Abstract

Background: The goal of this study was to synthesize evaluation results across the portfolio of CMS Innovation Center models to inform future model development. We examined 21 Medicare models and demonstrations with at least two years of impact estimates that cover interventions operating between 2012 and 2020.

Methods: In this synthesis, we looked at measures that were available across multiple model evaluations. These included gross and net Medicare spending and measures of utilization (e.g., inpatient admissions, emergency department visits, post-acute care, inpatient readmissions) and quality of care (e.g., self-reported satisfaction with care, mortality). Results were summarized across models to gain a broader understanding of themes by interventions, care settings, provider types, and beneficiary target populations. Many of the results presented are either final estimates or represent at least four years of performance, which is generally sufficient time to start drawing conclusions. However, results for a few models represent only two to three years of performance and should be considered preliminary; conclusions may change over time.

Results: Across the 21 Medicare models, more than half (fourteen) demonstrated gross savings to Medicare. Changes in spending were driven by improvements in inpatient admissions (ten models) and/or post-acute care (fourteen models). For models that paid financial incentives to participants, six had net savings, six incurred net losses, and six models had no notable impacts on net spending. Beneficiary or caregiver self-reported experience of care remained relatively unchanged by the majority (nine) of models, improved in two models, and had unfavorable results (which were small in magnitude) for one model. Mortality remained unchanged for most (eight) models, suggesting that these interventions did not cause any harm, and there were notable improvements in mortality in four models. Models that focus on reducing acute or specialty care or that targeted specific populations[1] (e.g., terminal illness, lower extremity joint replacements) were more likely to show gross savings and generally had larger, more favorable impacts on utilization relative to models focused on primary care and population management[2] which generally serve broader, healthier populations.

Conclusions: It is possible that the higher baseline spending of sicker beneficiaries, the inclusion of institutional and specialty care providers, and the more narrowly focused target populations in the Acute or Specialty Care and Targeted Population models provided more room to reduce spending. Primary Care and Population Management models served large panels of relatively healthy, mostly low-cost Medicare beneficiaries and focused on preventing disease and improving care coordination. Longer time windows for investments in care coordination, staffing, clinical workflow redesign, health information technology, and data analytics, as well as greater engagement of primary and specialty care providers, may be needed to reduce spending in Primary Care and Population Management models. Even with successful evaluation results and transformation efforts, models may face other barriers to national expansion. Generous financial incentive payments, which helped ensure robust participation in models, made it difficult for many models to demonstrate net savings. Voluntary models also were encumbered by participants exiting the model prior to Medicare being able to realize returns or savings.

---

[1] **Acute or Specialty Care and Targeted Population models**: Bundled Payments for Care Improvement Initiative; Bundled Payments for Care Improvement Advanced; Comprehensive End Stage Renal Disease Care (ESRD) Model; Comprehensive Joint Replacement Model; Home Health Value-Based Purchasing Model; Maryland All-Payer Model; Medicare Care Choices Model; Oncology Care Model; and Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport.

[2] **Primary Care and Population Management models:** Accountable Care Organization (ACO) Investment Model; Advance Payment ACO Model; Comprehensive Primary Care Initiative; Comprehensive Primary Care Plus; Financial Alignment Initiative for Medicare-Medicaid Enrollees, Washington; Independence at Home Demonstration; Medicare Advantage Value-Based Insurance Design Model; Million Hearts®: Cardiovascular Disease Risk Reduction Model; Next Generation ACO Model; Part D Enhanced Medication Therapy Management Model; Pioneer ACO Model; and Vermont All-Payer ACO Model.

Table of Contents

Purpose

The objective of this study was to synthesize evaluation results across 21 CMS Innovation Center models. Evaluation results for similar outcome measures were compiled for current and former Innovation Center models. Similarities were identified across interventions by focusing on the primary mechanism of action through which the model provided care. Themes in findings are provided among models with similarities in their participant types, interventions used, and type of beneficiaries or health care providers touched by the model.

Introduction

In 2010, the Centers for Medicare & Medicaid Services' (CMS') Innovation Center was created by section 1115A of the Social Security Act (as added by section 3021 of the Affordable Care Act) to test new methods for health care service delivery that would reduce program expenditures, while preserving or enhancing the quality of care furnished to individuals. The goal was to move the health care system away from fee-for-service (FFS), which incentivizes quantity of care, towards value-based care that rewards better outcomes, better health and lower costs. The Secretary of the Department of Health and Human Services has the authority to expand the duration and scope of the testing of successful models, including implementation on a nationwide basis, through rulemaking.[3] Ten years later, we've synthesized the evaluation results from many of the Innovation Center's Medicare models and demonstrations (here forth simply referred to as "models") in order to identify themes which can guide setting the future course of value-based care.

In 2021, the CMS Innovation Center released a strategic refresh with a renewed vision to help build a health system that achieves equitable outcomes through high quality, affordable, person-centered care[4]. This will require designing a more harmonized and streamlined portfolio of models that can drive system transformation. To inform this effort, the Research and Rapid Cycle Evaluation Group (RREG) within the CMS Innovation Center undertook an in-depth review of select CMS Innovation Center models. The goal of this synthesis was to review the most recently available evaluation results for Medicare-focused models, the largest share of the CMS Innovation Center's portfolio. Medicaid-focused models were not included in this analysis as they serve a different beneficiary population and themes from evaluation results of models focused on Medicaid may differ from those related to Medicare models.

We examined 21 Medicare models with at least two years of impact estimates beyond their respective implementation periods (Table 1).[5] The 21 models cover interventions that operated from calendar years 2012-2020. For this reason, the performance years (PY) reported for each model may represent different calendar years. These estimates also represent different maturity levels with some models having final results for all performance years available while other models' evaluations cover the first two years of performance.

---

[3] To exercise this authority, the Secretary and CMS actuaries must review the CMS evaluations and determine that a model must either reduce spending without reducing the quality of care, or improve the quality of care without increasing spending, and must not deny or limit the coverage or provision of any benefits. To date, the following certifications for potential model expansion have been prepared: Certification of Home Health Value-Based Purchasing, Certification of the Medicare Prior Authorization Model for Repetitive Scheduled Non-Emergent Ambulance Transport, Medicare Diabetes Prevention Program, Pioneer Accountable Care Organization (ACO) Model. https://www.cms.gov/Research-Statistics-Data-and-Systems/Research/ActuarialStudies/CMMI-Model-Certifications

[4] CMS Innovation Center 2021 Strategy Refresh, Strategic Direction: https://innovation.cms.gov/strategic-direction

[5] This analysis was conducted in 2021 using the most recent data available at that time. Estimates were updated for some models using data available in early 2022 or footnotes were added to indicate whether findings in the paper were similar or different from more recently available evaluation reports released in early 2022.

Table 1. 21 Medicare Models & Demonstrations, Evaluation Data Sources

CMS Innovation Center Model or CMS Demonstration	Data Source	Performance/ Periods Years covered
Accountable Care Organization (ACO) Investment Model (AIM)	Final Evaluation Report	1-3
Advance Payment (AP) ACO Model	Final Evaluation Report	1-3
Bundled Payments for Care Improvement (BPCI) Initiative	Final Evaluation Report	1-5
Bundled Payments for Care Improvement Advanced (BPCI-A) Model	Third Evaluation Report	1-2
Comprehensive ESRD Care (CEC) Model	Fifth Annual Evaluation Report	1-5
Comprehensive Joint Replacement (CJR) Model	Fourth Annual Report	1-4
Comprehensive Primary Care (CPC) Initiative	Final Evaluation Report	1-4
Comprehensive Primary Care+ (CPC+)	Fourth Annual Report	1-4
Financial Alignment Initiative for Medicare-Medicaid Enrollees (FAI) Demonstration, Washington[6]	Fifth Annual Report	1-6
Home Health Value-Based Purchasing (HHVBP) Model	Fifth Evaluation Report	1-5
Independence at Home (IAH) Demonstration	Year Five Evaluation Report [7]	1-5
Maryland (MD) All-Payer Model	Final Evaluation Report	1-4
Medicare Advantage (MA) Value-Based Insurance Design (VBID) Model	Report First Three Years	1-3
Medicare Care Choices Model (MCCM)	Fourth Annual Report	1-4
Million Hearts® (MH): Cardiovascular Disease Risk Reduction Model	Fourth Annual Report	1-4
Next Generation ACO (NGACO) Model	Fourth Evaluation Report	1-4
Oncology Care Model (OCM)	Performance Periods 1-5 [8]	1-5
Part D Enhanced Medication Therapy Management (Part D Enhanced MTM) Model	Third Evaluation Report	1-3[9]
Pioneer ACO Model	Final Evaluation Report	Final
Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport (RSNAT)	Final Evaluation Report	1-5
Vermont (VT) All-Payer ACO Model	First Evaluation Report	1-2

In this synthesis, we looked at similar measures that were available across each of the models. These included gross and net Medicare spending, measures of utilization (e.g. inpatient admissions, emergency department (ED) visits, post-acute care (PAC), inpatient readmissions), and quality of

---

[6] FAI involves **t**wo demonstrations that have used a Managed FFS model and 11 that use a capitated model. FAI Washington has the most demonstration years of data available at this time and that, coupled with its successful results, led to its inclusion in this paper. A high-level summary of evaluation results across the other FAI models is provided in Appendix B.

[7] Evaluation results for IAH are available through the sixth year when two fewer practices participated and results for outcome measured examined in this paper were non-significant (Sixth Evaluation Report).

[8] Evaluation results for OCM for spending through 2020 (sixth performance period) are available in a more recent evaluation report (Payment Impacts Evaluation Report: Periods 1-6), although that report does not include the other utilization measures examined in this synthesis. Results for spending outcomes are largely the same in the more recent report as are presented in this paper.

[9] Evaluation results for Part D Enhanced MTM through the fourth performance year are available in a more recent evaluation report (fourth evaluation report). The fourth evaluation report for Part D Enhanced MTM assessed the impact of the model on Low Income Subsidy and Medically Complex Subgroups. These subgroups were more likely to have been targeted for Enhanced MTM services but there were no significant expenditure impacts for these subgroups.

care (e.g. beneficiary or caregiver self-reported satisfaction with care, mortality). Results were summarized across models to gain a broader understanding of themes and impacts by care settings, health care provider types, and beneficiary target populations. The synthesis is also helping to identify where new approaches to measurement may be needed to more fully understand the CMS Innovation Center's impacts on system-wide transformation.

Methods

Each of the models had their own unique payment and care delivery transformation efforts. However, similarities were identified across interventions by focusing on the primary mechanism of action through which the model provided care. Themes across models fell within two broad categories: 1) **Acute or Specialty Care & Targeted Populations (N=9)** and 2) **Primary Care and Population Management (N=12)**. These two groupings had similarities across the types of model participants and health care providers involved, intervention used, and the type of beneficiaries that were touched by the intervention. Additional information related to the analyses used for this paper are available in Appendix A. All estimates discussed throughout the paper were statistically significant at the $p<0.05$ level unless otherwise noted.

Results

Acute or Specialty Care & Targeted Populations

Model Grouping Characteristics

The grouping of Acute or Specialty Care & Targeted Populations includes nine Medicare models. Seven models are voluntary: BPCI, BPCI-A, CEC, MD All-Payer, MCCM, OCM, and RSNAT[10]. Two are mandatory: HHVBP and CJR. Model participants in this grouping included hospitals, post-acute care facilities, specialty care providers (e.g. oncology, nephrology), and targeted populations such as those served in home health and hospice agencies. These models used interventions that focused on reducing or preventing unnecessary or avoidable care through acute-care episodes and/or management of specific diseases and conditions through specialty care. The payment models in this grouping are provided in the table below as well as described further in Appendix Table 1.

Acute or Specialty Care & Targeted Populations Models	Payment
Bundled Payments for Care Improvement (BPCI) Initiative	30, 60, or 90-day episode following a hospitalization or initiation of post-acute care with a discount to target price tied to quality
Bundled Payments for Care Improvement Advanced (BPCI-A) Model	90-day episode-based bundle (discount to prospective target price) tied to quality
Comprehensive ESRD Care (CEC) Model[11]	ACO with one and two-sided risk tied to quality

---

[10] While RSNAT was a voluntary model, ambulance suppliers in selected states must participate if they bill RSNAT codes by either going through prior authorization or prepayment review.

[11] The accountable care relationships in CEC incentivize care coordination and population management, similar to other ACO models in the Primary Care and Population Management grouping of models. However, the participants and health care providers involved in CEC include dialysis facilities and nephrologists, which is more similar to models in the Acute or Specialty Care & Targeted Populations grouping that include specialty care providers. The high-cost nature of beneficiaries in this model as well as the specific

Comprehensive Joint Replacement (CJR) Model	Episode-based bundle (discount to prospective target price) tied to quality
Home Health Value-Based Purchasing (HHVBP) Model	Budget-neutral quality-based bonus payments
Maryland (MD) All-Payer Model[12]	CMS waiver exempted Maryland hospitals from inpatient and outpatient prospective payment systems and shifted state's hospital payment structure to an all-payer annual global budget using rate-setting
Medicare Care Choices Model (MCCM)	Monthly care management payments
Oncology Care Model (OCM)	Prospective care management and performance-based incentive payments
Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport (RSNAT)	No financial incentives from CMS to participate in this model; Prior authorization or prepayment review used to decrease Medicare expenditures without affecting beneficiaries' access to or quality of care

These models served beneficiaries with acute-care episodes, those needing specialist care, targeted populations, and/or chronically ill patients. Put another way, these models served beneficiaries who, without proper care management, could need to use institutional care or care that would be otherwise potentially avoidable, increasing Medicare costs. The average baseline total Medicare expenditures for beneficiaries served by the majority of these models[13] tended to range between $15,000 to $30,000, with some exceptions. Beneficiaries served by the MD All-Payer Model encompassed all inpatient admissions in the state, which included a wider range of health statuses; therefore, the model had average baseline expenditures around $12,000. Beneficiaries served by three models had higher baseline total Medicare expenditures, including some episodes in BPCI and BPCI-A that ranged up to $62,000 and $47,000, respectively, HHVBP ($50,307), and CEC ($78,480). Beneficiaries in this grouping were considered sicker based on average baseline Hierarchical Condition Category[14] (HCC) scores around 2.0[15] (with some exceptions[16]) relative to the general Medicare FFS population that are considered healthier with an average score of 1.0. These models ranged in size from as small as 89 hospices serving around 6,500 beneficiaries in MCCM to as large as 2 million episodes in HHVBP and BPCI. Appendix Table 1 provides details for each of the nine models in this grouping, including the years covered by the impact estimates contained within in this

---

targeted population served (e.g., patients with end-stage renal disease) is also more similar to models in this grouping, most of whom have high baseline Medicare expenditures and have a specific condition or disease. These characteristics fit more closely with themes in the results for the Acute or Specialty Care and Targeted Populations model grouping, which included successful results for cost, utilization, and quality, likely driven by the model's participants, health care providers, and patient population.

[12] The MD All-Payer Model was included in this grouping as it focused on hospital services. The Care Redesign Program, which began towards the end of the model's performance period, provided hospitals with the opportunity to partner with specialists, primary care physicians, and other community-based providers to improve care coordination outside of the hospital.

[13] These include BPCI, BPCI-A, CJR, MCCM, and RSNAT.

[14] HCC scores are based on health conditions and diagnoses in health care claims that are summarized into an algorithm to calculate the patient's health complexity and assign a risk factor to the beneficiary that can be used to predict future patient costs. Higher scores indicate the beneficiaries has more health conditions and more costly health care utilization relative to beneficiaries with lower scores.

[15] Baseline HCC scores were at least 2.0 for HHVBP, MCCM, RSNAT, OCM. HCC scores summarized across BPCI and BPCI-A episodes were not available.

[16] Baseline HCC scores were lower for beneficiaries served by CJR (ranging from 1.31 to 1.6), MD All-Payer Model (ranging between 0.99 to 1.14), and CEC (ranging from 1.05 to 1.20). Baseline HCC scores were higher than 2.0 for MCCM beneficiaries at enrollment (mean HCC 5.6).

paper, the intervention, payment model, health care providers involved, beneficiaries touched, and relative size of the model.

Cross-group Results for Acute or Specialty Care & Targeted Populations

Overview

All nine models in the Acute or Specialty Care & Targeted Populations grouping had significant reductions on gross Medicare spending, driven in large part by significant reductions in inpatient admissions and post-acute care (PAC) spending or utilization. Only two models[17] reduced emergency department (ED) visits, but about half of the models[18] reduced inpatient readmissions. All but two models[19] assessed patient experience of care relative to a comparison group; five of the models[20] had no significant changes, one model[21] had significant improvements, one model[22] had unfavorable results. Three[23] of the eight models that examined mortality had significant reductions and the remaining[24] had non-significant changes.

Many of the results presented here represent final evaluation results[25] or at least four performance years[26]. Results for BPCI-A represent only two performance years and caution should be used when interpreting these findings as results may change as future evaluation results are available, particularly given the changes to the model design after these performance years[27]. Impact estimates across the nine models are reviewed below for each measure examined in the cross-model synthesis.

Spending & Utilization

Table 2 below summarizes impact estimates across these nine Medicare models that addressed changes in health care delivery within the Acute or Specialty Care & Targeted Populations grouping.

- All nine of the models demonstrated significant reductions in gross spending ranging from as low as -1% of the baseline mean in Oncology Care Model (OCM), -1.3% in Comprehensive ESRD Care (CEC), and -1.6% for Home Health Value-Based Purchasing (HHVBP) to moderate levels in Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport (RSNAT -2.4%), Maryland (MD) All-Payer Model (-2.8%), Bundled Payments for Care Improvement (BPCI) Model 2 (-3.1% and -4.9%), BPCI-A[28] (-2.1% Hospital medial episodes; -2.0 Physician Group Practice [PGP] medical episodes; -4.1 Hospital surgical episodes; -4.7% PGP surgical episodes), and

---

[17] MD All-Payer Model and RSNAT
[18] BPCI-A surgical episodes, CEC, CJR, MCCM
[19] MD All-Payer Model and RSNAT
[20] BPCI-A, CEC, CJR, HHVBP, OCM
[21] MCCM
[22] BPCI
[23] CEC, HHVBP, MD-All-Payer
[24] BPCI, BPCI-A, CJR, OCM, RSNAT
[25] BPCI, CEC, MD All-Payer, RSNAT
[26] CJR, HHVBP, MCCM, OCM
[27] CMS made significant design changes to BPCI-A starting in 2021 (fourth performance year), in order to meet statutory requirements, that intends to improve the model's target pricing based on evidence that target prices were too high for medical episodes but were more accurate for surgical episodes.
[28] BPCI-A results pooled across all episodes found gross reductions in expenditures of -2.7% (-2.2% for medical episodes, and -4.5% for surgical episodes).

Comprehensive Joint Replacement (CJR -5.2%), to higher levels in BPCI Model 3 (-7.5% and -7.6%)[29] and Medicare Care Choices Model (MCCM -17%).

- After accounting for model related incentive payments for the six models where this was relevant[30], only two models demonstrated significant decreases in net spending, BPCI-A (PGP surgical episodes -3.5%; Hospital surgical episodes -3.9%) and MCCM (-14.0%). BPCI (1.3% for Model 2 and 3.1% for Model 3) and BPCI-A (Hospital medical episodes 2.8%; PGP medical episodes 1.3%) demonstrated significant increases in net spending.

Underlying these changes in spending were significant changes in utilization.

- Five of six models[31] had significant reductions in inpatient admissions ranging from as low as -1.6% in HHVBP and -3% in CEC, to as high as -7.2% in MD All-Payer Model and -26% in MCCM.

- Most models in this grouping didn't significantly impact emergency department (ED) visits, with the exception of significant reductions in MCCM (-14%) and moderately significant increases in BPCI Model 3 Skilled Nursing Facilities (SNF) (1.0% at p<0.1 level).

- While there were significant reductions in inpatient admissions (-3%) and ED visits (-2.6%) observed in RSNAT, we do not expect that the model caused these findings.

Five of eight models[32] significantly reduced inpatient readmissions including BPCI Model 3 Home Health Agencies (HHAs) (-2%), CEC (-2%), CJR (-3.5% at p<0.1), BPCI-A PGP Surgical episodes (-5.4% at p<0.1 level), and MCCM (-28%).

---

[29] BPCI Model 4 had non-significant increases in gross spending of 2.3%.
[30] RSNAT and MD All-Payer Model do not have model-related incentive payments and therefore only report gross spending results. Net results for the fifth performance period of CEC were not available at the time the report was created although they were not statistically significant in the fourth performance period.
[31] Inpatient admissions are not a relevant measure for BPCI, BPCI-A, and CJR models as episodes may be triggered by inpatient care.
[32] The RSNAT evaluation did not include readmissions as an outcome measure.

Table 2. *Spending & Utilization for Acute or Specialty Care & Targeted Population Models*

Model	Spending		Utilization		
	Gross	Net	Inpatient admissions	ED Visits	Inpatient readmissions
BPCI Model 2 Hospitals	**-3.1%**	**1.3%**	--	0.0%	0.0%
BPCI Model 2 PGPs	**-4.9%**		--	0.0%	0.2%
BPCI Model 3 SNFs	**-7.6%**	**3.1%**	--	1.0%§	0.6%
BPCI Model 3 HHAs	**-7.5%**		--	-0.1%	**-2.0%**
BPCI Model 4 Hospitals[33]	2.3%		--	0.8%	-1.2%
BPCI-A Hospitals - Medical	**-2.1%**	**2.8%**	--	--	-0.2%
BPCI-A Hospitals - Surgical	**-4.1%**	**-3.9%**	--	--	-4.3%
BPCI-A PGP - Medical	**-2.0%**	**1.3%**	--	--	0.6%
BPCI-A PGP - Surgical	**-4.7%**	**-3.5%**	--	--	-5.4%§
CEC	**-1.3%**	--[34]	**-3.0%**	-0.2%	**-2.0%**
CJR * [35]	**-5.2%**	-1.8%	--	1.0%	-3.5%§
HHVBP *	**-1.6%**	--	**-1.6%**	0.5%	--
MD All-Payer Model	**-2.8%**	--	**-7.2%**	1.5%	-0.3%
MCCM	**-17.0%**	**-14.0%**	**-26.0%**	**-14.0%**	**-28.0%**
OCM	**-1.0%**	**2.0%**[36]	0.6%	0.0%	-1.2%
RSNAT	**-2.4%**	--	**-3.0%**	**-2.6%**	--

Table Key: Bolded estimates are statistically significant at least at the p<0.05 level. " § " indicates an estimate that was statistically significant at the p<0.10 level. Results that were in a favorable direction and statistically significant have a light green cell shading. Results that were unfavorable and statistically significant have red/orange cell shading. Cells with gray shading illustrate non-significant results. "--" indicates the measure was not relevant or available. " * " indicates the model included a randomized design. BPCI=Bundled Payments for Care Improvement; BPCI-A=Bundled Payments for Care Improvement Advanced; CEC=Comprehensive End-Stage Renal Disease Care; CJR= Comprehensive Joint Replacement; HHVBP=Home Health Value-Based Purchasing; MD= Maryland; MCCM= Medical Care Choices Model; OCM=Oncology Care Model; PGP=Physician Group Practices; RSNAT= Prior Authorization of Repetitive, Scheduled Non-Emergency Ambulance Transport.

[33] Net results were not available as the gross results were not statistically significant.

[34] Net results for CEC were not available at the time the final report was created. Net losses in the fourth performance period were not statistically significant.

[35] Results for CJR were reported for the mandatory hospitals only, and do not include results for the voluntary hospitals.

[36] The losses as a percentage of baseline average total episode payment (2.04%), including monthly enhanced oncology services and performance-based payments, was calculated by the independent evaluation contractor and was not included in the evaluation report. As of the fifth performance period, the model resulted in net losses of $377.1 million when accounting for the model's financial incentives (Payment Impacts Evaluation Report: Periods 1-6).

Post-Acute Care

Table 3 summarizes post-acute care (PAC) impact estimates for the nine models that addressed changes in health care delivery within the Acute or Specialty Care & Targeted Populations grouping.

- Most models (eight of nine) showed significant reductions in PAC utilization or spending in various settings, including institutional PAC (skilled nursing facilities [SNF], institutional rehabilitation facility [IRF]) or any PAC, and/or use of home health (HH).

- The Bundled Payments for Care Improvement (BPCI) Initiative had significant improvements in post-acute care utilization:
  - Reductions in any post-acute care utilization:
    - Model 2 Physician Group Practices [PGP] -5.6%;
    - Model 3 Skilled Nursing Facilities [SNF] -3.5%;
  - Reductions in institutional post-acute care utilization:
    - Model 2 Hospitals -3.6%;
    - Model 2 PGP -1.8% (at p<0.1 level)
  - Reductions in skilled nursing facility utilization or spending:
    - Model 2 Hospitals -2.9 days;
    - Model 2 PGPs -2.3 days;
    - Model 3 SNF -13.2% expenditures
  - Reductions in institutional rehabilitation facility expenditures:
    - Model 2 Hospital: -16.6% expenditures;
    - Model 2 PGP -32.3% expenditures;
    - Model 3 Home Health Agencies [HHAs]: -29.2% expenditures
  - Increases in home health utilization or expenditures:
    - Model 2 Hospitals 4.6% expenditures;
    - Model 2 PGP 12.7% expenditures;
    - Model 4 Hospitals 17.6% (at p<0.1 level)
    - Model 3 SNF: 9.2% visits

- The Bundled Payments for Care Improvement Advanced (BPCI-A) Model had improvements in institutional PAC, SNF, and IRF utilization or expenditures within many episodes within the hospital setting (medical and surgical episodes) with only one episode unfavorably increasing HH expenditures as well as some non-significant changes in certain episodes. Physician group practices participating in the medical episodes had improvements in some episodes for institutional PAC utilization and non-significant changes in SNF, IRF, and HH utilization. Physician group practices participating in surgical episodes had some improvements in institutional PAC utilization, SNF

utilization, IRF expenditures, and HH expenditures, but many episodes had non-significant changes.

- The Comprehensive ESRD Care (CEC) Model reduced PAC per beneficiary per month (PBPM) spending (-5.5%) but did not change HH per beneficiary per month spending.

- First episodes of PAC in the Comprehensive Joint Replacement (CJR) Model significantly reduced IRF (-28.1%) and increased HH (20.5%), but did not change SNF.

- The Home Health Value-Based Purchasing (HHVBP) Model significantly reduced SNF utilization (-6.9%) but did not significantly change HH use.

- The Maryland (MD) All-Payer Model had significant reductions in PAC spending (-5.9%).

- The Medicare Care Choices Model (MCCM) reduced PAC days (13%).

- The Oncology Care Model (OCM) did not significantly change IRF spending or SNF and HH utilization.

- The Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport (RSNAT) Model increased HH (1.8%) and SNF expenditures (1.3%), which was considered unfavorable.

Table 3. Post-acute Care for Acute or Specialty Care and Targeted Population Models

Model	Any PAC	Institutional PAC	SNF	IRF	HH
BPCI Model 2 Hospitals	-0.4%	**-3.6%**	**-2.9 days**	**-16.6% $**	**4.6% $**
BPCI Model 2 PGPs	**-5.6%**	-1.8%§	**-2.3 days**	**-32.3% $**	**12.7% $**
BPCI Model 3 SNFs	**-3.5%**	--	**-13.2% $**	2.5% $	**9.2%**
BPCI Model 3 HHAs	-1.8%	--	-7.0% $	**-29.2% $**	-3.9%
BPCI Model 4 Hospitals[37]	**7.1%**	2.2%	-1.1 days	31.5% $	17.6% $ §
BPCI-A Hospitals Medical [38]	--				
BPCI-A Hospitals Surgical [39]	--				
BPCI-A PGP Medical [40]	--				
BPCI-A PGP Surgical [41]	--				
CEC	**-5.5% $**	--	--	--	-0.4% $
CJR *	--	--	-6.5%	**-28.1%**	**20.5%**
HHVBP *	--	--	**-6.9%**	--	-2.1%
MD All-Payer Model	**-5.9% $**	--	--	--	--
MCCM	**-13.0%**	--	--	--	--
OCM	--	--	-0.1%	0.0% $	-0.1%
RSNAT	--	--	**1.3% $**	--	**1.8% $**

Table Key: Bolded estimates are statistically significant at least at the p<0.05 level. " § " indicates an estimate that was statistically significant at the p<0.10 level. Results that were in a favorable direction and statistically significant have a light green cell shading. Results that were unfavorable and statistically significant have red/orange cell shading. Cells with gray shading illustrate non-significant results. "--" indicates the measure was not relevant or available. " * " indicates the model included a randomized design. If utilization measures were not available, payment categories were used and denoted with "$" to make it clear it was a payment and not a utilization measure. Many of these measures are conditional on discharge to a post-acute care setting and calculated among a subset of the sample that used post-acute care services. BPCI=Bundled Payments for Care Improvement; BPCI-A=Bundled Payments for Care Improvement Advanced; CEC=Comprehensive End-Stage Renal Disease Care; CJR= Comprehensive Joint Replacement; HHVBP=Home Health Value-Based Purchasing; MD= Maryland; MCCM= Medical Care Choices Model; OCM=Oncology Care Model; PGP=Physician Group Practices; RSNAT= Prior Authorization of Repetitive, Scheduled Non-Emergency Ambulance Transport.

---

[37] Model 4 results represent 30-day post-discharge period (90-day post-discharge period results were also available).

[38] Two of 10 episodes had significant reductions in discharges to institutional PAC while eight had non-significant changes. Of 10 episodes, 9 had significant reductions in SNF days and one had non-significant changes. Of the five episodes examined, only two (stroke and UTI) had significant reductions in IRF payments. HH payments increased for all five medical clinical episodes examined but only one was statistically significant.

[39] MJRLE had significant institutional PAC reductions while hip & femur had nonsignificant reductions. Both MJRLE and hip & femur episodes had statistically significant reductions in SNF days and payments. MJRLE and hip and femur procedures had significant reductions in IRF payments. Hip and femur procedures (except major joint) had a significant increase in HH payments but MJRLE had non-significant decreases.

[40] Of the 11 episodes examined, six had significant reductions in institutional PAC at p<0.1 and 5 were non-significant. Of the 11 episodes examined, none had significant changes in SNF utilization.

[41] Of the 7 episodes examined, 4 had significant reductions in institutional PAC at p<0.05 level and three had non-significant declines. Of the 7 episodes examined, six had significant reductions in SNF days for SNF users (p<0.1) and one had no statistically significant changes. Of the six episodes examined, four had significant reductions in IRF payments (p<0.05) and two had no statistically significant changes. Of the six episodes examined, three had significant reductions in HH payments (p<0.05) and three had no statistically significant changes.

Quality of Care Measures

Many of these models examined similar quality of care measures, including patient experience and mortality. There were no significant improvements in beneficiaries' self-reported experience of care for seven(BPCI (Model 2 PGPs, Model 3 SNF, Model 3 HHA), BPCI-A, OCM, CEC[42], HHVBP, CJR) of the models in this grouping with results on this measure[43]. One model, MCCM, had significant improvements in caregivers' experience of care[44]. Fewer BPCI[45] beneficiary respondents reported the highest levels of satisfaction with their care relative to similar beneficiaries not the in model, although these differences were generally small in magnitude and were not accompanied by worse functional status outcomes. Three of eight models[46] had significant reductions in mortality, including MD All-Payer Model (-8.8%)[47], CEC (-2% at p<0.1) and HHVBP (-37%) while BPCI[48], BPCI-A[49], OCM, RSNAT, and CJR had no significant changes.

Lessons Learned & Iterative Testing

The successful results from many of the models in this grouping have informed the CMS Innovation Center's development of new models, choice to continue existing models, as well as whether to seek certification for potential expansion of a model's scope or duration, including implementation on a nationwide basis. To date, two models in this grouping have been certified for national expansion (HHVBP[50], RSNAT[51]). Successes in the MD All-Payer Model led to the development and implementation of the Maryland (MD) Total Cost of Care Model, which is currently operating through 2026. Lessons learned from CEC informed the Kidney Care Choices model[52] that began in 2022. A prior synthesis paper on episode payment models[53] noted that results from the Major Joint Replacement Lower Extremity bundle in BCPI Model 2 informed the design of CJR and BPCI-A. The length of the CJR model has been extended to allow for changes to the model design

---

[42] Overall, the differences in the beneficiary survey composite scores between CEC and comparison groups were small in magnitude and did not suggest clinically meaningful associations between model participation and beneficiaries' quality of life. Although there were statistically significant differences between participants in CEC and the comparison group for three of the composite scores, none of the estimates were deemed clinically meaningful.

[43] RSNAT did not have experience of care survey results. Impact estimates based on difference-in-differences regression comparing the intervention group to a comparison group were not available for MD All-Payer, although overall the trend analyses did not show any declines in care.

[44] Approximately 9 out of 10 caregivers of MCCM enrollees who transitioned to the Medicare Hospice Benefit reported care consistent with the beneficiary's wishes and favorable experiences of shared decision making about hospice enrollment. Caregiver perceived supportive services alongside treatment for terminal conditions prior to transitioning to hospice was beneficial and reported that MCCM enrollees who transitioned to the Medicare hospice benefit generally experienced very good care that was comparable to care experienced by those enrolled directly into hospice.

[45] Results were from the model's Fifth Evaluation Report (all other estimates in the table for this model come from the Seventh Evaluation Report).

[46] The MCCM evaluation matched on survival time to construct the comparison group. Therefore, examining overall mortality was not feasible or relevant.

[47] The percentage of Medicare admissions classified as having major or extreme severity or risk of mortality in Maryland hospitals increased less than in the comparison group. This finding was likely driven by a decrease in the intensity of services provided to the sickest hospital patients in Maryland.

[48] Mortality results for the overall (pooled) analyses show no changes but some individual clinical episodes showed significant changes.

[49] Mortality results for the overall (pooled) analyses show no changes but some individual clinical episodes showed significant changes.

[50] https://www.cms.gov/newsroom/press-releases/cms-takes-action-improve-home-health-care-seniors-announces-intent-expand-home-health-value-based

[51] https://www.cms.gov/newsroom/press-releases/cms-expand-successful-ambulance-program-integrity-payment-model-nationwide

[52] Boyer G, Duvall T, Wells C. 2022. The Medicare CEC Model: Using Lessons Learned to Improve Kidney Care, Health Affairs Forefront: https://www.healthaffairs.org/do/10.1377/forefront.20220116.500107

[53] CMS Episode Payment Models, January 2020: https://innovation.cms.gov/files/reports/episode-payment-models-wp.pdf

to be tested further. Learnings from OCM were used in the development of the Enhanced Oncology Model[54] that will begin in 2023. Finally, aspects of MCCM are being used in the development of new CMS Innovation Center models related to palliative care.

Primary Care and Population Management

Model Grouping Characteristics

The grouping of Primary Care and Population Management models includes 12 Medicare models, all of which had voluntary participation. This category of models encompassed interventions focused on prevention (primary, secondary, tertiary) and care coordination, inclusive of the management of diseases targeting primary care providers to coordinate care in FFS or provide disease management to beneficiaries with chronic conditions either through health homes or managed care networks. The payment models in this grouping are provided in the table below as well as described further in Appendix Table 2.

Primary Care and Population Management Models	Payment
Accountable Care Organization (ACO) Investment Model (AIM)	Up-front and monthly payments recouped from shared savings
Advance Payment (AP) ACO Model	Up-front and monthly payments recouped from shared savings
Comprehensive Primary Care (CPC) Initiative	Prospective care management fees and regional shared savings payments
Comprehensive Primary Care+ (CPC+)	Prospective care management fees, performance-based incentive payments, and Track 2 hybrid payment (prospective quarterly population-based payment and reduced FFS)
Financial Alignment Initiative for Medicare-Medicaid Enrollees (FAI) Demonstration, Washington[55]	Shared savings with State in managed FFS model, with care coordination as a Medicaid covered benefit through health homes, which receive a per-member-per-month Medicaid payment.
Independence at Home (IAH) Demonstration	Practices can earn incentive payments if their patients' expenditures are below target expenditures and they meet quality standards
Medicare Advantage (MA) Value-Based Insurance Design (VBID) Model	No financial incentives from CMS to participate in this model; CMS waiver of the MA uniformity requirement allows MA plans to structure cost-sharing and other benefit design elements to encourage beneficiaries with chronic conditions to use high-value care.
Million Hearts® (MH): Cardiovascular Disease Risk Reduction Model	Prospective care management tied to risk stratification and change in risk scores
Next Generation ACO (NGACO) Model	Tests whether strong financial incentives, flexible payment options, and tools to support care management improve value and lower costs.

---

[54] https://innovation.cms.gov/innovation-models/enhancing-oncology-model
[55] FAI involves **t**wo demonstrations that have used a Managed FFS model and 11 that use a capitated model. FAI Washington (a Managed FFS model) has the most demonstration years of data available at this time and that, coupled with its successful results, led to its inclusion in this paper. A high-level summary of evaluation results across the other FAI models is provided in Appendix B.

Part D Enhanced Medication Therapy Management (Enhanced MTM) Model	Prospective and performance-based payments (premium subsidy)
Pioneer ACO Model	Shared savings and losses tied to quality
Vermont All-Payer ACO Model	ACO two-sided risk tied to quality and optional population-based payments similar to financial methodology used in NGACO

Beneficiaries served by this group of model participants had mostly average health relative to the previous grouping of models, with a few exceptions[56]. Beneficiaries had baseline Hierarchical Condition Category[57] (HCC) scores around 1.0[58] or an average count of 4 or 5 chronic conditions[59] and lower average baseline total Medicare spending (ranging from $6,890 to $16,000) among eight of the eleven models in this grouping. Four models in this grouping served beneficiaries with more complex health needs.[60] Models in this grouping ranged in size from as small as 14 home-based primary care sites serving approximately 10,000 beneficiaries in Independence at Home (IAH) Demonstration, to as large as 3,070 Comprehensive Primary Care Plus (CPC+) practices serving approximately 2 million Medicare beneficiaries. Appendix Table 2 provides details on each of these models and two demonstrations including the years covered by the impact estimates contained within in this paper, the intervention, payment model, health care providers involved, beneficiaries touched, and relative size of the model.

Cross-Group Results

Overview

Results for gross Medicare spending across the 12 models in the Primary Care and Population Management grouping had a mixture of non-significant results (six models[61]), significant reductions (five models[62]), and one significant increase[63]. The five models[64] with reductions in spending tended to have significant reductions in inpatient admissions, inpatient readmissions, and/or post-acute care (PAC). Similarly, most of the six models[65] with non-significant changes in spending had limited or

---

[56] Three models (IAH, FAI Washington, and MA VBID) served beneficiaries with higher HCC scores.

[57] HCC scores are based on health conditions and diagnoses in health care claims that are summarized into an algorithm to calculate the patient's health complexity and assign a risk factor to the beneficiary that can be used to predict future patient costs. Higher scores indicate the beneficiaries has more health conditions and more costly health care utilization relative to beneficiaries with lower scores.

[58] These include CPC, CPC+, MH, and AIM.

[59] AP ACO, NGACO, Pioneer ACO Model, and VT All-Payer ACO did not report HCC scores but provided an average number of chronic conditions that would be similar to a low HCC score.

[60] The IAH (average HCC score: 3.5; average baseline Medicare expenditures: $52,764) and FAI Washington (average HCC score ranged from 2.0-2.3; average baseline Medicare expenditures: $14,576) demonstrations served beneficiaries with high needs who had multiple complex chronic conditions. MA VBID had HCC scores ranging from 1.5 to less than 2.0 but did not report average baseline expenditures. Part D Enhanced MTM served beneficiaries with an average of $11,144 in baseline Medicare expenditures with an HCC score of 1.16 for the entire set of beneficiaries who were enrolled in the Part D plans (results from the third evaluation report). However, plan sponsors targeted their beneficiaries based on health care needs and costs with low-income subsidy and medically complex beneficiaries having a range between $13,501 to $19,986 in average baseline Medicare Parts A & B expenditures with an HCC ranging from 1.42 to 1.87 (results from the fourth evaluation report: https://innovation.cms.gov/data-and-reports/2022/mtm-fourth-evalrept).

[61] CPC, CPC+, IAH, MA VBID, MH, Part D Enhanced MTM

[62] AIM, FAI Washington, NGACO, Pioneer ACO Model, VT All-Payer ACO

[63] AP ACO

[64] AIM, FAI Washington, NGACO, Pioneer ACO Model, VT All-Payer ACO

[65] CPC, CPC+, IAH, MA VBID, MH, Part D Enhanced MTM

no impacts on utilization outcomes. Where patient experience of care was assessed (five models[66]), there were no significant changes aside from one model[67]. Four models[68] examined mortality and one[69] had significant reductions while the remaining showed no significant changes[70]. Impact estimates across the 12 models are reviewed below for each measure examined in the cross-model synthesis.

Many of the results presented here represent final evaluation results[71] or at least four performance years[72]. Results for Part D Enhanced Medication Therapy Management (Part D Enhanced MTM), Medicare Advantage (MA) Value-Based Insurance Design (VBID), and the Vermont All-Payer ACO models represent fewer performance years and caution should be used when interpreting these findings as results may change as future evaluation results are available.

Spending & Utilization

Table 4 summarizes overall spending and utilization impact estimates for 12 Medicare models that addressed changes in health care delivery within the primary care setting or through population management. These models served beneficiaries with mostly average health (with a few exceptions[73]).

- Five of the 12 models had significant reductions in gross spending, including Next Generation Accountable Care Organization (ACO) (NGACO -1.2%), Pioneer ACO Model (-2.5%), ACO Investment Model (AIM -3.4%), VT All-Payer ACO (state-level analyses -6.8%; ACO level analyses -5.5%, both significant at $p<0.1$), and Financial Alignment Initiative (FAI) Washington (-$384.7 million cumulative across the six demonstration years).

- There was a significant increase in gross spending for Advance Payment (AP) ACO but no significant changes in gross spending for Independence at Home (IAH)[74], Comprehensive Primary Care (CPC), Comprehensive Primary Care Plus (CPC+), Million Hearts® (MH), Part D Enhanced MTM, and MA VBID.

- Three of the 11 models that had incentive payments[75] showed significant net savings; AIM (-2.5%), VT All-Payer ACO (state-level analyses -6.5%, ACO level analyses were not significant), and FAI Washington (-$297 million, cumulatively).

- Three of the 11 models that included incentive payments had significant net losses (CPC+ Track 1: 1.5%; Track 2: 2.6%; AP ACO: $242 million; NGACO: 0.4% at $p<0.1$ level).

The limited changes in spending were reflective of limited improvements in costly utilization within the inpatient setting.

---

[66] AIM, CPC, CPC+, MA VBID, Pioneer ACO Model
[67] Pioneer ACO Model
[68] AIM, CPC+, IAH, MH
[69] MH
[70] AIM, CPC+ IAH
[71] AP ACO, AIM, CPC, Pioneer ACO Model
[72] CPC+, FAI Washington, IAH, MH, NGACO
[73] FAI, IAH, and MA VBID served beneficiaries with complex chronic conditions.
[74] Spending and utilization results were not significant for IAH in the sixth performance year, where two fewer practices were participating than in the fifth performance year.
[75] MA VBID did not have financial incentives to participate in the model.

- Four models (CPC+ Track 2: -1.1%, Pioneer ACO Model: -3.5%, VT All-Payer [ACO: -17.9% and state-level: -9.3%[76]]) had significant reductions in inpatient admissions or inpatient spending (AIM: -3.3%) at the $p<0.05$ level and two models (CPC: -2.0% and CPC+ Track 1: -0.9%) had reductions significant at the $p<0.1$ level.

- One model, MH had significant increases in inpatient admissions (+3.8%).

- Six models (IAH, AP ACO, NGACO, FAI Washington, Part D Enhanced MTM, and MA VBID) had non-significant changes in inpatient admissions.

- Five of 11 models had significant reductions in ED visits (IAH: -4.9%[77], CPC: -2.0%, CPC+ Track 1: -1.8%, Track 2: -1.7%, Pioneer ACO Model: -0.04% AIM: -2.1%) while two models had significant increases (MH: 2.9% significant at $p<0.1$ level, Part D Enhanced MTM: 2.5%).

- Three models demonstrated significant reductions in inpatient readmissions (AIM: -4.2%, VT All-Payer model state-level analysis only -22.4%, Part D Enhanced MTM: -3.4%) while the remaining nine models that examined this measure showed non-significant changes (IAH, CPC, CPC+ Years 1-3, Pioneer ACO Model, AP, NGACO, VT All-Payer ACO level analyses, FAI Washington Years 4-6).

---

[76] Inpatient admissions in the VT All-Payer ACO model were not statistically significant in the first PY but became significant in the second PY.
[77] Spending and utilization results were not significant for IAH in the sixth performance year, where two fewer practices were participating than in the fifth performance year.

Table 4. Spending & Utilization for Primary Care & Population Management Models

Model	Spending		Utilization		
	Gross	Net	Inpatient admissions	ED Visits	Inpatient readmissions
AIM Test 1 ACO [78]	**-3.4%**	**-2.5%**	**-3.3% $**	**-2.1%**	**-4.2%**
AP ACO [79]	**0.4%**	**$242 million**	-0.7% $	--	--
CPC	-1.0%	1.0%	-2.0%§	**-2.0%**	-1.4%
CPC+ Track 1	0.2%	**1.5%**	-0.9%§	**-1.8%**	
CPC+ Track 2	0.1%	**2.6%**	**-1.1%**	**-1.7%**	
FAI WA Demonstration [80]	**-12.6%**	**-$216.8 million**	-2.9%	0.9% [81]	-0.4%
IAH Demonstration [82]	-4.6%	-$44.3 million	-4.6%	**-4.9%**	-7.0%
MA VBID [83]	--	--	--	--	--
MH * [84]	0.3%	0.5%	**3.8%**	2.9%§	--
NGACO	**-1.2%**	0.4%§	-0.1%	-- [85]	0.3%
Part D Enhanced MTM[86]	-0.25%	$146.7 million	-0.2%	**2.5%**	**-3.4%**
Pioneer ACO Model [87]	**-2.5%**	**-$254 million**	**-3.5%**	**-0.04%**	0.4%
VT ACO [88]	-5.5%§	-4.7%	**-17.9%**	4.9%	-12.4%
VT state-level [88]	-6.8%§	**-6.5%**	**-9.3%**	2.6%	**-22.4%**

Table Key: Bolded estimates are statistically significant at least at the p<0.05 level. " § " indicates an estimate that was statistically significant at the p<0.10 level. Results that were in a favorable direction and statistically significant have a light green cell shading. Results that were unfavorable and statistically significant have red/orange cell shading. Cells with gray shading illustrate non-significant results. "--" indicates the measure was not relevant or available. " * " indicates the model included a randomized design. AIM=ACO Investment; AP=Advance Payment; ACO=accountable care organization; CPC=Comprehensive Primary Care; CPC+=Comprehensive Primary Care Plus; MTM=Medication Therapy Management; FAI=Financial Alignment Initiative for Medicare-Medicaid Enrollees; IAH=Independence at Home Demonstration; MA=Medicare Advantage; MH=Million Hearts®: Cardiovascular Disease Risk Reduction Model; NGACO=Next Generation ACO Model; VBID=Value-Based Insurance Design Model; VT=Vermont.

---

[78] Evaluation results for AIM are provided for the Test 1 ACOs only from 2016-2018 (Test 2 ACOs were not included in this table). Cumulative results were not available; reported results are averages for the three performance years.

[79] Impact estimates represent performance during 2012 – 2014.

[80] Spending and utilization results represent the most recent demonstration years (4-6) only. Relative differences were not available for net expenditures.

[81] Relative differences for the probably of inpatient admissions, ED visits, and all-cause 30-day readmissions for Years 4-6 were not included in the evaluation report (as these estimates were not statistically significantly), however they were calculated internally. Results for Years 1-3 were consistent with results for Years 4-6. Regression-adjusted estimates are available in the evaluation report for Years 4-6 as well as prior demonstration years.

[82] Cumulative IAH results are through the fifth year since two fewer practices participated and results for all outcomes were null in the sixth year (see Sixth Evaluation Report).

[83] Relative differences were not reported for gross spending, inpatient admissions, or ED visits. The evaluation report provides regression estimates in percentage points.

[84] Evaluation results are provided for the MH high- and medium- risk beneficiaries, which were the target population of the model.

[85] Cumulative estimates for ED visits and observation stays were not interpretable (failure of parallel trends assumption for baseline years). PY4 estimates show a non-significant change (-1.2%).

[86] Results presented in this table for Part D Enhanced MTM were from the third evaluation report. The fourth evaluation report for Part D Enhanced MTM, which is now available, assessed the impact of the model on Low Income Subsidy and Medically Complex Subgroups. These subgroups were more likely to have been targeted for MTM services. Similar to the all-enrollee analysis in the third evaluation report, these subgroups did not experience a statistically significant reduction in Medicare Parts A and B expenditures.

[87] Cumulative results were not available; reported results are averages of the first and second performance years (2012 – 2013). Results compare beneficiaries in Pioneer ACO Model to beneficiaries in their "near" market.

[88] Cumulative results for gross and net spending were available across PY1 and PY2 (2018-2019). Cumulative results were not available for utilization, so results presented for PY2 (2019). However, utilization results for PY1 were not statistically significant.

Post-Acute Care Impact Estimates

Table 5 summarizes changes in post-acute care (PAC) utilization and spending for the 12 models that addressed changes in care delivery using primary care or population management.

- No models in this grouping examined the measure combining all PAC utilization and in general PAC was less relevant for this set of models given the larger focus on prevention of the development or progression of disease in the community setting.

- Only one model examined institutional PAC, Next Generation Accountable Care Organization (NGACO), which had significant reductions in spending in this setting (3.9%).

- Eleven of the 12 models in this group examined skilled nursing facility (SNF) spending or utilization. Three of the four accountable care organization (ACO) models (Pioneer ACO Model, ACO Investment Model [AIM], NGACO) had significant reductions in SNF utilization or spending as did Financial Alignment Initiative [FAI] Washington (Years 4-6), Part D Enhanced Medication Therapy Management (Part D Enhanced MTM), and Medicare Advantage (MA) Value-Based Insurance Design (VBID while the remaining four models (Independence at Home [IAH], Comprehensive Primary Care [CPC], Comprehensive Primary Care Plus [CPC+], Vermont [VT] All-Payer ACO) had non-significant changes.

- Three models looked at changes in institutional rehabilitation facility (IRF) utilization or spending. Pioneer ACO Model had significant declines in IRF/long-term care days (-0.05%) but AP ACO did not have significant changes in IRF spending. CPC+ Track 1 and 2 both had unfavorable increases in IRF expenditures.

- Increases in home health (HH) utilization and spending can be viewed as favorable or unfavorable depending on the model's theory of action. For the ACO models, decreases in HH were typically viewed as favorable with AP ACO, Pioneer ACO Model, and VT All-Payer ACO demonstrating significant declines in either utilization or spending. There were significant declines in HH spending in CPC+ (Track 1 and Track 2).[89] Four models had non-significant changes in HH utilization or spending (IAH, CPC, VT All-Payer state level analyses, MA VBID).

---

[89] Given that the CPC+ model focuses on improving primary care, reductions in HH were not expected and could be viewed unfavorably, although the intended direction for this measure did not have a strong hypothesis ahead of time.

Table 5. Post-acute care for primary care & population management models[90]

Model	Any PAC	Institutional PAC	SNF	IRF	HH
AIM Test 1 ACO (2016 – 2018)[91]	--	--	**-5.5% days**	--	--
AP ACO (2012-2014)	--	--	**3.5% $**	-4.0% IRF/LTC $	**-3.7% $**
CPC	--	--	-5% $	--	-1% $
CPC+ Track 1 [92]	--	--	-0.4% $	**5.4% $**	**-3.0% $**
CPC+ Track 2 [61]	--	--	0.4% $	**6.3% $**	**-2.6% $**
FAI WA Demonstration	--	--	**-24.2%**	--	--
IAH Demonstration	--	--	-0.8% $	--	-0.7% $
MA VBID [93]	--	--	-0.4 pp§	--	
MH [94]	--	--	--	--	--
NGACO	--	**-3.9% $**	**-2.0% $**	--	--
Part D Enhanced MTM	--	--	**-4.0% LOS** -1.0% admissions	--	--
Pioneer ACO Model [95]	--	--	**-0.01% days**	-0.05% IRF/LTC days	**-0.02% visits**
VT ACO (PY2 – 2019) [96]	--	--	-3.5% days	--	-25.2% visits§
VT state-level (PY2 – 2019) [58]	--	--	-7.5% days	--	1.4% visits

Table Key: Bolded estimates are statistically significant at least at the $p<0.05$ level. " § " indicates an estimate that was statistically significant at the $p<0.10$ level. Results that were in a favorable direction and statistically significant have a light green cell shading. Results that were unfavorable and statistically significant have red/orange cell shading. Cells with gray shading illustrate non-significant results. "--" indicates the measure was not relevant or available. " * " indicates the model included a randomized design. LOS=length of stay; LTC=long-term care; pp=percentage point. "$" indicates the measure represents payments instead of utilization. AIM=ACO Investment; AP=Advance Payment; ACO=accountable care organization; CPC=Comprehensive Primary Care; CPC+=Comprehensive Primary Care Plus; MTM=Medication Therapy Management; FAI=Financial Alignment Initiative for Medicare-Medicaid Enrollees; IAH=Independence at Home Demonstration; MA=Medicare Advantage; MH=Million Hearts®: Cardiovascular Disease Risk Reduction Model; NGACO=Next Generation ACO Model; VBID=Value-Based Insurance Design Model; VT=Vermont.

[90] Many of these measures are conditional on discharge to a post-acute care setting and are calculated among a subset of the sample that used post-acute care services. If utilization measures were not available, payment categories were used and denoted with "$" to make it clear it was a payment and not a utilization measure.

[91] Evaluation results for AIM are provided for the Test 1 ACOs only (Test 2 ACOs were not included in this table). Cumulative results were not available, so reported results are averages for the three performance years.

[92] Results presented in this table for CPC+ were from the third evaluation report. The fourth evaluation report for CPC+ had similar results for SNF, IFR, and HH spending.

[93] Relative differences were not reported for SNF or HH visits. The evaluation report provides regression coefficients.

[94] The Million Hearts® evaluation did not examine post-acute care as it was not hypothesized as a primary outcome for the model.

[95] Cumulative results were not available, so reported results are averages of the first and second performance years. Results compare beneficiaries in Pioneer ACOs to beneficiaries in their "near" market.

[96] Cumulative results were not available for utilization, so results presented for PY2. However, utilization results for PY1 were not statistically significant.

Quality of Care Measures

Among the five models[97] with experience of care survey results relative to a comparison group, one showed significant changes (Pioneer ACO Model[98]), while the remaining showed no significant or substantial changes (CPC, CPC+[99], AIM, MA VBID). For the four models[100] that examined mortality, only one model showed significant improvements (Million Hearts®: 4.0% which amounts to 2.8 fewer deaths per 1,000 people over three years in the intervention group compared to the control group) while the remaining showed no significant changes (IAH, CPC+, AIM).

Lessons Learned & Iterative Testing

The Pioneer ACO Model, as tested during the first two performance years of the model, was certified in 2015 by the CMS Office of the Actuary; it was the first model that met the expansion criteria[101] Early successes in the Pioneer ACO Model were incorporated into the Medicare Shared Savings Program as well as used in the development of NGACO. Lessons from NGACO were subsequently built into the Global and Professional Direct Contracting Model that launched April 2021, which has been redesigned and renamed the ACO Realizing Equity, Access, and Community Health (REACH) Model starting in 2023. Lessons from AP ACO were used to inform the AIM model. Successful results from AIM were used to inform changes to the 2023 Physician Fee Schedule proposed rule that would allow qualifying new Shard Savings Program ACOs to receive up-front funds that could be recouped from future shared savings[102]. Similarly, the experience gained in CPC was used to design CPC+. The same underlying principles in CPC+ were used in the design of the Primary Care First Model that began in 2021. Substantial changes to the MA VBID Model and participation began in 2021, including but not limited to supplemental benefit targeting by socioeconomic status, and including the Medicare hospice benefit as a covered benefit within the Medicare Advantage benefits package. Other models in the grouping have ended (MH, Part D Enhanced MTM) or have been extended (FAI Washington) although implementation learnings may be used in the planning for future models.

Discussion

Looking across results of the 21 Medicare models included in this synthesis there are a number of considerations for planning future CMS Innovation Center models. More than half (fourteen) of

---

[97] AIM, CPC, CPC+, MA VBID, and Pioneer ACO Model had experience of care survey results relative to a comparison group. AP ACO, MH, NGACO, VT All-Payer ACO, and Part D Enhanced MTM did not include experience of care survey results. Beneficiaries and caregivers were surveyed in the IAH evaluation with results presented in the Year Four report, however these were not tested against a comparison group. Experience of care survey results for FAI WA were only provided for the demonstration group and were not assessed relative to a comparison group.

[98] Beneficiaries served by Pioneer ACOs were significantly more satisfied on average with the timeliness of care, appointments, information, how well their health care provider communicated, and overall rating of health care provider relative to FFS beneficiaries. Beneficiaries served by Pioneer ACOs were significantly lower on access to specialists and ease of getting care. However, the magnitudes of these significant effects do not appear materially significant as the performance scores were similar to beneficiaries not touched by the model.

[99] Compared to Medicare FFS beneficiaries in similar practices, beneficiaries served by CPC+ practices in the third performance year reported similar experiences with most aspects if care covered in the beneficiary survey. The expectation was that beneficiaries served by Track 2 CPC+ practices were more likely to report that they received timely follow-up after a hospitalization than beneficiaries in similar practices that were not participating in CPC+. These findings were consistent with those in the second performance year.

[100] CPC, Pioneer ACO Model, NGACO, VT All-Payer ACO, FAI WA, Part D Enhanced MTM, and MA VBID did not examine mortality.

[101] https://wayback.archive-it.org/3926/20170127185820/https://www.hhs.gov/about/news/2015/05/04/affordable-care-act-payment-model-saves-more-than-384-million-in-two-years-meets-criteria-for-first-ever-expansion.html

[102] https://www.cms.gov/newsroom/fact-sheets/calendar-year-cy-2023-medicare-physician-fee-schedule-proposed-rule-medicare-shared-savings-program

these models demonstrated gross savings to Medicare driven by reduced utilization and spending in inpatient admissions and/or more efficient post-acute care. For models with financial incentives paid to participants, about six realized net savings while six incurred net losses to CMS. Ten models reduced inpatient admissions and fourteen improved post-acute care. Seven models reduced emergency department visits and/or inpatient readmissions. Four models had unfavorable increases in care in these settings, although in some cases these increases could have been the result of increased access and engagement with the health care system that may dissipate over time. Many of the results presented are either final estimates or represent at least four years of performance suggesting we have more sufficient time to draw conclusions. However, results for a few models represent only two to three years of performance and should be considered preliminary. Future reports from these models may provide changes to the themes or conclusions presented at the point in time this was written.

Based on the data examined in this synthesis paper, beneficiaries' quality of care is being maintained with a few examples of improvements. Beneficiaries self-reported experience of care remained relatively unchanged by the majority of models with data on this outcome aside from one model with unfavorable findings that were small in magnitude and two models with favorable improvements. While experience of care did not necessarily improve across most models, it was generally similar for beneficiaries within the models relative to beneficiaries not being seen by providers in the models, and therefore did not decline in the nine models that had non-significant findings. Medicare beneficiaries in most models tend to report high levels of satisfaction with care that can be difficult to improve upon further by model participants. There were notable mortality benefits for beneficiaries in four models. Eight models had non-significant changes in mortality, suggesting these models did not cause any harm. There are limited quality of care measures available that are similar to examine across the 21 models as most evaluations focus on quality measures that are uniquely relevant to the model's theory of action, which are designed to drive model-specific changes in care delivery. More details on changes in model-specific quality measures are available in Appendix B. Quality measurement in some of the evaluations presented in this synthesis may have been constrained by the need to capture changes relative to a comparison group and because of the cost and burden of collecting patient-reported outcome measures.

Model design features such as mandatory or voluntary participation played important roles in detecting improvements in care delivery and spending. Mandatory and randomized models (two models) all had significant reductions to Medicare expenditures as well as reductions in utilization categories to drive savings. Many of the voluntary models (twelve of nineteen) produced gross savings and reductions in institutional care, particularly post-acute care, but did not produce net savings due to generous financial incentives that ensured robust participation in the model. Voluntary models were also encumbered by participants entering and exiting the model when it is in the participants', and not necessarily Medicare's, best financial interest. Voluntary models suffer from potential selection bias in terms of which participants choose to participate, making it more challenging for the model evaluation to construct a comparison group of similar health care providers and beneficiaries to create unbiased impact estimates.

The two groupings of models essentially serve two distinct purposes, with the majority of models in the first grouping structured as episodic or focused on high cost and specialty care and the second grouping structured as population-based models. These purposes directly relate to the differences between the groups in the types of participants, health care providers, and beneficiaries served, all of which were associated with the evaluation findings. The two groupings of models serve different purposes to effectively manage both the complex and heathier populations that make-up the Medicare population. The Acute or Specialty Care and Targeted Populations models produced large

effects on outcomes likely due to the high-cost beneficiaries served by these models using the targeted and specialized services. The Primary Care and Population Management models operating through primary care practices, accountable care organizations (ACOs), health homes, or health plan networks typically had smaller average declines in spending and other outcomes. Prevention-based interventions are expected to show benefits over longer time horizons. This coupled with the large numbers of relatively healthy beneficiaries served by these models makes it difficult to show improvements in the short term. Below we review themes from evaluation results within the two groupings of models, factors that could have influenced results, and highlight limitations to this analysis.

Considerations for Acute or Specialty Care & Targeted Populations Models

Models within this grouping served beneficiaries that may have used costly institutional or care that would be otherwise potentially avoidable if their condition was not properly managed. The overwhelming majority of these models were able to demonstrate significant reductions in utilization such as inpatient admissions and post-acute care (PAC) that drove down Medicare spending, with some improvements in quality. All of these models had reductions in gross Medicare spending and most had reductions in inpatient admissions and PAC. Two models had reductions in emergency department (ED) visits and less than half (four models) reduced inpatient readmissions. Many of these models covered short time-windows of acute care or serious illness, while a few managed costly, chronic diseases over longer periods of time. These models were run by participants operating within the institutional setting or were targeting specific patient populations through specialty care providers that manage the care of patients with complex chronic conditions. The involvement of these providers allowed for management and prevention of beneficiaries' use of inpatient services. The high cost nature of these beneficiary populations affords more opportunities to reduce expenditures and utilization relative to population-based models that include larger proportions of relatively healthy beneficiaries managed over longer time windows. Two key contextual factors separated this grouping of models. First, there were differences in results based on model design features such as mandatory or voluntary participation and whether the model provided financial incentives for participation. Second, results also varied based on whether the model served targeted or broader patient populations.

Participation and Financial Incentives

This grouping of models included model designs that were randomized with mandatory participation. Mandatory and randomized models such as Home Health Value-Based Purchasing (HHVBP), and Comprehensive Joint Replacement (CJR) both had significant reductions to Medicare expenditures as well as reductions in key utilization categories to drive savings. Voluntary models tend to include more efficient health care providers that, in general, are more likely to volunteer to participate in the model after having already found ways to be cost-efficient and provide better quality of care. The mandatory models required eligible health care providers in geographic locations to participate in the model, allowing for the inclusion of less efficient and higher cost health care providers. This potentially allows more savings to be realized as these less efficient health care providers respond to model incentives designed to reduce Medicare utilization and expenditures.

Two other models in this grouping, Maryland (MD) All-Payer and Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport (RSNAT), also had successful evaluation results and geographic factors that could have influenced participation in the model. The MD All-Payer Model and RSNAT were not randomized nor mandatory, however the geographic location of these models influenced participation. In RSNAT, ambulance suppliers in selected states may have been affected by the model because their reimbursement was tied to receiving prior authorization or prepayment review to transport beneficiaries. Within one year of model implementation, 36% of RSNAT suppliers exited the market. While the MD All-Payer Model was voluntary, all regulated hospitals within the state participated. An alternative to the model would have involved converting the state's payment rates to match Medicare's inpatient and outpatient payment systems which would have drastically cut payment rates for the state's hospitals, a less favorable scenario that could have affected the decision to choose to participate in the model in a different way than other voluntary models. Hospitals regularly monitored their volume and adjusted their rates during the course of the model to meet global budget targets.

The interventions of select models created incentives for efficiency or improved quality without providing incentive payments to participants. Both RSNAT, through prior-authorization/prepayment review, and MD All-Payer, through a Medicare waiver, do not incur a cost to CMS related to providing financial incentives to participants. While HHVBP provides bonus payments to qualifying participants, it is budget neutral. Therefore, value-based incentives in HHVBP[103], RSNAT[104], and MD All-Payer[105] each had the advantage of not having the added costs to CMS for financial incentives related to participation, allowing for improvements in care delivery to directly reduce spending to the Medicare program.

Voluntary models with financial incentives for participants such as the Oncology Care Model (OCM), Comprehensive End-Stage Renal Disease (ESRD) Care (CEC), Bundled Payments for Care Improvement (BPCI) Initiative, and preliminary results from Bundled Payments for Care Improvement Advanced (BPCI-A) demonstrated significant declines in gross spending but large value-based incentive payments made it difficult for these models to demonstrate net savings, and in some cases, there were significant net losses. Generous financial incentives in both CEC and OCM eliminated the opportunity for net reductions in expenditures but incentivized participants to remain in the model. OCM distributed $464.9 million in monthly payments to qualifying practices and $106.5 million in incentive payments through the fifth performance period[106]. The CEC Model distributed $197 million in shared savings payments to qualifying ESRD Seamless Care Organizations through the fourth performance year[107]. Financial incentives in the BPCI and BPCI-A models were also substantial. BPCI waived downside risk early in the initiative due to technical challenges, which contributed to the net losses as participants with episodes above the target price would have had to pay the difference to CMS, thus inhibiting Medicare savings. BPCI also had a large number of participants (40-60%) exit the model that tended to be high cost, which also contributed to net losses. CMS made significant design changes to BPCI-A starting in 2021 (fourth performance year), in order to meet statutory requirements, that intends to improve the model's target pricing based on evidence that target prices were too high for medical episodes but were more accurate for surgical episodes.

The Medicare Care Choices Model (MCCM), which was also a voluntary model, produced significant net savings, even after accounting for hospice expenditures which were directly increased by the model, but faced a large number of participants exiting the model (60%) with enrollment concentrated in a small number of hospices. This, coupled with the small number of beneficiaries served by the model (representing less than 1% of those who lived near participating hospices and met the claims-based MCCM eligibility criteria), limits the generalizability of the model to the broader CMS beneficiary population.

Targeted versus Broader Patient Populations

The targeted nature of beneficiary populations served by specialty providers in CEC, CJR, HHVBP, MCCM, RSNAT likely resulted in improvement in utilization or quality. These models focused on a specific set of diseases or conditions that may have allowed participants to focus on

---

[103] HHVBP has financial incentives but the budget neutral design allows for these incentives to be distributed to the eligible HH agencies while reducing spending to ineligible HH agencies in the mandatory states.

[104] RSNAT uses prior-authorization (no other financial incentives) to determine whether eligible services can receive payment in selected states or resulting claims are subject to prepayment review.

[105] MD All-Payer Model uses a waiver (no other financial incentives from CMS).

[106] Results from the sixth evaluation report.

[107] Shared savings payments for the fifth performance year were not available in the final CEC evaluation report.

aligning incentives around condition-specific care delivery. The patient populations within each of these models are fairly homogenous relative to the broader, general Medicare population covered in the population-based models, which allows for a discrete set of care delivery practices and quality measures for health care providers to focus on.

Conversely, OCM, BPCI, BPCI-A, and MD All-Payer Model covered a broader range of populations with a mixture of more expensive and less expensive patient case mixes. The varying patient mix diluted significant effects seen among more costly episodes/patients relative to less costly episodes/patients, resulting in more modest model-wide effects. Although OCM is solely focused on cancer treatment, it included a broad range of cancer types with varying levels of severity, including lower-risk cancers episodes[108] with lower total Medicare expenditures at baseline (around $7,000), relative to higher-risk cancer episodes[109] that had higher baseline expenditures (about $40,000). In OCM, higher-risk episodes had significant reductions in gross spending (-1.3%) but lower-risk episodes on average increased spending (2.1%) resulting in an overall smaller model-wide effect on gross spending (1%) and consequent net losses. Both BPCI and BPCI-A episodes include a range of procedures and conditions, some of which are cost drivers, while others have shown more limited effects on outcomes as is evident in the differential effects by the various model types and subgroups. For example, BPCI Model 3, serving beneficiaries requiring post-acute care, had much larger effects on gross spending (-7.6% and -7.5%) relative to Model 2 (-3.1%), which was focused on both acute and post-acute care episodes. Similarly, there was variation in the level of gross savings for BPCI-A episodes depending on whether they were medical (-2.2%) or surgical episodes (-4.5%). Even within broader beneficiary populations, there may be important subgroups of vulnerable populations with complex health conditions whose health care utilization may vary relative to healthier beneficiaries. For example, beneficiaries with multiple chronic conditions or who are dually-eligible tend to be higher-cost patients with more complex health care needs and are an important subgroup that may be targeted by model participants for disease management. Hospitals in the MD All-Payer Model served a broad range of Medicare beneficiaries as the model included all Medicare beneficiaries admitted to most Maryland hospitals. Interestingly, beneficiaries with multiple chronic conditions or who were dually-eligible for Medicare and Medicaid served by hospitals in the MD All-Payer Model had greater reductions in expenditures and utilization relative to their healthier counterparts. These results suggest that Maryland hospitals may have prioritized high-cost, high-need patients as they changed their care delivery practices.

Considerations for Prevention and Population Management Models

The majority of the twelve models in this grouping included large panels of relatively healthy Medicare beneficiaries and were focused on prevention and improving care coordination. These models had less pronounced improvements in Medicare spending, utilization, and quality relative to the models serving more targeted populations with higher need patients. These models included participants such as primary care practices, health homes, convening entities, and managed care organizations. Half of the 12 models in this grouping showed significant improvements in inpatient admissions relative to the majority of applicable models in the prior group showing significant improvements for this measure. Half of these models significantly reduced emergency department (ED) visits (a theme not predominately seen in the Acute or Specialty Care and Targeted Populations grouping). Two models in this grouping showed unfavorable changes in inpatient admissions and/or

---

[108] Lower risk cancer episodes in OCM included low-risk breast, low-intensity prostate and low-risk bladder cancer.
[109] Most common high-risk cancer episodes in OCM included lung, myeloma, lymphoma, colorectal, and high-risk breast cancer.

ED visits. There were many models with improvements in post-acute care within this group (seven of the 12 models) but fewer models had improvements in inpatient readmissions (three of the twelve). The focus on primary care and prevention resulted in improvements focused within the outpatient care setting. Participants in primary care and accountable care organization (ACO) models made a number of financial investments to transform their care delivery practices that may be yet realized in later years, potentially after the models end. These included investments in staffing for care coordination, redesigning clinical workflows, health IT, data analytics, engaging health care providers with financial incentives, and beneficiary engagement through care management and preventive services. However, some model participants noted that the model incentives payments were not always viewed as large enough to compensate for the amount of work required in the model. The lack of strong relationships between model participants and specialty providers, hospitals, and post-acute care settings may have contributed to the limited effects on those settings and downstream expenditures, both gross and net. Additionally, three model design factors created methodological issues for these models in demonstrating more robust results. These include the voluntary nature of all of the models, small effect sizes that may require longer time window to observe changes, and sample sizes that were either too small to be powered to detect meaningful changes or too large to produce savings when accounting for financial incentives.

Financial Incentives

Four models in this groups (Comprehensive Primary Care [CPC], Comprehensive Primary Care Plus [CPC+], Independence at Home [IAH], Million Hearts® [MH]) largely focused on prevention of greater health complications and spending solely or predominately through primary care physicians. As the fee-for-service (FFS) system does not currently have widely-used mechanisms to pay for enhanced primary care, model participants received an incentive payment to cover the costs of up-front investments in additional staffing, health IT infrastructure, and workflow changes needed to achieve the model's goals. Some model participants noted that model incentives payments were not always large enough to compensate for the amount of work required in the model. Similarly, the Financial Alignment Initiative (FAI) Washington Model incents care management entities to coordinate care for dually-eligible individuals by addressing data and workflow concerns that prevent successful integration of Medicare and Medicaid services. The FAI Washington Model used a shared savings arrangement with the state that also included care management fees to health homes to provide coordinated care.

Five ACO models in this grouping (Advance Payment [AP] ACO, ACO Investment Model [AIM], Pioneer ACO Model, Next Generation ACO [NGACO], Vermont [VT] Medicare ACO Initiative) used risk arrangements to incent better coordinated care across delivery settings for Medicare FFS beneficiaries. Accountable care organizations in this grouping participated in upside-only shared savings as well as two-sided shared savings/shared losses. ACOs typically include individual clinicians, physician practices, hospitals, skilled nursing facilities (SNFs) and other health care providers. ACOs use various business arrangements with their health care providers that may include financial or non-financial incentives such as the opportunity to receive shared savings distributions or data-based performance feedback. These arrangements provide the impetus for improved communication and care coordination among health care providers serving ACO beneficiaries. Some of the models also included prospective or advance incentive payments designed to facilitate investment in infrastructure, care management, and other ACO activities. Similar to the primary care models, participants often invested in staffing (care coordinators), redesigning clinical workflows, health information technology, data analytics, engaging health care providers with financial and non-financial incentives, and beneficiary engagement through care management and

preventive services. In theory, such enhancements should result in accumulations in improved beneficiary health over time. However, the panels of health care providers participating in the ACO, and their aligned beneficiary populations, often change from year to year. As a result, evaluation impact estimates will not reflect reductions in Medicare spending or measured improvements in quality that would otherwise have been realized in the absence of this turnover. While the magnitude of this difference is unknown, beneficiary turnover (driven by provider turnover) in the Pioneer ACO Model over the first three years was 70%. Historically, ACOs have experienced barriers to effective care delivery transformation, either because of challenges engaging specialty care physicians or from incomplete access to patient data which would allow more sophisticated care management interventions. However, many participant ACOs earned shared savings payments.

Two models in this grouping involved health plans (Part D Enhanced Medication Therapy Management [Part D Enhanced MTM], Medicare Advantage [MA] Value-Based Insurance Design [VBID]) where the model intervention occurs through plan networks, although each of these models targeted different patient populations using different incentives. The Part D Enhanced MTM model serves FFS beneficiaries enrolled in participating Part D plans. It provides prospective payments for implementation of the Part D Enhanced MTM model and performance-based payments[110] contingent on reductions in Medicare expenditures of at least 2% relative to a benchmark. In MA VBID, there are no financial incentives from CMS for model participants (beyond those already existing in the MA program) although the participants may offer financial incentives to beneficiaries to encourage them to use high-value care. Examples include reduced copays for hypertension medications or reduced primary care cost-sharing for beneficiaries who agree to participate in care management programs.

Model Design and Evaluation Issues

Notably, all of the models in this grouping had voluntary participation and most had non-randomized designs. Voluntary models suffer from potential selection bias in terms of which participants choose to participate, making it more challenging for the model evaluation to construct a comparison group of similar health care providers and beneficiaries to create unbiased impact estimates. Self-selection also means that participants can choose when to exit a model and often do so when it is in the participants', and not necessarily Medicare's, best financial interest. These concerns could be alleviated through the use of more mandatory and/or randomized models. Using a randomized design in a voluntary model alleviates selection bias within the evaluation design although it leaves open the self-selection issue more broadly for CMS. A mandatory model with random assignment to the treatment and control group would ensure that all types of organizations participate, enhancing the generalizability of the findings and ensuring that the health care system as a whole is being transformed by CMS Innovation Center's initiatives. Where mandatory models are not feasible, the use of randomization in a voluntary model, such as the Million Hearts® model, greatly improves the internal validity of the model's design and creates more confidence in evaluation estimates. However, the benefits to using randomization or mandatory participation need to be considered in light of the challenges and complexity involved with operating these models. Models also need to be tested in the methodology that would ultimately be used for expansion if that becomes

---

[110] The performance payment is distributed as a fixed $2 PBPM amount in the form of an increase in Medicare's contribution to the Plan Benefit Package's Part D premium (i.e., an increase in the direct subsidy component of Part D payment), thus decreasing the plan premium paid by beneficiaries.

an option; if a model is created with the intention to expand on a voluntary basis, it needs to be tested under that scenario for the evaluation estimates to be useful for expansion decisions.

These concerns coupled with the small magnitude in results (effect size) that is typically found among prevention-based interventions make it difficult to show improvements in outcomes, particularly in the short-term. These types of interventions require longer time windows for chronic disease to be effectively managed. For example, the CPC Initiative (a four-year model that preceded CPC+) showed no changes in gross or net expenditures and only significantly improved ED visits (-2%) with more marginal changes in inpatient admissions. However, in a study on the longer-term effects of CPC[111], the greatest relative declines in hospitalizations were found in the two years after the model ended and when most practices were in CPC+. Hospitalizations declined by 3.1% and 3.5% in Years 5 and 6 compared with 1.7% in Year 1. The effect on hospitalizations grew over time, with stronger effects demonstrated after the model's observation window. Alternatively, two models (MH, Part D Enhanced MTM) showed increased inpatient admissions and/or ED visits, suggesting that the screenings and disease management activities could have created more engagement with the health care system, prompting more visits in the short-term as beneficiaries became more aware of their symptoms and conditions.

The size of the prevention model is also important to consider, as the small anticipated effect sizes from these types of interventions require a larger sample size for analyses to be adequately powered. For example, while IAH may be serving a sicker population than the other models in this grouping, it is underpowered to be able to detect significant, reasonable changes in outcomes. Alternatively, generous model incentive payments in CPC+ and NGACO ensured robust participation but made it difficult to demonstrate net savings and ultimately resulted in significant net losses, thus far. The large size of many of the models in this group is indicative of population management models that assume responsibility for a broad panel of patients that are not necessarily touched as many are relatively healthy. However, the entire panel of beneficiaries is still included in the model and evaluation analyses, ultimately diluting any effects that may have been seen among beneficiaries that were touched by model participants.

Limitations

This analysis has a number of notable limitations worth considering in relation to the study findings. Each model's implementation period covered different amounts of time and different calendar years. Some models had few years of performance while others represented the final evaluation results after the model had ended, giving more time to produce a more robust model implementation relative to results only for the first early years of the model. Therefore, the conclusions from this synthesis are subject to change as we receive more updated findings for many of the models included. Some models were launched at the beginning of the creation of the CMS Innovation Center prior to the presence of multiple models in some markets. Usual care has evolved over the past decade and models operating today have to out-perform other initiatives also seeking to reduce readmissions and improve the quality of care, suggesting an even higher bar for these models moving forward. Most recently, care patterns may have changed due to the Public Health Emergency, which is captured for three models[112] in this analysis. Future evaluation reports may document additional changes in care delivery, such as the increased use of telehealth. While this analysis focused on a parsimonious set of outcomes in order to tell a story across models, there are a

---

[111] Available in the Appendix of the Fourth Evaluation Report for CPC+.
[112] Comprehensive ESRD Care Model, Home Health Value-Based Purchasing Model, and Medicare Care Choices Model

number of additional study measures that are important to each respective model that were not presented, creating a less complete picture of how a given model's effectiveness may be viewed. Additional measures such as preventive and professional services as well as model-specific quality measures were also summarized but not presented here, as there were not enough common measures across all of the models included in this analysis. [Appendix B](#) includes more information on quality measures of note for individual models. In general, model evaluations have a limited view of quality of care because they need to examine changes against a comparison group, for which there may be data available for model participants but not non-participants. Collecting patient-reported outcomes is important and valuable but also costly and potentially burdensome to participants and beneficiaries. The factors are typically weighed when decisions are made about the types of quality data collected and available for analyses in model evaluations. Therefore, evaluations typically examine quality measures that can be assessed by claims data only. Lastly, there are additional CMS Innovation Center models that were not included in this analysis either because they are too new, focused on Medicaid beneficiaries[113], or did not have sufficient results available.

Conclusion

Lessons learned from prior models will ensure the CMS Innovation Center can build upon past successes to design a more harmonized and streamlined portfolio of models that can drive system transformation. The review will also inform efforts to assess the impacts of CMS Innovation Center models on transformation including for beneficiaries, health care providers, and the health system more broadly. Models focused on acute care or preventing institutional care through specialty providers and targeted populations all demonstrated gross saving, relative to only half of the models focused on primary care and population management. In many cases, the acute or specialty care and targeted population models had larger effect sizes on changes in utilization and spending relative to the prevention and population management models. It is possible that the higher baseline spending of beneficiaries in the acute or specialty care and targeted population models group provided the room for which to cut costs further. Longer time windows may be needed for prevention efforts to show returns among models serving broader, healthier populations in the primary and population management models focused on prevention, which may not produce appreciable reductions in spending until after the model ends. Even with successful evaluation results and transformation efforts, models may face other barriers to national expansion. This review will guide development of future CMS Innovation Center models, including informing target participants for different types of models, quality and beneficiary experience measurement, integration of health care providers and care settings, scale of models, and approaches to evaluation to measure success through health care transformation and health equity.

---

[113] Some of these models (e.g., Financial Alignment Initiative for Medicare-Medicaid Enrollees, Washington) include beneficiaries that are dually eligible for Medicare and Medicaid. Other models (e.g., Comprehensive Primary Care Initiative, Comprehensive Primary Care Plus, Oncology Care Model, Vermont All-Payer ACO Model) are multi-payer and may include Medicaid as a payer partner and thus may assess Medicaid utilization and quality of care measures as resources allow. However, evaluation results for these models using Medicaid data were not included in this analysis.

Appendix A. Analysis

Model-specific characteristics and results

The similarities of each grouping of models are initially reviewed in the results sections to help frame the justifications for why these models were included together in a grouping. Within a grouping, results were presented qualitatively for each individual model in Appendix B (Model-specific results) within the context of the model's theory of action to familiarize the reader with each of the models at a high-level. Key model-specific results were also reviewed and some of those results may not be summarized in the cross-model tables as they were only relevant to the given model and not applicable across all models examined.

Cross-model results

Results were summarized quantitatively by specific measures across models to understand how models fit together in relation to favorable, unfavorable, and non-significant results. Impact estimates for spending, utilization, and post-acute care results for each model were provided in Tables 2-4 where available. If a measure was not available to report, the cell was left blank. Bolded estimates were statistically significant at least at the $p < 0.05$ level. The expectation was that all models would lower gross spending (and net spending, if relevant) based on reductions in costly utilization settings such as hospitalizations and other inpatient care. However, there was not always a strong hypothesis ahead of time on whether a specific measure would increase or decrease (e.g. home health utilization) and this could vary depending on the models' theory of action. Statistically significant results considered to be favorable (e.g. decreased spending) have a light green cell shading. Statistically significant results considered unfavorable (e.g. increased emergency department visits) based on a model's theory of action have red/orange cell shading. Cells with gray shading illustrate non-significant results, regardless of the direction of the point estimate, as the non-significance indicates either no change relative to a comparison group or insufficient power to detect meaningful differences. If a model touched commercial or Medicaid beneficiaries in addition to Medicare (e.g. Comprehensive Primary Care, Comprehensive Primary Care Plus, Maryland All-Payer, Vermont All-Payer Accountable Care Organization), results were only reported for the Medicare beneficiaries touched by the model. Quality of care measures (mortality and self-reported experience of care) were reported qualitatively in the narrative after the impact estimates were presented for spending and utilization measures.

Regression analyses

To majority of model evaluations used in this paper included results that were based on difference-in-differences (DiD) regressions. DiD analyses are considered the gold standard method to evaluate quasi-experiments, particularly in the absence of the ability to conduct a randomized controlled trial. These methods compare changes among participants in the intervention period relative to baseline trends and a well-matched comparison group. The only evaluation that did not use DiD regressions was the Medicare Care Choices Model [114], which used a comparison of means in the intervention period. Models with a randomized design further reduces potential sources of

---

[114] A DiD evaluation design was judged to be unviable in the fourth evaluation report for the Medicare Care Choices Model.

bias. Three[115] of the model designs included in this analysis used randomized that aided the evaluation, which is noted in the tables displaying results.

The majority of models included in this paper used voluntary participation and therefore suffer from inherent selection bias related to the participant's decision and ability to participate in the model.[116] While the DiD study design attempts to account for selection bias related to internal validity and the construction of a comparison group, it cannot correct for design issues that inhibit external validity. Mandatory models, where participation is required, do not have the same concerns related to selection bias in the applicability of evaluation results to the broader Medicare program and therefore provide stronger external validity. Two models (Comprehensive Joint Replacement Model[117] and Home Health Value-Based Purchasing Model [HHVBP]) included in this paper used mandatory participation.

In keeping with the focus on evaluation results that met the most rigor, this paper highlights estimates that are statistically significant at p<0.05 (although results that are significant at the p<0.10 are noted for policy considerations). All estimates discussed throughout the paper were statistically significant unless otherwise noted. Where available, we report cumulative estimates to date across the performance years examined (e.g. we do not report impact estimates for each performance year unless cumulative estimates were not available) and across all model participants, aside from a few exceptions where estimates are reported by track or other meaningful groupings[118] due to inherent model design features.

Impact estimates were pulled from each respective evaluation report using a similar unit of analysis, relative differences, where feasible to make results comparable across models. This format has historically been more useful to policymakers within the CMS Innovation Center than DiD regression coefficient themselves. Estimates are bolded where the differences were determined to be statistically significant at the p<0.05 level in regression analyses relative to a well-matched comparison group.

Measures

Gross and net total Medicare spending was provided for each model where applicable. Gross spending was measured as the total allowed expenditures for all Part A and B Medicare services except for the Medicare Advantage (MA) Value-Based Insurance Design Model (VBID) which measures Medicare Advantage costs to Medicare due to the model's focus on MA plans and beneficiaries. Net spending results included any model incentive payments, such as shared savings or losses or upfront infrastructure and care delivery payments, in addition to the gross spending impact. Some models, for example Maryland All-Payer, Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport, and MA VBID, did not have additional payments to account for,

---

[115] Comprehensive Joint Replacement, Home Health Value-Based Purchasing, and Million Hearts®. While randomization was used in the Medicare Care Choices Model design to assign participants to cohorts, those methods did not benefit the evaluation design.
[116] Smith, B. 2021. CMS Innovation Center at 10 Years – Progress and Lessons Learned. The New England Journal of Medicine. 384(8):759-764.
[117] The Comprehensive Joint Replacement Model has been partly voluntary during some time periods.
[118] We report Comprehensive Primary Care Plus model impact estimates for Track 1 and Track 2 practices separately. Separate models were reported for Bundled Payments for Care Improvement Initiative as results were provided separately for each and not across all models. Similarly, we report Bundled Payments for Care Improvement Advanced by hospital and PGP combinations with medical and surgical clinical episodes.

in which case only gross spending results were provided. Gross spending is reported also for HHVBP, which is a budget neutral model. The expectation is that all of these models would lower gross and net spending based on reductions in costly utilization settings such as hospitalizations and other inpatient care.

Utilization and quality measures tend to be assessed as similar concepts across evaluations, even if the specific definitions and names vary. For utilization measures, we included four commonly measured care settings that are often used to examine changes in care delivery. These included inpatient admissions, emergency department (ED) visits, post-acute care (PAC), and inpatient readmissions. For inpatient admissions, this measure is typically defined as any inpatient admissions (binary measure) that occurred within the years examined with a few expectations that measure the number of admissions per a unit of analysis such as 1,000 beneficiaries. Some evaluations call this measure unplanned acute care hospitalizations, hospital admissions, or simply hospitalizations. ED visits are also typically measured as a binary outcome of any visits in the years examined. Most evaluations define this measure as outpatient ED visits and observation stays that do not result in an admission/hospitalization although some evaluations break those two concepts a part and/or measure the number of visits that resulted in an admission (which was not included in this paper). For the inpatient readmissions measure, most model evaluations defined this measure as any 30-day readmissions although a handful assessed this based on a count per hospital discharges. The expectations are that all of these models should lower all forms of institutional care. PAC is a broader category of utilization measures that could include skilled nursing facility (SNF) visits, institutional rehabilitation facility [IRF] visits, home health [HH] visits. Most of these measures were assessed as binary indicators of the presence of any utilization in a given category but in some cases these measures were assessed as counts of the number of days of care in a given setting. Where PAC utilization measures were not available or reported for a model, PAC setting specific payment results were provided. Changes in HH expenditures and utilization were considered favorable if they increased for some models while others hypothesized decreases to remove wasteful overuse.

Quality of care is often measured using a self-reported measure of beneficiaries' experience of care and sometimes assessed using patient mortality. Experience of care, as assessed by some version of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey, is typically measured as the summation of a composite of measures such as shared decision making, access to care, affective communication, exchanging information, enabling patient self-management, and symptom or pain management. As this information does not lend itself to being summarized into one value, these results were described qualitatively in the results section. Mortality was measured in a variety of different time periods (e.g. 12 months, 18 months) and analysis types (e.g. survival analyses, DiD) across model evaluations. These results, where available, are summarized in the narrative. In keeping with the model's theory of action, we would expect experience of care to improve and mortality to decline (where relevant) or at least not to increase.

Appendix B. Model-specific results

Accountable Care Organization (ACO) Investment Model (AIM)

AIM**,** which operated from 2015-2018, distributed up-front monthly payments (pre-paid shared savings) to encourage new Medicare Shared Savings Program (SSP) ACOs to form in rural and underserved areas (AIM Test 1 ACOs) and to encourage existing Shared Savings Program ACOs to transition to greater financial risk (AIM Test 2 ACOs). Test 1 ACOs received an up-front, fixed payment, an up-front variable payment based on the number of assigned beneficiaries and a monthly payment based on the size of the ACO; Test 2 ACOs only received up-front, variable payments for the number of assigned beneficiaries and monthly payments for the size of the ACO. AIM funds were used by ACOs to hire more clinical staff, upgrade health information technology (IT), support data analysis, and perform ACO management. None of the AIM Test 2 ACOs had consistently lower or higher spending across the performance years and did not change quality of care (results are not included in the cross-model synthesis). The 41 AIM Test 1 ACOs lowered Medicare FFS spending and utilization with no decrements in quality of care. The model saved $526.4 million in Medicare spending across three performance years ($381.5 million in net savings). Many AIM ACOs largely relied on relationships with management companies to support ACO operations, share performance feedback with health care providers, and provide stability given rural workforce challenges. Most AIM ACOs opted not to assume financial risk and noted needing more time in an upside-only arrangement or greater financial incentives to remain in the Medicare Shared Savings Program. As of 2018 shared savings reconciliation, 54.2% of AIM funds had been recouped.

Advance Payment (AP) ACO Model

AP ACO, operating between 2012-2015, was designed to provide new ACOs with up-front monthly payments to invest in care coordination infrastructure. The advance payments were intended to be recouped against shared savings payments. The model, which included 36 SSP ACOs, was designed to help smaller ACOs with less capital participate in the Medicare Shared Savings Program. The model did not produce improvements in utilization or quality of care and had significant increases in Medicare fee-for-service (FFS) spending ($242 million in net losses). Of the $68 million in advance payments distributed to the AP ACOs, $30 million had not been recouped against shared savings by the end of the three-year participation agreement period. The majority of the advance payment funding to ACOs went to personnel (care management, office staff, ACO leadership, health care providers serving ACO roles such as medical directors) and benefits costs such as IT infrastructure to build analytic capacity, enabling communication across health care providers, and creating tools to support care management. Many participating organizations had little to no experience working with and analyzing claims data and incorporating population-level data into their care planning or management; some were also newly formed and had no history working as a single entity, potentially creating barriers to implementation. Advance payments recouped against shared savings were used in AP ACO to motivate participation in the Medicare Shared Savings Program. About 47% of these ACOs were able to repay their advance payments by the end of 2015[119] and 15 of the 36 participating ACOs earned shared savings. Many of the AP ACOs cited inadequate IT as a barrier to making practice changes that depended on improved population health data and needing more time in an

---

[119] Two additional ACOs repaid an additional $5.5 million through earned shared savings in subsequent years.

upside-only arrangement or greater financial incentives to remain in the Medicare Shared Savings Program.

Bundled Payments for Care Improvement (BPCI) Initiative

BPCI, which occurred from 2013-2018, coordinated care across a wide range of health care providers and settings (hospitals, Physician Group Practices [PGP], Skilled Nursing Facilities [SNFs], Home Health Agencies [HHAs]); and conditions or procedures with 48 clinical episodes in total across 4 different model designs[120] with the goal of reducing expenditures and maintaining or improving quality. Models 2 and 3 consisted of retrospective bundled payment arrangements where actual expenditures were reconciled against a target price for an episode of care. In Model 2, the episode included the inpatient stay in an acute care hospital plus the post-acute care (PAC) and all related services up to 90 days post-hospital discharge. In Model 3, the episode of care was triggered by an acute care hospital stay but began at initiation of post-acute care services with a skilled nursing facility, inpatient rehabilitation facility, long-term care hospital, or home health agency, and lasting up to 90 days after episode initiation. In Model 4, CMS made a single, prospectively determined bundled payment to the hospital that encompassed all services furnished by the hospital, physicians, and other practitioners during the episode of care, which lasted the entire inpatient stay and included related readmissions for 30 days after hospital discharge. Physicians and other practitioners had the option to submit "no-pay" claims to Medicare and be paid by the hospital out of the bundled payment. In total the model included 1.4 million episodes across Models 2-4. BPCI was designed to achieve savings to Medicare ranging from 2% to 3.25% of the baseline expenditures that averaged from $15,000 to $58,000 in the baseline. Participants with episodes below the target price were eligible to receive the difference in reconciliation payments, which summed to $1,611 million for Model 2 participants and $342 million for Model 3 participants. Despite gross reductions in FFS payments in BPCI Model 2 ($1,193 million) and 3 ($232 million), Medicare experienced net losses of $418 million for Model 2 and $110 million for Model 3 after accounting for reconciliation payments to model participants. Model 4 had non-significant changes in spending. BPCI Models reduced unnecessary PAC use for many of the episodes examined but generally did not improve emergency department (ED) visits and most Models had non-significant changes in readmissions aside from Model 3 HHAs that had significant reductions. While self-reported beneficiary care experience declined slightly in Model 2, self-reported functional outcomes were not affected. There were also no changes in mortality overall across all BPCI models.

Bundled Payments for Care Improvement Advanced (BPCI-A)

BPCI-A, built from lessons learned in BPCI, began in 2018 and only has evaluation results through its second performance year to date. The model tests whether linking payments across health care providers for an episode of care can reduce expenditures while maintaining or improving quality. Episode payments are compared to a risk-adjusted target price and participants can earn a reconciliation payment if episode payments are below their target price, after considering quality of care. In Model Year 2, BPCI-A included 13 medical and 19 surgical clinical episodes that hospitals and PGPs can trigger. Episodes are initiated with a hospitalization or outpatient procedure through 90

---

[120] Model 1, which included all inpatient stays, was substantially different from the other BPCI model designs and did not produce any statistically significant changes in the majority of health outcomes or in total expenditures. Therefore, results from Model 1 were not included in the synthesis.

days post-discharge or the end of the procedure. In BPCI-A, benchmark prices are discounted 3% of the baseline expenditures that averaged $25,000 across all episodes, which is intended to be Medicare savings under the model. As of 2019, $616.5 million in reconciliation payments have been distributed to participants. Despite gross reductions in FFS payments across all episode types ($551 million), net results for BPCI-A were mixed, with surgical episodes showing promising net savings ($204 million) and medical episodes demonstrating significant losses ($275 million). The model resulted in relatively small, statistically non-significant net losses model-wide for the first two performance years ($65.5 million), although recent changes to the model design and benchmarks should produce more favorable outcomes in later model years. Overall, both surgical and medical clinical episodes had improvements in PAC. BPCI-A reduced readmissions for surgical episodes during the 90 days following a discharge or procedure by 4.1% of the BPCI-A mean for Model Years 1 and 2 (2018-2019). Estimates were similar by episode initiator type, though only the PGP estimate was statistically significant. As of 2019, there are no significant overall changes in mortality, self-reported functional status, care experience, or satisfaction with care.

Comprehensive ESRD Care (CEC) Model

CEC, which operated from October 2015 through March 2021, created End-stage Renal Disease (ESRD) Seamless Care Organizations (ESCOs) comprised of dialysis facilities, nephrologists, and other health care providers, that were financially accountable for the quality and cost of care for their attributed beneficiaries with ESRD. Evaluation results are available through 2020. Seven dialysis organizations participated through 37 ESCOs reaching 13% of Medicare beneficiaries with ESRD. CEC participants expanded access to dialysis care, provided more consistent education to patients and caregivers, implemented ED notifications to follow-up with beneficiaries after an acute event, and assisted in medication management, especially following a hospitalization through intense care coordination and interdisciplinary team discussions. CEC reduced gross Medicare FFS spending by reducing the number of hospitalizations, readmissions, and post-acute care for aligned beneficiaries, while improving a number of model-specific quality of care measures as well as mortality. Despite these improvements, the model resulted in non-significant aggregate net losses of $46 million (through the fourth performance year) after accounting for shared savings payments made to ESCOs.

Comprehensive Joint Replacement (CJR) Model

CJR is a mandatory model[121] testing whether an episode-based payment approach for lower extremity joint replacements (LEJR) can lower payments while maintaining or improving quality that has been operating since 2016 with evaluation results available through 2019.[122] In CJR, hospitals are held financially accountable for lowering costs for LEJRs and providing improved quality of care. Actual episode payments are compared to the hospital's quality-adjusted target price and hospitals can receive an extra payment if their episode payment is below their target price. CJR participating hospitals were subject to a target price for LEJR episodes that was $1,511 lower on average than payments to control group hospitals (or a 5.2% reduction from baseline). Hospitals with spending below the target price had the opportunity to receive reconciliation (incentive) payments to invest in care redesign and coordination and increase efficiency and quality of care provided to patients. However, in 2019, only half of mandatory hospitals received these payments which totaled to $126.1

---

[121] CJR has been partly voluntary during some time periods.
[122] The model was extended through rulemaking through 2024, representing eight performance years.

million. Hospitals used strategies to influence care during the episode such as improving guidelines or directives to consider when determining whether to perform a LEJR, providing performance feedback reports or data to surgeons to modify their care practices, and coordinated with SNFs on patient care. These efforts resulted in $251.8 million gross Medicare FFS savings ($76 million in net savings was not statistically significant) with improvements in readmissions and PAC as well as model-specific quality measures such as complication rates and unplanned readmissions. However, patients with hip fractures self-reported poorer outcomes.

Comprehensive Primary Care (CPC) Initiative

CPC, a multi-payer initiative operating between 2012-2016, was aimed at strengthening primary care using population-based care management fees and shared savings as well as a set of requirements for practices to provide comprehensive care.[123] Care management fees for Medicare FFS attributed beneficiaries averaged $20 per beneficiary per month (PBPM) for the first two years and $15 PBPM for the last two years. In CPC, CMS care management fees to practices accounted for between 10-20% of practice revenue and averaged $175,000 per practice ($50,000 per clinician) each year. Participants in CPC could also earn shared savings based on regions meeting performance targets. However, a few practices reported that Medicare's shared savings methodology held them responsible for reducing costs incurred by specialists or hospitals that they felt were outside of their control. The model, which included nearly 500 primary care practices across seven regions in the United States, had modest effects on hospitalizations and ED visits that resulted in non-significant changes to gross expenditures, which would not have been large enough to cover Medicare's financial incentives totaling $479.1 million over the entire initiative. The model had little impact on claims-based quality of care and beneficiaries' experience of care and no impact on mortality. Practices engaged in substantial, challenging transformation and improved how they delivered care through risk-stratified care management, expanded access to care, and continuous improvement driven by data. However, they also faced barriers to change, including burden associated with quality monitoring and reporting, existing FFS incentives that encourage volume, and lack of infrastructure for comprehensive and efficient health information exchange between health care providers.

Comprehensive Primary Care Plus (CPC+)

CPC+, which began in 2017 and ended in 2021, was built off of CPC and tested similar concepts[124] but involved two separate participation Tracks based on the level of requirements and model payments to gradually shift participating practices from FFS towards population-based payments. Instead of using shared savings to reward practices for cost and quality performance, participating CPC+ practices received a prospective performance-based incentive payment that was paid at the beginning of the year and was partially or fully recouped at the end of the year, depending on their performance. CPC+ was much larger in size relative to CPC with 18 regions and 3,070 practices. CPC+ reduced acute care utilization such as outpatient emergency department visits, hospitalizations, and acute care expenditures as well as improved some claims-based quality of care measures. The large size of the CPC+ model resulted in net losses to Medicare when accounting for financial incentives ($14 and $25 PBPM for Track 1 and Track 2 practices, respectively as of 2020). Practices increased use of on-site behaviorists each year of CPC+. This investment in behavioral

---

[123] There were five care delivery functions: 1) access and continuity, planned care for chronic conditions and preventive care, 3) risk-stratified care management, 4) patient and caregiver engagement, and 5) coordination of care across the medical neighborhood.

[124] CPC+ is testing whether multipayer payment reform, actionable data feedback, robust learning supports, and health information technology vendor support enables primary care practices to transform how they deliver care and improve patient outcomes. CPC+ requires practices to transform across five care delivery functions: (1) access and continuity, (2) care management, (3) comprehensiveness and coordination, (4) patient and caregiver engagement, and (5) planned care and population health.

health integration was particularly valuable in 2020 as the Public Health Emergency (COVID-19) increased mental health care demand. Further, between 2016 and 2020, long-term opioid use and potential overuse decreased among CPC+ beneficiaries; the population included people who had a disability and were dually eligible for Medicare and Medicaid.

Financial Alignment Initiative (FAI) for Medicare-Medicaid Enrollees, Washington

The FAI demonstration aims to provide full benefit individuals dually enrolled in Medicare and Medicaid with better care experience and better align the financial incentivizes across the two programs through either a capitated model or a managed FFS (MFFS) model. Of the twelve FAI demonstrations with Medicare cost estimates, one shows reductions, six show increases, and five show neither increases nor decreases in Medicare expenditures. Service utilization results are mixed overall; however, six demonstrations increased the number of physician visits. Washington has the most demonstration years of data available at this time and that, coupled with its successful results, led to its inclusion in this paper. Many of the other FAI demonstrations continue to be evaluated. FAI Washington (WA), which uses the MFFS model and adds care coordination as a Medicaid-covered benefit, launched in 2013 and has been extended through 2023. The model targets high-cost, high-risk beneficiaries and leverages health homes for care coordination. Health homes contract with the state to provide comprehensive care coordination services and receive a per-member per-month Medicaid payment for engaged enrollees. This is a shared savings initiative where the state can earn savings as a result of the demonstration. The state pays health homes for delivery of health home services per member per month (composite rate across three payment tiers: $256 as of 2020) for enrolled beneficiaries for care coordination across their networks. These rates have increased over time and agencies report they can now operate without taking losses, although other agencies have noted the amount allocated for administrative costs remains too low to cover overhead. Health homes collaborate with community-based organizations, Area Agencies on Aging, and managed care organizations. They also established a network of care coordination organizations representing health care providers of primary care, mental health services, long-term services and supports, chemical dependency services, and specialty providers. The demonstration has resulted in shared savings distributed to the state in each year, ranging from $11 to $18 million over the six years. The demonstration has resulted in $385 million in gross Medicare savings and $297 million in net savings to Medicare through the sixth demonstration year. In the most recent years, the demonstration has reduced SNF admissions and long-stay nursing facility use although has not significantly changed inpatient admissions or ED visits.

Home Health Value-Based Purchasing (HHVBP) Model

The original HHVBP model was a randomized mandatory model implemented in nine states beginning 2016, with evaluation results through 2020 included in this paper. The model is currently being expanded nationwide[125]. HH agencies in the original HHVBP model were able to earn financial incentives for quality improvement based on performance relative to other agencies within their state. HH agencies used data analytics, staff training, and frontloading practices (scheduling more skilled nursing visits early in an episode of care) to improve performance and care delivery that resulted in reductions in hospitalizations and SNF use which were important drivers of reductions in Medicare

---

[125] In November 2021, CMS finalized the expansion of the HHVBP Model nationally starting in January 2022 and ended the original HHVBP Model one year early with the CY 2021 payment year (i.e., in 2019).

FFS spending, resulting in $949 million in aggregate savings. Although three of five measures of patient experience declined, the model significantly reduced mortality and improved beneficiaries' functional status on a number of quality measures.

Independence at Home (IAH) Demonstration

IAH is a Congressionally mandated demonstration that has been operating since 2012 testing whether a payment incentive for home-based primary care reduces health care expenditures and improves quality of care for chronically ill and functionally limited beneficiaries. Practices can earn incentive payments if their patients' Medicare expenditures are below the practice's target expenditures and the practice meets required standards for a set of quality measures. IAH sites have tried to reduce hospital use by making care more comprehensive and responsive to patient's needs and by providing follow-up contacts for patients within 48 hours of a hospital discharge or ED visit. In its first five years, the demonstration has been able to reduce ED visits and avoidable hospital admissions in some demonstration years but total admissions in only one year. Although sites have trended towards reductions in total spending, the effect size and small number of participants have not been large enough to demonstrate significant savings. Coupled with the $37 million in incentive payments to IAH sites, the demonstration has resulted in $44.3 million in losses to the Medicare program. In 2019, total incentive payments were $11.1 million distributed to 12 demonstration sites serving an average of 580 enrollees per site. IAH practice attrition in later years of the demonstration was in part motivated by the perceived onerous participation requirements.

Maryland (MD) All-Payer Model

The MD All-Payer Model, operating from 2014-2018, used global budgets and all-payer rate setting for hospital services with hospitals accountable for total hospital cost of care and quality. This model included all patients hospitalized within all 46 general acute care hospitals in the state. With the state's historically high Medicare FFS expenditures, the use of the unique waiver and rate setting allowed for guaranteed reductions in spending. There are no financial incentives from CMS to participate in this model. Nearly all hospitals invested in care coordination, discharge planning, social work staffing, patient care transition programs, and systematic use of patient care plans in response to the model. The multi-payer model resulted in $975 million in Medicare savings, largely driven by slower growth in hospital expenditures ($796 million in savings). The model had significant reductions in inpatient admissions, potentially avoidable hospitalizations, PAC spending, and mortality but there were no significant changes to ED visits or unplanned readmissions.

Medicare Advantage (MA) Value-Based Insurance Design (VBID) Model

The MA VBID model, which started in 2017 and ends in 2024, grants a limited waiver of the Medicare Advantage (MA) uniformity requirement, allowing MA plans to structure cost-sharing and other benefit design elements to encourage beneficiaries with chronic conditions to use high-value care. Participating MA plans can offer reduced cost sharing for high-value services and additional supplemental benefits and can require disease management of other activities as a condition of receiving the benefits with the goal of improving beneficiary health, care quality, and saving money for plans and Medicare. There are no financial incentives from CMS to participate in this model. In the first three years, 11 Parent Organizations served over 100,000 MA beneficiaries with conditions such coronary artery disease, congestive heart failure, chronic obstructive pulmonary disease, diabetes, and hypertension. MA VBID increased the use of most services targeted, such as diabetes monitoring, visits to health care providers, and drug fills for select prescriptions. MA VBID has not yet led to changes in costs to Medicare or plans but is also not costing Medicare additional money.

Substantial changes to the model and participation began in 2021, including but not limited to supplemental benefit targeting by socioeconomic status, and including the Medicare hospice benefit as a covered benefit within the MA benefits package.

Medicare Care Choices Model (MCCM)

MCCM[126], operating from 2016-2021 with results available through 2020, tested whether offering eligible beneficiaries[127] supportive services[128] without forgoing treatment for their terminal conditions would improve quality of life and care and reduce Medicare FFS expenditures. MCCM participants, which included hospices, palliative care providers such as nurses, care coordinators, social workers, clergy, and bereavement counselors, received a fixed monthly payment to provide care coordination and supportive services similar to those provided under the Medicare hospice benefit. The model provided participants with a payment (typically $400 per beneficiary per month, averaging $1,827 per enrollee to date totaling $8.4 million) for enrolled beneficiaries. Despite low initial participation (141 hospices) and a large number of drop-outs (49 hospices remained in the model as of 2021) that prevented the ability to use the randomized model design in the evaluation, MCCM significantly reduced Medicare expenditures ($33.2 million net savings) by reducing hospital services (inpatient admissions, ED visits, readmissions, and PAC) while increasing beneficiaries' use of the Medicare hospice benefit. MCCM beneficiaries were significantly more likely to elect the Medicare hospice benefit and elect the benefit earlier relative to the comparison group. Beneficiaries in MCCM typically remained with the same care team once they elected the hospice benefit, which helped ease the transition to hospice. Caregivers of MCCM enrollees who transitioned to hospice reported highly positive experiences in the model; caregivers of MCCM enrollees who did not transition held less positive views of the model as did caregivers of those enrolled less than 30 days.

Million Hearts® (MH): Cardiovascular Disease Risk Reduction Model

MH Cardiovascular Disease (CVD) Risk Reduction Model, was a randomized controlled trial that operated from 2017-2021, providing incentives to health care providers to reduce CVD risk and lower the incidence of first-time heart attack and stroke among Medicare beneficiaries ages 40-79 who have not had a previous heart attack or stroke. Participating health care providers, including 173 practices and 273,730 beneficiaries in the intervention group, used a standardized risk assessment tool to calculate patient's risk of heart attack or stroke within 10 years and received financial incentives to reduce CVD risk among high-risk beneficiaries. Organizations in the MH intervention group received payments for risk stratifying eligible Medicare beneficiaries, providing cardiovascular care management to beneficiaries at high risk of having a heart attack or stroke during the first year of the model, and then for reducing risk of those beneficiaries in later years. Average payments ranged from a high of $15,251 in the second 6-month performance period, to a low of $3,478 in the seventh 6-month performance period. Over the first four years, the model has not resulted in changes to gross or net expenditures, even though there were significant increases in inpatient admissions and ED visits. There was also a small but significant mortality benefit for beneficiaries in the intervention group. It's possible the model made beneficiaries more aware of worrisome symptoms and more

---

[126] While randomization was used in the MCCM model design to assign participants to cohorts, those methods did not benefit the evaluation design.

[127] Model enrollees were Medicare FFS beneficiaries at the end of life (expected to live less than six months) with a diagnosis of cancer, congestive heart failure, chronic obstructive pulmonary disorder, or HIV/AIDS.

[128] These services included care coordination and case management, round-the-clock access to health care professionals, patient- and family-centered care planning, shared decision making, symptom management, and counseling.

engaged with the health care system, promoting more utilization. CVD risk scores decreased by a small amount among the intervention group driven by reductions in blood pressure and cholesterol. Beneficiaries in the intervention group were more likely to start or intensify statin or anti-hypertensive therapy. Over the first four years, CMS paid $7.6 million to participants that reduced CVD risk scores, although it has not resulted in a reduction of the rate of first-time heart attack or stroke. Nearly all of the MH organizations interviewed stated they did not think the payments covered the costs of implementing the model requirements. Organizations that stayed in the model did so because they were more motivated by the opportunity to improve patients' health outcomes or because the model aligned with other quality incentive programs. Among organizations that withdrew early from the model, the most common reasons for leaving included not perceiving the financial incentives as commensurate with the work required and not having adequate staff to comply with model requirements, particularly uploading the data to the registry.

Next Generation ACO (NGACO) Model

NGACO, which began in 2016 and ended in 2021, is a voluntary Medicare ACO model with ACOs more experienced in coordinating care and managing financial risk. The model's features included two-sided risk (80% or 100%); quality reporting; payment mechanisms designed to facilitate infrastructure improvement; and benefit enhancements for flexibility in care delivery and beneficiary engagement. Its goal was to test whether strong financial incentives and tools to support better patient engagement and care management could improve health outcomes and lower costs for FFS beneficiaries. Organizationally, NGACOs in the model were organized, around physician practices; an integrated delivery system (IDS) or hospital system; or as a partnership between a physician organization and a hospital or hospitals. NGACOs responded to the model's incentives by: 1) investing in improved data analytic capacity to manage prospectively aligned populations; 2) engaging beneficiaries through care management and annual wellness visits; 3) engaging physicians using financial and non-financial incentives; and 4) developing stronger ties with select SNFs to improve delivery of PAC. Collectively, the effect of these changes were associated with $667 million (p<0.01) in lower gross Medicare spending but an increase of $243 million (p<0.1) in net spending by the fourth performance year (2019). Most declines in cumulative spending were from PAC services, acute hospital spending, professional services, and hospice[129]. The model was not associated with changes in quality of care as measured by hospital readmissions, ambulatory care sensitive admissions or readmissions following a SNF stay. Declines in spending associated with the model increased over time from -$134 PBPY (2016) to -$258 PBPY (2019) reflecting exit by poorer performing NGACOs and improvement among NGACOs remaining. Spending effects varied by a NGACO's organization type; risk and payment mechanism; and market characteristics. Physician practice affiliated NGACOs reduced acute care hospital and outpatient facility spending, while hospital affiliated NGACOs tended to reduce professional spending. NGACOs choosing 100% risk as well as entities electing population-based payment mechanisms were associated with larger average spending reductions relative to entities electing lower risk levels or those electing FFS with or without infrastructure payments, respectively.

---

[129] The spending declines noted here for hospice services reflects smaller relative increases in the NGACO population's hospice spending compared to hospice spending in the comparison group over time.

NGACOs located in markets with higher per capita Medicare expenditures achieved higher spending reductions, as they had greater opportunities to improve efficiency.

Oncology Care Model (OCM)

OCM, which began in 2016[130] with evaluation results through 2019[131] presented in this paper, was designed to improve care coordination and access to care for Medicare FFS beneficiaries receiving chemotherapy treatment. Oncology practices (190 originally down to 176 by 2019) received a $160 per beneficiary monthly payment and the potential to earn performance-based payments for meeting quality and cost goals. Practices transformed cancer care to be more person-centered and standardized care pathways that benefited all patients. The model had an impact on gross spending[132], driven by higher-risk episodes, but saw increased net spending (-$377.1 million through the fifth performance period)[133] driven by incentive payments. There were no significant changes in inpatient admissions, ED visits, readmissions, PAC, mortality, or experience of care. OCM lead to higher-value (more cost-conscious) use of Part B non-chemotherapy drugs, which are supportive care drugs, but there are no signs that OCM is driving value-oriented chemotherapy or radiation treatment.

Part D Enhanced Medication Therapy Management (Part D Enhanced MTM) Model

Part D Enhanced MTM, which began in 2017 and concluded in 2021, tested whether modifications to traditional medication therapy management (MTM) requirements incentivize Part D sponsors to right-size their MTM services, leading to improved therapeutic outcomes and reducing net Medicare expenditures. The model provided Part D stand-alone prescription drug plans with additional flexibilities and financial incentives not present under traditional MTM. This included flexibility to target enrollees and offer services tailored to their enrollee characteristics, prospective payments to support implementation of interventions, performance-based payments in the form of a premium subsidy for reducing Medicare costs relative to a benchmark, and new eligibility and encounter data reporting requirements. Six Part D prescription drug plans representing 22 plan benefit packages with 1.9 million enrollees participated in this model. By year three, the model did not have significant impacts on gross or net Medicare expenditures. Reductions in inpatient institutional PAC spending were partially offset by increases in ED, outpatient, and ancillary service spending. Inpatient expenditures related to ambulatory care sensitive conditions decreased as did readmissions. There were modest improvements in medication use for diabetes, but measures of potentially unsafe medication use did not improve. Through 2019, CMS paid sponsors an average of about $70.5 million to cover the anticipated model implementation costs ($3.5 PBPM for most sponsors) and an average of about $22.7 million in performance-based payments each year ($1.1 PBPM). These payments have exceeded gross savings as Medicare gross savings have been relatively small and not statistically significant[134].

---

[130] OCM is anticipated to end in June, 2022.
[131] Evaluation results for spending through 2020 (sixth performance period) are available in a more recent evaluation report, although that report does not include the other utilization measures examined in this synthesis. Results for spending outcomes are largely the same in the more recent report as are presented in this paper.
[132] Gross spending results in the sixth evaluation report were consistent with those in the fifth report.
[133] Results from the sixth evaluation report including episodes that began between July 1, 2016 and January 1, 2019, all of which had ended by June 30, 2019
[134] Evaluation results through Model Year 4 are now available. They continue to show no significant cumulative impact on gross Medicare expenditures.

Pioneer ACO Model

The Pioneer ACO Model, which began in 2012 with 32 ACOs, was designed for health care organizations and health care providers already experienced in coordinating care for Medicare FFS patients across care settings. The model used a shared saving payment policy during the first two years with higher levels of savings and risk relative to the Medicare Shared Savings Program, which is what is reflected in the evaluation results included in this paper. In Pioneer ACO Model, 10 ACOs produced savings in the first two performance years (ranging from as high as $30 million to as low as $3 million), 10 ACOs had savings in at least one year (ranging from as high as $22 million to as low as $5 million), although two ACOs had significant losses ($33 million and $8.5 million), and 12 had no significant savings. The model saved $384 million over the first two performance years ($254 million in net savings), driven by reductions in inpatient admissions and PAC utilization. The Pioneer ACO Model, as tested during the first two performance years of the model, was certified in 2015 by the CMS Office of the Actuary; it was the first model that met the expansion criteria[135]. Elements of Pioneer ACO Model were subsequently incorporated into the Medicare Shared Savings Program through rulemaking. Many ACOs experienced barriers to health care provider engagement as shared savings were too small (or non-existent) and not immediate enough to motivate physicians, particularly PCPs, who work with other payers that provide more immediate, direct, and larger incentives than in the model. Similarly, ACO leaders reported difficulty engaging specialists in the model, as many specialists did not see a role for themselves in a model of care delivery focused on population health management and specialists believe they stand to lose more in revenue than they can gain in shared savings because specialty care is often the target of utilization management.

Prior Authorization of Repetitive, Scheduled Non-Emergent Ambulance Transport (RSNAT)

RSNAT, which operated between 2014-2020, was a voluntary model in nine states that aimed to reduce improper use of RSNAT services[136]. Ambulance suppliers in these states had to obtain prior authorization for these services or the resulting FFS claims would be subject to prepayment review. Among beneficiaries with end-stage renal disease (ESRD) and/or stages 3-4 pressure ulcers who need these services, RSNAT resulted in approximately $1 billion in savings to the Medicare Trust Fund. These savings were almost entirely due to reductions in use of non-emergent ambulance transports. The model had little to no impact on quality or access to care, and did not increase mortality. The model was certified for national expansion[137].

Vermont (VT) All-Payer ACO Model

VT All-Payer ACO Model, which began in 2017 and is scheduled to end in 2022, tests whether scaling an ACO program across all payers (Medicare, Medicaid, commercial) in the state can reduce expenditures while preserving or improving quality by generating sufficient incentives and alignment across payers for broad delivery system transformation. The model builds on nearly two decades of primary care and population health investments in Vermont, strong regulatory oversight, and a statewide culture of reform. The Vermont Medicare ACO Initiative (ACO model) was built off of

---

[135] https://wayback.archive-it.org/3926/20170127185820/https://www.hhs.gov/about/news/2015/05/04/affordable-care-act-payment-model-saves-more-than-384-million-in-two-years-meets-criteria-for-first-ever-expansion.html

[136] Transports for medically necessary, scheduled, non-emergency ambulance transportation for three or more round trips in a 10-day period or at least once a week for three weeks.

[137] Even though the model continued through 2020 under Innovation Center authority, CMS decided not to evaluate the model beyond 2019 because the agency had adequate evidence to support an expansion decision. The model transitioned to Section 515 of the Medicare Access and CHIP Reauthorization Act of 2015 authority as of December 2020 and is in the process of being expanded nationwide.

the NGACO methodology that had been tailored for the state with one ACO currently participating. The evaluation captures the Medicare ACO-level results, which reflect the impact of the all-payer ACO framework beyond payer-specific ACO models that operated previous as well as results for state-wide changes from payment and delivery system reform initiatives in the Vermont All-Payer Model (VTAPM). In the first two years, the Vermont Medicare ACO Initiative achieved gross savings of $56.85 million ($p<0.1$) with non-significant net savings of $48.91 million after accounting for shared savings and other pass-through payments. The state-level analyses reveal gross savings of $127.52 million ($p<0.01$) and a cumulative net impact of $122 million ($p<0.01$). Net shared savings payments[138] totaled $7.94 million in the first two performance years. These results were likely driven by declines in acute care stays (at the ACO and state level) and in readmissions at the state level.

---

[138] These calculations took into account shared savings payments to VT All-Payer ACO health care providers in the baseline and comparison health care providers in the baseline and performance periods.

Appendix Table 1. Acute or Specialty Care and Targeted Population Medicare Models

Model	Years covered	Intervention	Payment	Health care providers	Beneficiaries	Size
BPCI Models 2-4	October 2013-September 2018	Coordinating care across multiple types of health care providers involved in episodic care that is initiated with a hospitalization through 30, 60, 90 days after discharge or the end of the procedure.	30, 60, or 90-day episode following a hospitalization or initiation of PAC with a discount to target price tied to quality	Hospitals, PGPs, SNFs, HHAs, IRFs, LTCH	FFS acute: 48 clinical episodes across all three payment models (Model 2 had 32 hospital episodes and 21 PGP episodes. Model 3 had 11 SNF and 3 HHA episodes. Model 4 had 2 hospital episodes). Baseline total annual Medicare expenditures ranged from $15,000 to $58,000.	1.4 million total BPCI episodes initiated across Models 2-4. Model 2 hospital episodes accounted for 46% of that total; PGP episodes accounted for 42.3%. Model 3 SNF episodes, accounted for 6.5%; HHA episodes accounted for 1.9%; PGP, LTCH, and IRF episodes accounted for 2.3%. Model 4 hospital episodes accounting for 1.1%. Across the five years of the model, 423 hospitals and 272 PGPs participated in Model 2; 873 SNFs and 117 HHAs participated in Model 3; 23 hospitals participated in Model 4, however only 2 hospitals remained in Model 4 in Year 5.
BPCI-A	October 2018-December 2019	Coordinating care across multiple types of health care providers involved in episodic care that is initiated with a hospitalization or an outpatient procedure through 90 days after discharge or the end of the procedure.	90-day episode-based bundle (discount to prospective target price) tied to quality	Hospitals, PGP, clinicians with surgical specialties	FFS acute and outpatient care: 30 inpatient, 3 outpatient, & 1 multi-setting clinical episodes Average baseline Medicare spending varied by episode but was around $25,000.	As of January 2020, 1,084 (33% of eligible) hospitals and 1,166 PGPs participated at some point during the first 3 model years. From the start of the model through the end of 2020, close to 1 million total BPCI-A episodes were initiated.
CEC	2015-2020	Providing coordinated care for Medicare beneficiaries with ESRD through dialysis facilities, nephrologists, and nephrology practices.	ACO with one and two-sided risk tied to quality	Dialysis facilities and nephrologists	FFS ESRD beneficiaries with average baseline total Medicare expenditures of $78,480 with an average HHC score ranging from 1.05 to 1.20.	There were 37 ESCOs that participated in the model with 33 remaining in the model as of 2020. Seven dialysis organizations participated in 2020 with an average of 35 dialysis facilities in each ESCO (1,290 in total) across 32 states and Washington, D.C. touching 62,501 FFS beneficiaries (13% of Medicare beneficiaries with ESRD).
CJR * ^	Episodes initiated from April 2016-December 2019	Hospitals responsible for coordinating care with health care providers involved in the surgical episode to provide better quality.	Episode-based bundle (discount to prospective target price) tied to quality	Hospitals, health systems, orthopedic surgeons, PAC providers	Lower extremity joint replacement (LEJR) with and without complications. FFS beneficiaries in CJR had an average of about $29,000 in baseline Medicare payments and average baseline HCC score of 1.6 for mandatory hospitals and 1.31 for voluntary hospitals.	As of 2019, there were 831 hospitals across 67 Metropolitan Statistical Areas (MSAs) in the mandatory CJR model covering over 153,000 FFS episodes. Beginning in PY3, the number of mandatory MSAs was scaled back from 67 to 34 MSAs (395 mandatory hospitals) with the highest average historical payments. There were 74 opt-in and 200 non-opt-in hospitals in the 33 voluntary MSAs where hospitals could elect to continue participation. Hospitals in mandatory MSAs designated low-volume or rural (N=126) had a choice to participate in the model after PY2 of which 15 chose to participate (but are not included in the evaluation analyses) and 111 did not.
HHVBP * ^	2016-2020	Home health agencies are given a financial incentive to prioritize quality over volume.	Financial incentives are budget neutral to CMS.	Home Health Agencies	The model serves FFS beneficiaries that are homebound with functional status limitations (average 2.7-3.1 HCC score; average baseline Medicare spending per day during and following the episode was $138.33 ($50,307 annually) and need regular assistance with the goal of improving beneficiaries' physical abilities to care for themselves.	As of 2020, there were 1,907 home health agencies serving 734,951 FFS beneficiaries (2 million episodes).

Model	Years covered	Intervention	Payment	Health care providers	Beneficiaries	Size
MD All-Payer	2014-2018	All-payer rate setting for hospital services with hospitals accountable for total hospital cost of care and quality.	None, Waiver exempted Maryland hospitals from inpatient and outpatient prospective payment systems and shifted state's hospital payment structure to an all-payer annual global budget using rate-setting	All regulated acute care hospitals in Maryland	All patients hospitalized at regulated Maryland hospitals. Average baseline beneficiary HCC scores ranged between 0.99 to 1.14 with around $12,000 in total annual Medicare expenditures.	By July 2014, all 46 general acute care hospitals in the state began operating under a global budget.
MCCM *	2016-2020	Hospices provide supportive services to terminally ill beneficiaries while these patients have the option to continue to receive curative treatment.	Prospective care management ranging from $200 to $400 per beneficiary	Hospices and palliative care providers including nurses, care coordinators, social workers, clergy, and bereavement counselors.	Terminal illnesses (cancer [72%], congestive heart failure [38%], chronic obstructive pulmonary disease [34%], or HIV/AIDS [<1%] in FFS. Baseline total Medicare expenditures were around $31,000 with an average HCC score of 5.6 at enrollment.	The model started with 141 hospices across 41 states. As of 2021, there were 49 hospices participating in 25 states. MCCM served over 6,500 FFS beneficiaries since the start of the model in 2016.
OCM	Episodes initiating July 2016-January 2019	Multi-payer model enhancing treatment for cancer patients by having oncology physician group practices enter into total cost of care responsibility for episodes of cancer care that include financial and performance accountability.	Prospective care management ($160 per month in episode) and performance-based incentive payments; 6-month episodes of care surrounding chemotherapy administration to cancer patients.	Oncologists (1-350+ oncologists per practice)	The average baseline Medicare expenditures were $28,681 across all episodes. HCC scores were 2.664 at baseline. High-risk cancers make up 66% of all episodes with the low-risk cancers comprising the remaining third.	As of 2019, the model included 176 oncology practices in 35 states initiating over 120,000 FFS episodes per performance period (987,332episodes total).
RSNAT	2014-2019	Test whether prior authorization for RSNAT services (medically necessary, scheduled, non-emergency ambulance transportation for three or more round trips in a 10-day period or at least once a week for three weeks) can decrease Medicare expenditures without affecting beneficiaries' access to or quality of care.	Requires independent ambulance suppliers in participating states to obtain prior authorization for RSNAT services or claims are subject to prepayment review.	Ambulance suppliers in participating states (N=3,177)	The evaluation examined beneficiaries with ESRD and/or stages 3-4 pressure ulcers account for 85% of all Part B RSNAT claims, which were the focus of the evaluation (N=603,818) with average beneficiary baseline HCC scores of 4.3 and total annual Medicare expenditures of $15,918.	Phase 1: Three states (New Jersey, Pennsylvania, South Carolina) with high historical improper payment and use (started December 1, 2014). Phase 2: Six states (District of Columbia, Delaware, Maryland, North Carolina, Virginia, and West Virginia) with improper payment and use closer to the national average (started January 1, 2016). The model ended December 1, 2020 under Innovation Center authority but continues to operate under MACRA authority.

Notes. HIV/AIDS= human immunodeficiency virus/acquired immunodeficiency syndrome; BPCI=Bundled Payments for Care Improvement; BPCI-A=Bundled Payments for Care Improvement Advanced; CEC=Comprehensive ESRD Care Model; CJR= Comprehensive Joint Replacement Model; ESCO=ESRD Seamless Care Organization; ESRD= End Stage Renal Disease; FFS=fee-for-service; HHA=Home Health Agencies; HHVBP=Home Health Value-Based Purchasing Model; HCC=Hierarchical Condition Category; IRF=Inpatient Rehabilitation Facility; LTCH=Long-Term Care Hospital; MACRA=Medicare Access and CHIP Reauthorization Act of 2015. MD= Maryland; MCCM= Medical Care Choices Model; OCM=Oncology Care Model; PGP=Physician Group Practices; PY1= Performance Year 1; RSNAT= Prior Authorization of Repetitive, Scheduled Non-Emergency Ambulance Transport; SNF= Skilled Nursing Facilities. ^ Indicates the model was mandatory (otherwise models without this symbol were voluntary). * Indicates the model included a randomized design (otherwise models without this symbol were not randomized and can be considered quasi-experimental).

Appendix Table 2. Primary Care and Population Management Medicare Models and Demonstrations

Model or demonstrat	Years covered	Intervention	Payment	Health care providers	Beneficiaries	Size
AIM Test 1 ACO	2016-2018	Pre-paid shared savings to encourage new Medicare Shared Savings Program ACOs to form in rural and underserved areas.	Up-front and monthly payments to SSP ACOs recouped from shared savings.	Primary care practices, small hospitals, Federally Qualified Health Centers or Rural Health Clinics	Traditional FFS with plurality of primary care received by ACO health care providers. Average beneficiary baseline HCC scores ranged from 0.99 to 1.17 with average total annual Medicare expenditures of $11,000 to $16,000.	492,114 Medicare beneficiaries were served by 41 AIM ACOs (5,422 practitioners; 691 facility-based providers) spanning 37 states in 2018.
AP ACO	2013-2015	Pre-paid shared savings to encourage new Medicare Shared Savings Program ACOs to form	Up-front and monthly payments to SSP ACOs recouped from shared savings	Primary care practices, hospitals, Federally Qualified Health Centers or Rural Health Clinics	Traditional FFS with plurality of primary care received by ACO health care providers. Average beneficiary baseline chronic condition count was 4.4 and averaged $10,112 in total annual Medicare expenditures.	290,292 Medicare beneficiaries were served by 36 AP ACOs in 2014.
CPC	October 2012-December 2016	Multi-payer primary care medical home	Prospective care management fees and regional shared savings payments	Primary care practices and health care providers with some having formal agreements with specialty care providers	Traditional FFS with plurality of primary care received by CPC practices. Average beneficiary baseline HCC scores were 1.15 with $11,000 in total Medicare expenditures.	In 2016, 439 practices with 2,159 clinicians participated across 7 regions serving 320,173 Medicare FFS patients.
CPC+	2017-2019	Multi-payer primary care medical home	Prospective care management fees, performance-based incentive payments, and Track 2 hybrid payment (prospective quarterly population-based payment and reduced FFS)	Primary care practices	Traditional FFS with plurality of primary care received by CPC+ practices. Average beneficiary baseline HCC scores were 1.0 with $6,890 in total Medicare expenditures.	As of 2019, there were 3,070 CPC+ practices across 18 regions with 13,739 practitioners serving 2.4 million Medicare FFS beneficiaries.
FAI – Washington	2013-2019 (2017-2019 for utilization and quality measures)	Demonstration providing full benefit dually enrolled beneficiaries better care through aligned financial incentives of Medicare and Medicaid programs. Managed FFS with care coordination through health homes as a Medicaid covered benefit.	Shared savings with State in managed FFS model, with care coordination as a Medicaid covered benefit through health homes, which receive a per-member-per-month Medicaid payment.	Network of care coordination organizations representing primary care, mental health, long-term services and supports, chemical dependency services and specialty providers.	Dually-eligible for Medicare and Medicaid as well as considered high-cost, high-risk. The mean HCC scores ranged from 2.0-2.3 during the sixth demonstration year with an average annual Medicare spending of $14,576.	In 2019, there were 30,445 eligible Medicare-Medicaid beneficiaries were aligned with the Washington demonstration, of which 12,114 were enrolled in a health home.
IAH	June 2012-September 2017	Demonstration providing home-based primary care to beneficiaries with multiple chronic conditions.	Practices can earn incentive payments if their patients' expenditures are below target expenditures and meet quality standards	Physicians and nurse practitioners	Home-based primary care for Medicare FFS beneficiaries with at least two chronic conditions that need help from another person to complete two activities of daily living, have been admitted to the hospital and received acute or subacute rehabilitation in the past year. Long-term or hospice care exclude beneficiaries from enrollment. Average baseline beneficiary HCC scores were 3.5 with total Medicare expenditures of $52,764 per year.	Each practice has to serve at least 200 patients per year. The statute provided that the Secretary shall limit the number of practices so that the number of applicable beneficiaries that may participate in the demonstration does not exceed 10,000 for Years 1 to 5. In year 5, IAH had 14 home-based primary care sites, including a consortium of three practices.
MA VBID	2017-2019	Testing a broad array of complementary MA health plan innovations targeting enrollees based on chronic conditions to use high-value care (e.g. cost-sharing and supplemental benefits).	None from CMS; Waivers incentivize changes	N/A; MA Plans	All MA (no FFS). Average baseline beneficiary HCC scores ranged from 1.5-less than 2.0.	As of 2019, there were 11 MA parent organizations participating in 6 states serving 105,257 eligible MA beneficiaries, 62% of whom participated in an intervention.

Model or demonstrat	Years covered	Intervention	Payment	Health care providers	Beneficiaries	Size
MH *	2017-2019	Prevention-focused model where participants are paid to assess CVD risk and lower the incidence of first-time heart attack and stroke.	Prospective care management tied to risk stratification and change in risk scores	44 primary care practices, 30 specialty and multispecialty organizations, 4 federally qualified health centers, 1 critical access hospitals or rural hospital, 5 acute care hospitals and other health centers	FFS ages 40-79 without previous heart attack / stroke with average HHC score of 1.16 and around $8,000 in total annual Medicare expenditures at baseline.	The model has been implemented as a randomized control trial. As of 2019, there were 173 MH practices across 47 states serving 230,664 beneficiaries in the intervention group.
NGACO	2016-2019	Tests whether strong financial incentives, flexible payment options, and tools to support care management improve value and lower costs.	ACO two-sided risk tied to quality and optional population-based payments	Physician practices (both primary and multi-specialty); integrated delivery systems; hospital systems, partnerships between physician practices and hospitals. Includes physician and non-physician practitioners, hospitals, SNFs, IRFs, LTCHs, HHAs, hospices, and other institutional providers.	Traditional FFS with plurality of primary care delivered by ACO practitioners whose services are used in alignment. Average beneficiary baseline spending was $13,000 with a mean of 5 chronic conditions (however the number of conditions had a wide range with a standard deviation of 3.7).	In 2019, there were 41 NGACOs (down from 50 in 2018) across 29 states with an average of 4,546 participating and preferred practitioners serving 1.2 million eligible FFS beneficiaries across the model.
Part D Enhanced MTM	2017-2019	29 interventions (e.g. medication reconciliation, reviews, adherence)	Prospective and performance-based payments (premium subsidy)	Standalone prescription drug plan sponsors	Part D plans (for FFS beneficiaries). Average beneficiary baseline Medicare expenditures were $11,144 with an HCC score of 1.16 for the entire set of beneficiaries who were enrolled in the Part D plans. However, plan sponsors targeted their beneficiaries based on health care needs and costs with LIS and medically complex beneficiaries having a range between $13,501 to $19,986 in average baseline Medicare Parts A & B expenditures with an HCC ranging from 1.42 to 1.87.	As of 2019, there were 6 sponsors who operated 22 plan benefit packages in five Part D regions with 1.9 million FFS beneficiaries.
Pioneer ACO Model	2012-2013	Health care organizations and providers/suppliers experienced in coordinating care for patients across care settings.	Shared savings and losses tied to quality	Primary care practices with some specialty care, hospital, and PAC affiliations	Traditional FFS using qualifying evaluation and management services received by ACO health care providers. Average annual baseline spending was $11,238 per beneficiary.	32 ACOs, each averaging around 20,000 aligned beneficiaries in the first performance year and 50,000 in the second performance year (over 500,000 beneficiaries served in total).
VT All-Payer ACO	2018-2019	State-level population health and multi-payer ACO-level interaction.	ACO two-sided risk tied to quality and optional population-based payments similar to financial methodology used in NGACO	Primary care practices with some specialty care & hospital affiliations	Traditional FFS with plurality of primary care received by ACO health care providers. Average baseline Medicare expenditures was around $10,000 and beneficiaries had an average of 4 chronic conditions.	One ACO with 10-13 participating hospitals and 4,000 practitioners in VT serving 37,000-54,000 Medicare FFS beneficiaries.

Notes. AIM=ACO Investment Model; AP=Advance Payment Model; ACO=accountable care organization; CPC=Comprehensive Primary Care Initiative; CPC+=Comprehensive Primary Care Plus; CVD=Cardiovascular disease; MTM=Medication Therapy Management Model; FAI=Financial Alignment Initiative for Medicare-Medicaid Enrollees; FFS=fee-for-service; HCC=Hierarchical Condition Category; IAH=Independence at Home Demonstration; LIS=low income subsidy; MA=Medicare Advantage; MH=Million Hearts®: Cardiovascular Disease Risk Reduction Model; N/A=not applicable; NGACO=Next Generation ACO Model; SSP=Shared Savings Program; VBID=Value-Based Insurance Design Model; VT=Vermont. *Indicates the model included a randomized design.