pp	-$316*** (-$394; -$237)	-2.2pp
2020	-$292*** (-$404; -$180)	-2.2pp	-$135** (-$237; -$34)	-1.0pp
2021	-$28 (-$179; $122)	-0.1pp	$20 (-$128; $167)	0.2pp
Combined (2019-2021)	-$214*** (-$306; -$121)	-1.5pp	-$145*** (-$233; -$58)	-1.0pp
Hospital spending ($ per beneficiary per year)				
Early MDAPM period				
2014	-$197*** (-$268; -$125)	-2.8%	-$197*** (-$268; -$125)	-2.8%
2015	-$161*** (-$230; -$92)	-2.3%	-$161*** (-$230; -$92)	-2.3%
2016	-$349*** (-$417; -$280)	-4.9%	-$349*** (-$417; -$280)	-4.9%

	Difference-in-differences impact estimates: main estimate		Difference-in-differences impact estimates: controlling for COVID-19 hospitalizations and ED visits	
	Estimate (90% CI)	% Impact	Estimate (90% CI)	% Impact[a]
Later MDAPM period				
2017	-$256*** (-$355; -$158)	-3.5%	-$256*** (-$355; -$157)	-3.5%
2018	-$424*** (-$521; -$328)	-5.6%	-$424*** (-$521; -$327)	-5.6%
Combined (2017-2018)	-$341*** (-$434; -$248)	-4.6%	-$341*** (-$434; -$247)	-4.6%
MD TCOC period				
2019	-$580*** (-$693; -$467)	-7.4%	-$580*** (-$693; -$467)	-7.4%
2020	-$443*** (-$577; -$309)	-5.9%	-$334*** (-$460; -$208)	-4.5%
2021	-$507*** (-$660; -$354)	-6.3%	-$471*** (-$615; -$326)	-5.9%
Combined (2019-2021)	-$510*** (-$633; -$387)	-6.6%	-$461*** (-$579; -$343)	-6.0%
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-$240*** (-$299; -$180)	-2.8pp	-$240*** (-$299; -$180)	-2.8pp
2020	-$102* (-$192; -$13)	-1.3pp	$7 (-$73; $87)	0.1pp
2021	-$166** (-$278; -$54)	-1.7pp	-$130** (-$235; -$25)	-1.3pp
Combined (2019-2021)	-$169*** (-$241; -$97)	-2.0pp	-$121*** (-$187; -$54)	-1.4pp
All-cause admissions (number per 1,000 beneficiaries per year)				
Early MDAPM period				
2014	-7** (-12; -2)	-2.3%	-7** (-12; -2)	-2.3%
2015	-15*** (-21; -10)	-4.9%	-15*** (-21; -10)	-4.9%
2016	-18*** (-24; -12)	-6.0%	-18*** (-24; -12)	-6.0%
Later MDAPM period				
2017	-27*** (-34; -20)	-8.9%	-27*** (-34; -20)	-8.9%
2018	-34*** (-41; -27)	-11.4%	-34*** (-41; -27)	-11.4%
Combined (2017-2018)	-30*** (-37; -23)	-10.0%	-30*** (-37; -23)	-10.0%
MD TCOC period				
2019	-47 (-55; -39)	-15.7%	-47 (-55; -39)	-15.7%
2020	-46 (-55; -38)	-17.7%	-42 (-50; -34)	-16.4%
2021	-37*** (-46; -29)	-14.3%	-36*** (-44; -28)	-14.0%
Combined (2019-2021)	-44 (-52; -35)	-16.1%	-42 (-50; -34)	-15.5%

	Difference-in-differences impact estimates: main estimate		Difference-in-differences impact estimates: controlling for COVID-19 hospitalizations and ED visits	
	Estimate (90% CI)	% Impact	Estimate (90% CI)	% Impact[a]
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-17*** (-20; -13)	-5.7pp	-17*** (-20; -13)	-5.7pp
2020	-16*** (-21; -12)	-7.7pp	-11*** (-15; -7)	-6.4pp
2021	-7** (-11; -2)	-4.3pp	-5** (-10; -1)	-4.0pp
Combined (2019-2021)	-13*** (-17; -9)	-6.1pp	-11*** (-15; -7)	-5.5pp
Potentially preventable admissions (number per 1,000 beneficiaries per year)				
Early MDAPM period				
2014	0.1 (-1.1; 1.3)	0.2%	0.1 (-1.1; 1.3)	0.2%
2015	0.1 (-1.3; 1.5)	0.2%	0.1 (-1.3; 1.5)	0.2%
2016	-0.7 (-2.2; 0.9)	-1.4%	-0.7 (-2.2; 0.9)	-1.4%
Later MDAPM period				
2017	-3.5*** (-5.3; -1.7)	-6.6%	-3.5*** (-5.3; -1.7)	-6.6%
2018	-6.1*** (-8.7; -3.6)	-11.8%	-6.1*** (-8.7; -3.6)	-11.8%
Combined (2017-2018)	-4.8*** (-6.9; -2.7)	-9.2%	-4.8*** (-6.9; -2.7)	-9.2%
MD TCOC period				
2019	-7.4*** (-9.8; -5.0)	-14.3%	-7.4*** (-9.8; -5.0)	-14.3%
2020	-6.7*** (-9.0; -4.3)	-16.8%	-6.3*** (-8.7; -4.0)	-15.9%
2021	-6.8*** (-9.4; -4.3)	-17.4%	-6.7*** (-9.3; -4.2)	-17.2%
Combined (2019-2021)	-7.0*** (-9.3; -4.6)	-16.1%	-6.8*** (-9.2; -4.5)	-15.7%
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-2.6*** (-3.5; -1.7)	-5.1pp	-2.6*** (-3.5; -1.7)	-5.1pp
2020	-1.8** (-3.1; -0.6)	-7.6pp	-1.5** (-2.7; -0.3)	-6.7pp
2021	-2.0*** (-3.2; -0.8)	-8.2pp	-1.9** (-3.1; -0.7)	-8.0pp
Combined (2019-2021)	-2.1*** (-3.1; -1.2)	-6.9pp	-2.0*** (-3.0; -1.0)	-6.5pp

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$

[a] The percentage is calculated as the impact estimate for the year divided by the estimate of the counterfactual for the year. We estimated the counterfactual as the mean outcome observed that year in Maryland minus the difference-in-differences impact estimate for the year.

CI = confidence interval; pp= percentage point

C.2.2. Adjusting 2021 chronic condition flags to account for underreported diagnoses because of the COVID-19 pandemic

To test the sensitivity of our results to different assumptions about health condition measurement during the COVID-19 pandemic, we estimated impacts for a set of key outcomes in models that replace 2021 health condition values with values from 2020 (which were based on claims before the pandemic). We do this to smooth the prevalence over time under the assumption that most of the conditions we control for are designed to represent a chronic disease concept with little variation year to year.

We present results that use 2020 health conditions in 2021 for four key outcomes in Table C.3. Similar to the models in which we removed health conditions entirely, results were largely qualitatively consistent with our main impact findings in 2021 (and virtually identical in other years, as expected) and, in general, less different from our main impact results than when we remove conditions altogether. Reductions in hospital spending were moderately larger, and increases in non-hospital spending were moderately smaller, leading to larger reductions in total spending in 2021 that were statistically significant. Reductions in all-cause hospital admissions were largely similar to our main impact findings.

Although this test shows the results are not overly sensitive to how we measure conditions during the pandemic, it has its own limitations—which is why we prefer our main models over this test. First, any beneficiaries not in our analytic sample in 2020 do not have a value for health conditions to be replaced, which adds to the number of beneficiaries we assign a missing value for health conditions to. A second limitation of using 2020 health condition values in 2021 is that it discards information about beneficiaries who *are* correctly assigned the condition based on diagnosis codes in 2020—beneficiaries who might be among the most medically complex because they visited their health care provider despite pandemic risks. Finally, as controls in our regression models, by making this change, we've altered what it means to have a health condition in one specific year, 2021, but not the others. Under this approach, in 2021, condition flags represent the association between outcomes and having a condition at least two years prior, rather than last year, as is the case in all other years. We interact health conditions with year in our models so we can estimate these effects separately, but it still represents a conceptual disconnect from earlier years.

Table C.3. Impacts in 2021 adjusting 2021 chronic conditions to account for potential mismeasurement

	Difference-in-differences impact estimates: main estimate		Difference-in-differences impact estimates: replacing 2021 flags with 2020	
	Estimate (90% CI)	% Impact	Estimate (90% CI)	% Impact
Total Medicare FFS spending ($ per beneficiary per year)				
Early MDAPM period				
2014	-$186*** (-$278; -$94)	-1.5%	-$186*** (-$278; -$95)	-1.5%
2015	-$62 (-$152; $28)	-0.5%	-$62 (-$153; $28)	-0.5%
2016	-$179*** (-$266; -$92)	-1.4%	-$179*** (-$266; -$92)	-1.4%
Later MDAPM period				
2017	-$51 (-$181; $79)	-0.4%	-$51 (-$181; $79)	-0.4%
2018	-$216*** (-$347; -$85)	-1.6%	-$217*** (-$347; -$86)	-1.6%
Combined (2017-2018)	-$134* (-$258; -$9)	-1.0%	-$134* (-$259; -$10)	-1.0%
MD TCOC period				
2019	-$450*** (-$595; -$304)	-3.2%	-$450*** (-$596; -$305)	-3.2%
2020	-$426*** (-$592; -$261)	-3.2%	-$427*** (-$592; -$261)	-3.2%
2021	-$162 (-$362; $38)	-1.1%	-$316*** (-$504; -$128)	-2.2%
Combined (2019-2021)	-$348*** (-$504; -$192)	-2.5%	-$398*** (-$552; -$245)	-2.9%
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-$316*** (-$394; -$238)	-2.2pp	-$316*** (-$394; -$238)	-2.2pp
2020	-$292*** (-$404; -$180)	-2.2pp	-$293*** (-$405; -$181)	-2.2pp
2021	-$28 (-$179; $122)	-0.1pp	-$182** (-$316; -$47)	-1.2pp
Combined (2019-2021)	-$214*** (-$306; -$121)	-1.5pp	-$264*** (-$353; -$175)	-1.9pp
Hospital spending ($ per beneficiary per year)				
Early MDAPM period				
2014	-$197*** (-$268; -$125)	-2.8%	-$197*** (-$268; -$125)	-2.8%
2015	-$161*** (-$230; -$92)	-2.3%	-$161*** (-$231; -$92)	-2.3%
2016	-$349*** (-$417; -$280)	-4.9%	-$349*** (-$418; -$281)	-4.9%

	Difference-in-differences impact estimates: main estimate		Difference-in-differences impact estimates: replacing 2021 flags with 2020	
	Estimate (90% CI)	% Impact	Estimate (90% CI)	% Impact
Later MDAPM period				
2017	-$256*** (-$355; -$158)	-3.5%	-$257*** (-$355; -$158)	-3.5%
2018	-$424*** (-$521; -$328)	-5.6%	-$424*** (-$521; -$328)	-5.6%
Combined (2017-2018)	-$341*** (-$434; -$248)	-4.6%	-$341*** (-$434; -$248)	-4.6%
MD TCOC period				
2019	-$580*** (-$693; -$467)	-7.4%	-$581*** (-$693; -$468)	-7.4%
2020	-$443*** (-$577; -$309)	-5.9%	-$443*** (-$578; -$309)	-5.9%
2021	-$507*** (-$660; -$354)	-6.3%	-$583*** (-$732; -$435)	-7.2%
Combined (2019-2021)	-$510*** (-$633; -$387)	-6.6%	-$535*** (-$658; -$413)	-6.9%
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-$240*** (-$299; -$180)	-2.8pp	-$240*** (-$299; -$180)	-2.8pp
2020	-$102* (-$192; -$13)	-1.3pp	-$102* (-$192; -$13)	-1.3pp
2021	-$166** (-$278; -$54)	-1.7pp	-$243*** (-$348; -$137)	-2.6pp
Combined (2019-2021)	-$169*** (-$241; -$97)	-2.0pp	-$194*** (-$265; -$124)	-2.3pp
All-cause admissions (number per 1,000 beneficiaries per year)				
Early MDAPM period				
2014	-7** (-12; -2)	-2.3%	-7** (-12; -2)	-2.3%
2015	-15*** (-21; -10)	-4.9%	-15*** (-21; -10)	-4.9%
2016	-18*** (-24; -12)	-6.0%	-18*** (-24; -12)	-6.0%
Later MDAPM period				
2017	-27*** (-34; -20)	-8.9%	-27*** (-34; -20)	-8.9%
2018	-34*** (-41; -27)	-11.4%	-34*** (-41; -27)	-11.4%
Combined (2017-2018)	-30*** (-37; -23)	-10.0%	-30*** (-37; -23)	-10.0%
MD TCOC period				
2019	-47 (-55; -39)	-15.7%	-47*** (-55; -39)	-15.7%
2020	-46 (-55; -38)	-17.7%	-46*** (-55; -38)	-17.7%
2021	-37*** (-46; -29)	-14.3%	-40*** (-49; -32)	-15.3%
Combined (2019-2021)	-44 (-52; -35)	-16.1%	-45*** (-53; -36)	-16.4%

	Difference-in-differences impact estimates: main estimate		Difference-in-differences impact estimates: replacing 2021 flags with 2020	
	Estimate (90% CI)	% Impact	Estimate (90% CI)	% Impact
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-17*** (-20; -13)	-5.7pp	-17*** (-20; -13)	-5.7pp
2020	-16*** (-21; -12)	-7.7pp	-16*** (-21; -12)	-7.7pp
2021	-7** (-11; -2)	-4.3pp	-10*** (-15; -5)	-5.3pp
Combined (2019-2021)	-13*** (-17; -9)	-6.1pp	-14*** (-18; -10)	-6.4pp
Non-hospital spending ($ per beneficiary per year)				
Early MDAPM period				
2014	$11 (-$27; $48)	0.2%	$11 (-$27; $48)	0.2%
2015	$99*** ($58; $140)	1.8%	$99*** ($58; $140)	1.8%
2016	$170*** ($129; $211)	3.2%	$170*** ($129; $211)	3.2%
Later MDAPM period				
2017	$206*** ($148; $264)	3.8%	$205*** ($147; $263)	3.7%
2018	$208*** ($146; $270)	3.6%	$208*** ($146; $270)	3.6%
Combined (2017-2018)	$207*** ($148; $265)	3.7%	$207*** ($148; $265)	3.7%
MD TCOC period				
2019	$131*** ($61; $200)	2.2%	$130*** ($61; $200)	2.1%
2020	$17 (-$55; $88)	0.3%	$17 (-$55; $88)	0.3%
2021	$345*** ($262; $428)	5.5%	$268*** ($190; $345)	4.2%
Combined (2019-2021)	$162*** ($95; $229)	2.7%	$137*** ($70; $203)	2.3%
Difference in estimates during MD TCOC period and later MDAPM period				
2019	-$76*** (-$115; -$37)	-1.5pp	-$76*** (-$115; -$38)	-1.6pp
2020	-$190*** (-$237; -$143)	-3.4pp	-$190*** (-$237; -$143)	-3.4pp
2021	$138*** ($72; $204)	1.8pp	$61* ($2; $120)	0.5pp
Combined (2019-2021)	-$45* (-$84; -$5)	-1.0pp	-$70*** (-$109; -$31)	-1.4pp

* $p < 0.10$; ** $p < 0.05$, *** $p < 0.01$

[a] The percentage is calculated as the impact estimate for the year divided by the estimate of the counterfactual for the year. We estimated the counterfactual as the mean outcome observed that year in Maryland minus the difference-in-differences impact estimate for the year.

CI = confidence interval; pp= percentage point

Mathematica Inc.

Princeton, NJ • Ann Arbor, MI • Cambridge, MA
Chicago, IL • Oakland, CA • Seattle, WA
Woodlawn, MD • Washington, DC

Mathematica
Progress Together

mathematica.org

EDI Global, a Mathematica Company

Operating in Tanzania, Uganda, Kenya, and the United Kingdom