Project Report

Evaluation Report of the First Three Years (2017–2019) of the Medicare Advantage Value-Based Insurance Design Model Test

Christine Eibner, Dmitry Khodyakov, Erin Audrey Taylor, Christine Buttorff, Courtney Armstrong, Marika Booth, Kathryn E. Bouskill, Matthew Cefalu, Stephanie Dellva, Michael Dworsky, Federico Girosi, Ann C. Haas, Monique Martineau, Blen Eshete-Roesler, Alice Kim, Julie Lai, Afshin Rastegar, Daniel Schwam, Tisamarie Sherry, Shiyuan Zhang

RAND Health Care

www.rand.org/health-care

September, 2020

PR-4677-CMS

Prepared for the Centers for Medicare and Medicaid Services
Center for Medicare and Medicaid Innovation
Under RMADA Contract Number HHSM-500-2014-00036I
Order Number HHSM-500-T0003
Sarah Lewis, Contracting Officer's Representative (COR)

This project was funded by the Centers for Medicare & Medicaid Services under contract no. HHSM-500-2014-00036I, Task Order HHSM-500-T0003. The statements contained in this report are solely those of the authors and do not necessarily reflect the views or policies of the Centers for Medicare & Medicaid Services. RAND assumes responsibility for the accuracy and completeness of the information contained in this report.

Preface

This report presents the RAND Corporation team's findings from its evaluation of the first three years (2017–2019) of the Medicare Advantage (MA) Value-Based Insurance Design (VBID) Model Test, initiated by the Center for Medicare & Medicaid Innovation (CMMI). VBID allows MA plans to 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.

In this report, we describe findings from interviews with MA insurers (called Parent Organizations) and Medicare beneficiaries who were eligible to receive VBID benefits. We also report findings from analyses to estimate the effects of VBID on a range of outcomes, for various years after the model was implemented based on data availability. The outcomes studied include health care utilization, health care quality, health outcomes, beneficiary spending, MA bids, and costs to the Centers for Medicare & Medicaid Services. The results are useful to a range of audiences, including policymakers, health plans, and researchers interested in insurance benefit design.

The evaluation was funded by CMMI under Research, Measurement, Assessment, Design, and Analysis Contract Number HHSM-5002014-00036I, Order Number HHSM-500-T0003, for which Sarah Lewis is the contracting officer's representative. It was carried out within the Payment, Cost, and Coverage Program in RAND Health Care.

RAND Health Care, a division of the RAND Corporation, promotes healthier societies by improving health care systems in the United States and other countries. We do this by providing health care decisionmakers, practitioners, and consumers with actionable, rigorous, objective evidence to support their most complex decisions.

For more information, see www.rand.org/health-care, or contact
RAND Health Care Communications
1776 Main Street
P.O. Box 2138 Santa Monica, CA 90407-2138
(310) 393-0411, ext. 7775
RAND_Health-Care@rand.org

Contents

Figures

Tables

Executive Summary

Value-based insurance design (VBID) refers to health insurers' efforts to structure cost-sharing and other health plan design elements to encourage enrollees with chronic conditions to use the services that can benefit them the most. Examples include lower prescription drug copayments to increase medication adherence and reduced patient cost-sharing for visits with high-value providers to spur people to get recommended checkups. The goals of VBID are to improve patient health through better disease control and to save money by reducing costly complications that can occur when chronic conditions are poorly managed.

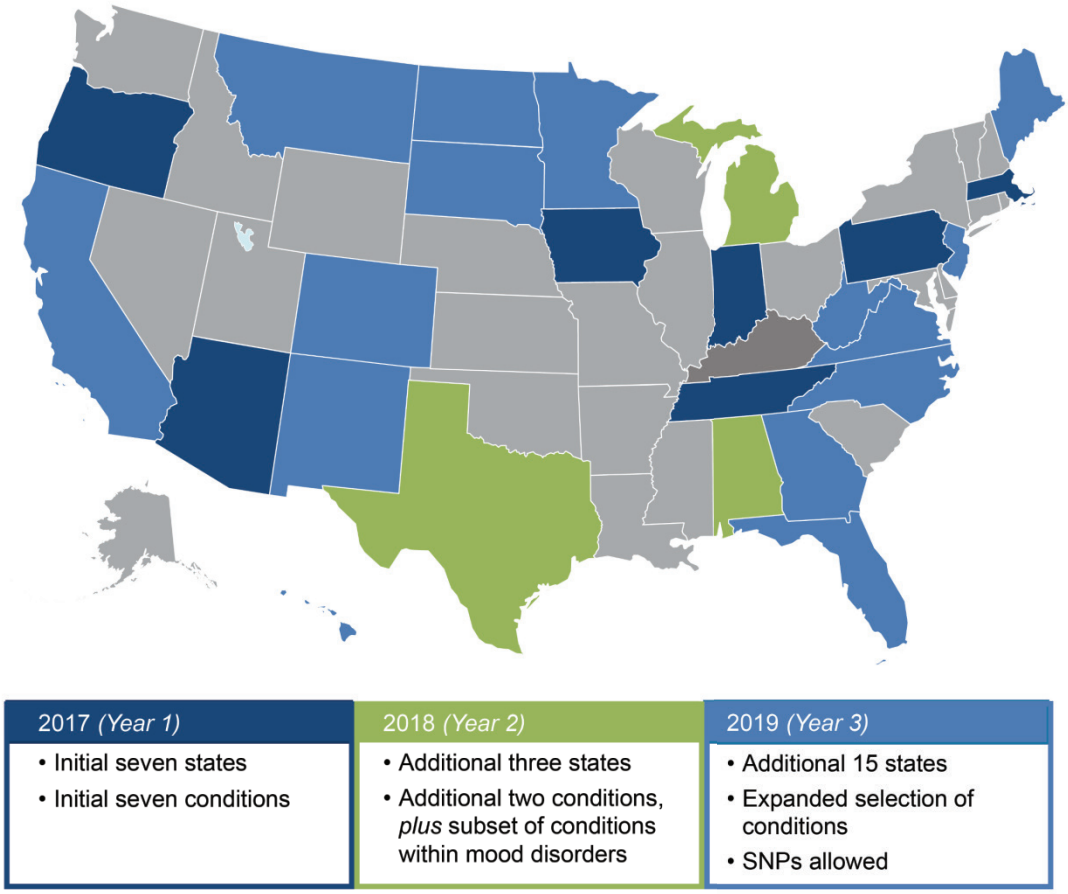
Until recently, VBID had not been implemented or tested among individuals aged 65 and older. However, under the authority granted to the Centers for Medicare & Medicaid Services (CMS) under section 1115A of the Social Security Act, the Center for Medicare & Medicaid Innovation (CMMI) within CMS is currently conducting a Medicare Advantage (MA) VBID model test. First implemented in 2017, the voluntary program allows Medicare Advantage Organizations (MAOs) and their parent organizations (POs) to alter specific benefit design elements within their plan benefit packages (PBPs) to encourage enrollees with targeted conditions to use high-value services and providers.

CMMI contracted with the RAND Corporation to evaluate the first three years of the MA VBID model test, which ran from January 1, 2017 through December 31, 2019. This report serves as our second and final evaluation of the first three years of the model test. The evaluation considers POs, beneficiaries, and, in some cases, providers' participation experiences and assesses the effect of VBID on key outcomes such as health care service utilization, cost, and quality. This report builds on our first evaluation report of the MA VBID model test (Eibner et al., 2018), published in early 2019, by adding data that have since become available and by providing information on how POs' experiences have changed over time.

Center for Medicare & Medicaid Innovation Value-Based Insurance Design Model Test

This evaluation estimated changes in PO and beneficiary participation, experiences, and outcomes during the first three years of the model test (2017–2019). Figure S.1 shows the states in which the MA VBID model test was offered, as well as how state eligibility changed over time. The figure also describes changes in the chronic conditions that POs could target with their VBID interventions.

Figure S.1. VBID Eligibility over Time



In 2017, the model test was available to POs in seven states: Arizona, Indiana, Iowa, Massachusetts, Oregon, Pennsylvania, and Tennessee. Participating POs could offer reduced cost-sharing to eligible beneficiaries who used high-value services or high-value providers, or who participated in care management/disease management (CM/DM) activities. POs could also offer supplemental benefits (e.g., nutrition services, transportation) to eligible beneficiaries. VBID could be offered to beneficiaries with one or more of the following conditions: chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), coronary artery disease (CAD), diabetes, hypertension, mood disorder, and past stroke.

In 2018, CMS expanded the model test to three additional states (Alabama, Michigan, and Texas) and added two new conditions: dementia and rheumatoid arthritis. In addition, CMS allowed POs to restrict eligibility to specific mood disorders such as depression, rather than all mood disorders. In April 2018, CMS announced that, beginning in the 2019 plan year, the long-standing uniformity rule requiring that MA PBPs offer the same benefits to all enrollees for the same premium and cost-sharing, was being reinterpreted. This reinterpretation now allows VBID-style flexibility outside the MA VBID model test; for example, POs can reduce copays

for specialist visits for enrollees with chronic conditions, even if the PO does not participate in the MA VBID model test. The new uniformity flexibility does not apply to Part D (prescription drug) benefits, so POs must remain in the model test to offer lower copays for prescription drugs.

CMS further expanded the VBID model in 2019 by making it available in 15 additional states, allowing POs to identify beneficiaries based on any chronic condition or to select their own set of diagnosis codes within the previous nine eligible conditions, and also allowing certain Special Needs Plans to participate.

Provider Organization Participants

Initially, nine POs from three states (Indiana, Massachusetts, and Pennsylvania) participated in the model test, offering benefits for four of the seven eligible conditions (COPD, CHF, diabetes, and hypertension). One additional PO joined the model test in 2018, offering benefits for CAD. Although one more PO joined the model test in 2019, a prior participant exited in the same year. Furthermore, one model participant expanded its VBID offerings to a new state in 2019. In total, nine POs participated in 2017 and ten participated in 2018 and 2019. Across all three years, 11 unique POs participated.

By 2019, VBID was offered in six states (Arizona, Indiana, Massachusetts, Michigan, Pennsylvania, and West Virginia) and was offered to beneficiaries with one or more of five conditions (CAD, CHF, COPD, diabetes, and hypertension). Most participating POs were state-based insurance providers; there was one national participant. Of the 11 POs participating at some point in Years 1 through 3, five were Blue Cross and/or Blue Shield affiliates.

Although this report focuses on the first three years of the model test, we discussed in interviews with POs their intentions to stay in MA VBID after 2019, given substantial changes to the model beginning in 2020. ***Four of the 11 POs that participated in the model test between 2017 and 2019 continue to participate in 2020.*** Additional POs have also joined the model test.

Intervention Designs

CMS allowed participating POs substantial flexibility in designing and implementing their VBID interventions, leading to variation in the VBID approaches that were adopted. Table S.1 summarizes the interventions implemented by the 11 POs that participated in the MA VBID model test at some point between 2017 and 2019. ***Most (7 of 11) POs required beneficiaries to complete certain requirements,*** such as participating in CM/DM activities, to receive VBID benefits; these POs are shown first in the table. ***CHF, COPD, and diabetes were the most commonly targeted conditions,*** with just one PO targeting hypertension and another targeting CAD. Specialist visits were the most frequently chosen high-value service (targeted by six POs). ***Three of the 11 POs reduced cost-sharing for high-value drugs,*** the VBID approach for which the literature provides the most robust support.

Most POs entered multiple PBPs into the model test: There were 45 unique PBPs participating in 2017, which grew to 48 by 2019.

Table S.1. VBID Approaches

PO	Beneficiary Participation Requirements	Condition(s) Targeted	Description	Prescriptions Targeted
A	Must complete scorecard	Diabetes	Provided rebates for incurred cost-sharing if beneficiaries completed a "scorecard" of specific preventive screenings	No
B	Must participate in CM/DM	Diabetes and/or COPD	Reduced copayments for high-value providers and supplemental benefits	No
C	Must participate in CM/DM	CHF and COPD or CHF and diabetes	Provided rebates[a] for incurred cost-sharing for beneficiaries who completed up to six CM activities	No
E	Must participate in CM/DM	COPD	Waived cost-sharing for select specialty care visits, tests, and durable medical equipment	No
F	Must participate in CM/DM	COPD and CHF	Waived or reduced copayments for primary and select specialty care visits	No
G	Must participate in CM/DM	CHF	Waived copayments for primary care, cardiologists, and select generic prescription drugs[b]	Yes
I	Must participate in CM/DM	CHF	Provided free blood pressure cuffs, pulse oximeters, and weight scales	No
D	None	Hypertension	Eliminated cost-sharing for select drugs	Yes
H	None	Diabetes and CHF	Reduced copayments for select specialty care visits	No
K	None	Diabetes	Eliminated cost-sharing for endocrinologist visits	No
J	None	CAD	Eliminated cost-sharing for select drugs	Yes

[a] PO C changed the rebate to an over-the-counter debit card that accrued funds every time a beneficiary completed a required activity.
[b] PO G added select supplemental benefits in 2019.

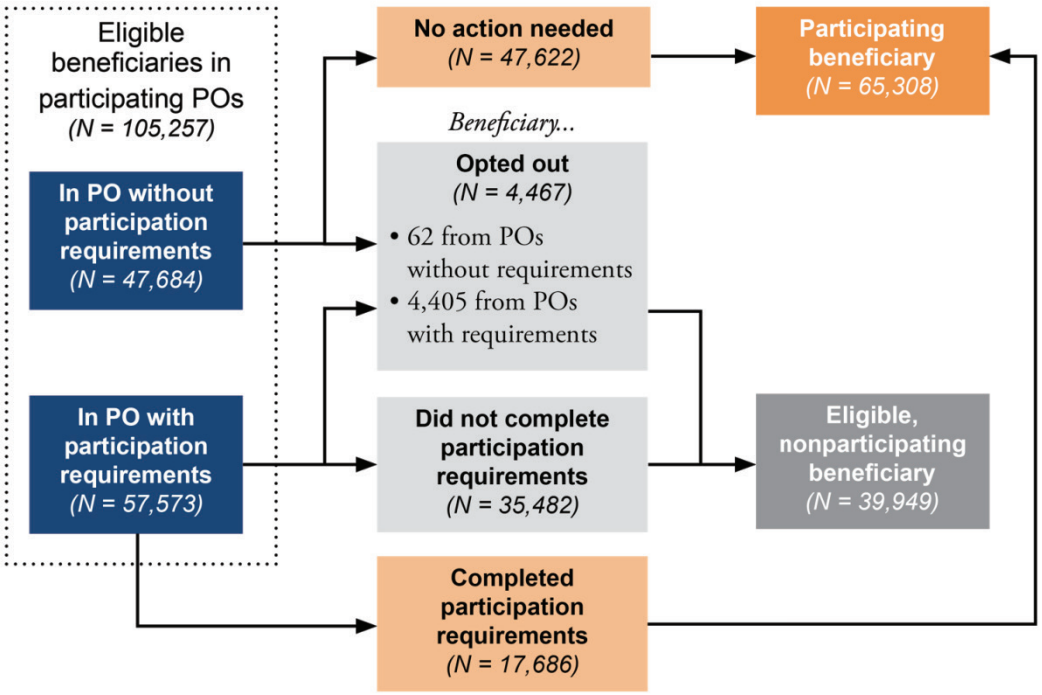
Most POs that continued participation in the model test did not change their intervention designs in their second or third year, despite changes by CMS to the model test in both 2018 and 2019 that allowed design adjustments. By 2019, POs were permitted to offer VBID for any condition; however, POs continued to target the same subset of conditions that were allowed at the beginning of the model test.

Participating Beneficiaries

The total number of beneficiaries who were eligible to participate in MA VBID grew from 98,740 beneficiaries in 2017 to 105,257 in 2018 (Figure S.2). Among beneficiaries in POs with

CM/DM participation requirements, about two-thirds opted out of VBID or did not complete participation requirements. More than 70 percent of participating beneficiaries were in the four POs without participation requirements.

Figure S.2. 2018 Participation Status Among VBID-Eligible Beneficiaries (*N* = 105,257)



Implementation Experiences

POs that were new to the model test identified several implementation challenges, including the need to reconfigure their information technology systems to track VBID beneficiaries and manage two sets of benefits within the same plan. Although many returning POs had worked through these early implementation difficulties, some POs expressed concern about several ongoing challenges.

Marketing restrictions: During the open enrollment period leading up to 2017 (the first year of the model test), POs were prevented from marketing VBID benefits to prospective enrollees. CMMI revised its guidance before the 2018 open enrollment period to allow POs to review VBID benefits with eligible beneficiaries, but POs were still prohibited from discussing the model test with prospective enrollees or beneficiaries without VBID conditions. For the 2019 open enrollment period, CMMI further clarified marketing restrictions, enabling POs to discuss VBID with everyone, including prospective enrollees.

POs perceived that limitations on their ability to communicate about VBID benefits during the 2017 and 2018 open enrollment periods may have negatively affected beneficiaries'

awareness of the model and their willingness to participate. For example, some POs believed that beneficiaries were more likely to review outreach materials provided during the open enrollment period than materials provided after they had already selected a PBP.

Identifying eligible beneficiaries: Some POs noted that ***data processing lags made it difficult to identify beneficiaries who became newly eligible for VBID (e.g., due to the onset of a chronic condition) in a timely manner.*** For example, after a beneficiary receives a qualifying diagnosis there may be a delay in receiving VBID benefits because POs' systems may not immediately recognize that the individual has become VBID eligible. Sometimes, POs described having to manually process refunds for copays paid after an eligible diagnosis was confirmed, but before their internal systems were updated to reflect a new VBID status.

Getting beneficiaries to engage with CM/DM: Several POs that implemented CM/DM programs found that ***it was difficult to get beneficiaries to agree to participate in and follow through with CM/DM activities.*** Perceived reasons for the difficulty in engaging beneficiaries varied across POs. Some POs reported that they had trouble engaging newly diagnosed beneficiaries who were less familiar with their conditions and might not be aware of their diagnoses. Other POs reported that sicker beneficiaries were more difficult to engage because they considered themselves to be too ill to participate in CM/DM activities and did not think their health could be improved through CM/DM activities, such as smoking cessation classes or nutrition programs.

Despite these concerns, several PO representatives expressed appreciation for participating in the model test as a learning environment in which they could work out operational issues and test new benefit designs. The knowledge gained in the process, many felt, would be helpful as the MA program allows more flexible benefit designs.

Value-Based Insurance Design's Effect on Key Outcomes

We assessed the effects of the VBID model test on a wide range of outcomes including health care utilization, health care costs, beneficiary enrollment and experiences, and other outcomes such as health care quality and health outcomes. Our results are based on interviews with POs and beneficiaries, as well as data analyses using difference-in-differences regressions that compare trends for VBID-participating POs and beneficiaries with matched comparison groups to test whether a PO's participation in the VBID model test affected outcomes. The regression analyses generally estimated VBID's impacts in the first year after the MA VBID model took effect, and, as a result, reflects short-term effects of VBID on key outcomes. For most analyses, we pool all VBID-participating POs and beneficiaries (and their matched comparators) into a single regression.

Of 80 outcomes considered, we found statistically significant results in 24 instances. Most of the statistically significant findings were for utilization outcomes, which is expected if changes in utilization precede changes in health outcomes or beneficiary experiences of care.

Changes were statistically significant for more than half of all utilization outcomes considered (18 of 27 that could be analyzed), most in the intended direction (15 of 18). Ten of

these increases were for services that POs specifically targeted with their VBID interventions, such as visits with specific types of specialists and use of targeted treatments like computed tomography (CT) scans and sleep studies. We also found increases in use of primary care services, specialist visits, prescription drugs, and several other service categories across all participating POs.

VBID is not yet generating savings to Medicare, and it is also not costing Medicare additional money, as expected. There were no changes in MA program costs to Medicare (i.e., actual payments made by CMS to plans for benefits provided to MA and Part D enrollees after accounting for final risk scores and rebates) nor in plans' own realized spending (i.e., actual spending by plans on medical and drug benefits for enrollees). We found statistically significant reductions in four Part D cost outcomes (beneficiary out-of-pocket [OOP] costs, premiums, costs to Medicare, and bids), and no change to MA-PD combined premiums.

There was only one statistically significant change among any of the patient experience, health care quality, or health outcomes measures that we considered. Specifically, we found a statistically significant improvement in patient-reported care coordination, but it is unclear whether this change can be directly attributable to VBID.

The finding that most of the changes were for utilization outcomes underscores the possibility that changes in other outcomes may be likely down the road. Use of services may change relatively quickly in response to reductions in cost-sharing, but it may take time for these changes in utilization to exert a meaningful effect on beneficiary health or costs.

Most of the changes that we estimated were small in magnitude; however, because data for many of the outcomes were not yet available for 2018 and 2019 (the second and third years after the model took effect) at the time of this report, it is possible that additional or larger effects would manifest over a longer duration period.

A limitation of our approach is that ***PBPs that were entered into the model test were inherently very different from nonparticipating PBPs.*** While our mixed-methods evaluation was intended to isolate the effect of VBID, this study is ultimately observational in nature. In the absence of a randomized controlled trial, it is difficult to establish whether these findings can be attributed to the VBID model alone, or in part.

Below we describe our findings in more detail, with a focus on the statistically significant changes observed.

Health Care Utilization

Our analyses considered changes in two types of utilization among VBID-participating beneficiaries and their matched comparators: VBID-targeted services and general services. Targeted services are the high-value health care services chosen by a specific PO for focused changes in cost-sharing or benefit design. Analyses of these outcomes are limited to within-PO utilization. We also considered utilization of certain health care services that could have changed due to VBID, regardless of whether they were explicitly targeted by POs (some were;

others were not). We called this category of outcomes "general utilization," and we analyzed changes in use for these services across all POs.

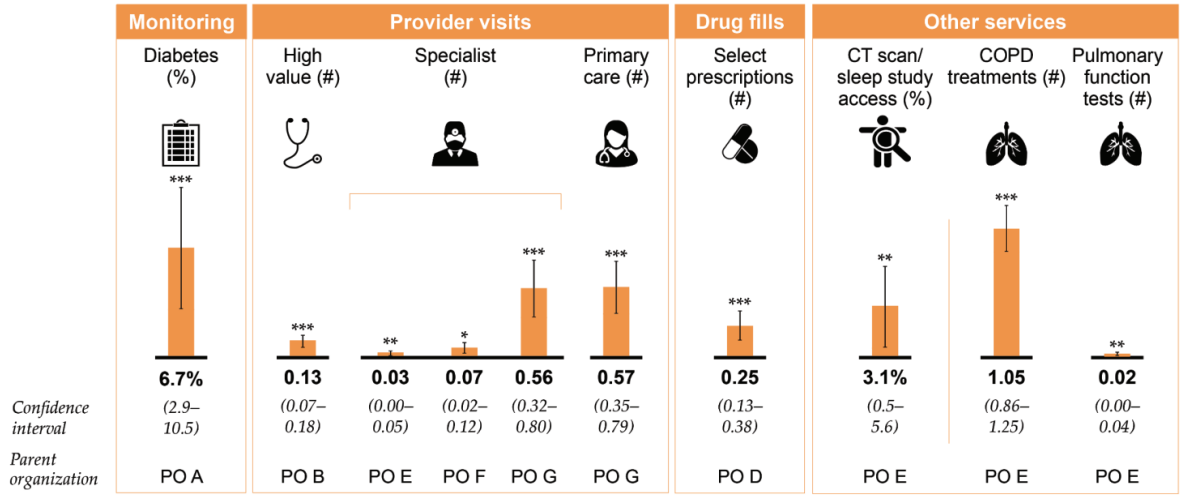
Targeted Utilization

We found statistically significant increases in use of targeted services among eligible beneficiaries in the following ten outcomes, relative to matched comparators:

- Proportion of beneficiaries with completion of a scorecard—a set of four diabetes monitoring activities—in PO A (6.7-percentage-point increase; confidence interval [CI]: 2.9–10.5 percent)
- High-value provider visits in PO B (0.13 visits; CI: 0.07–0.18)
- Selected prescription drugs in PO D (0.25 fills; CI: 0.13–0.38)
- Specialist visits in PO E (0.03 visits; CI: 0.00–0.05)
- Proportion of beneficiaries accessing CT scans and sleep studies in PO E (3.1 percentage points; CI: 0.5–5.6)
- COPD treatments in PO E (1.05 services; CI: 0.86–1.25)
- Pulmonary function tests in PO E (0.02 tests; CI: 0.00–0.04)
- Specialist visits for PO F (0.07; CI: 0.02–0.12)
- Primary care provider (PCP) visits in PO G (0.57 visits; CI: 0.35–0.79)
- Specialist visits in PO G (0.56; CI: 0.32–0.80).

These increases are consistent with expectations and align with prior research showing that utilization increases when cost-sharing is reduced. Figure S.3 summarizes the statistically significant findings among the targeted utilization outcomes.

Figure S.3. Summary of Statistically Significant Increases in Targeted Services Utilization in 2017, with Confidence Intervals



NOTES: Bars show changes in the number of services utilized except for diabetes monitoring and CT scan/sleep study Access, which reflect the change in the proportion of beneficiaries accessing the service. "Select Prescriptions" refers to the elimination of cost sharing for Tier 1 and Tier 3 hypertensive drugs by PO D. Statistical significance: ***p < 0.01; **p < 0.05; *p < 0.10.

Note that Figure S.3 shows only the ten services that had statistically significant changes, out of 18 services targeted. For six of the remaining targeted services, there was no statistically significant change. We were unable to test changes for two services because of a lack of data.

General Utilization

For general utilization, we looked at overall changes in the number of services used for several specific services, such as visits to primary care or specialist physicians. We also looked at the proportion of beneficiaries accessing certain services during our evaluation period.

Among all POs, we found an increase in primary care visits of 0.14 visits per-beneficiary per-year **(CI: 0.10–0.18)** ***and an increase in specialist visits of 0.19 visits per-beneficiary per-year*** **(CI: 0.12–0.27)** across all eligible beneficiaries in participating POs, relative to matched comparators. On average, beneficiaries in our sample had 4.5 primary care and 9.4 specialty care visits each year, so the changes represent a 2- to 3-percent increase in utilization. Although small, these effects could be meaningful for the subset of the population who use the services. For example, the 0.14 increase in PCP visits per-beneficiary per-year represents about 14,000 additional visits among VBID-eligible beneficiaries.

Prescription drug use increased among VBID-eligible beneficiaries, from 45.1 to 45.3 fills per-beneficiary per-year (an increase of 0.19 30-day medication fills, CI: 0.01–0.34). Although this is a less than 1-percent increase in prescription drug utilization, it represents an additional 19,000 prescriptions. Sensitivity testing with a restricted sample suggested that these increases were driven by beneficiaries in POs that targeted drug copayments as part of their intervention.

VBID was associated with small increases in ambulatory care sensitive (ACS) visits to inpatient and emergency department (ED) settings. ACS ED and inpatient visits are those that could potentially be avoided or reduced through timely treatment in outpatient settings and represent a small subset of overall inpatient and ED visits.[1] VBID beneficiaries experienced a 0.54-percentage-point increase in the probability of a beneficiary having any ACS ED visits (CI: 0.34–0.75) and a 0.67 percentage point increase in the probability of an ACS inpatient visit (CI: 0.40–0.95). Increases in ACS ED and inpatient use are not consistent with the goals of VBID, which aims to reduce costly complications that could lead to hospital stays and ED visits. Our interviews with POs did not suggest that their internal evaluations of their own VBID interventions identified statistically significant increases in ED and hospitalization among VBID participants, although some POs raised a general concern about lack of return on investment.

Interviews with beneficiaries revealed a high degree of confusion about what benefits were covered under VBID, so it is possible that utilization of noncovered services (like ED and

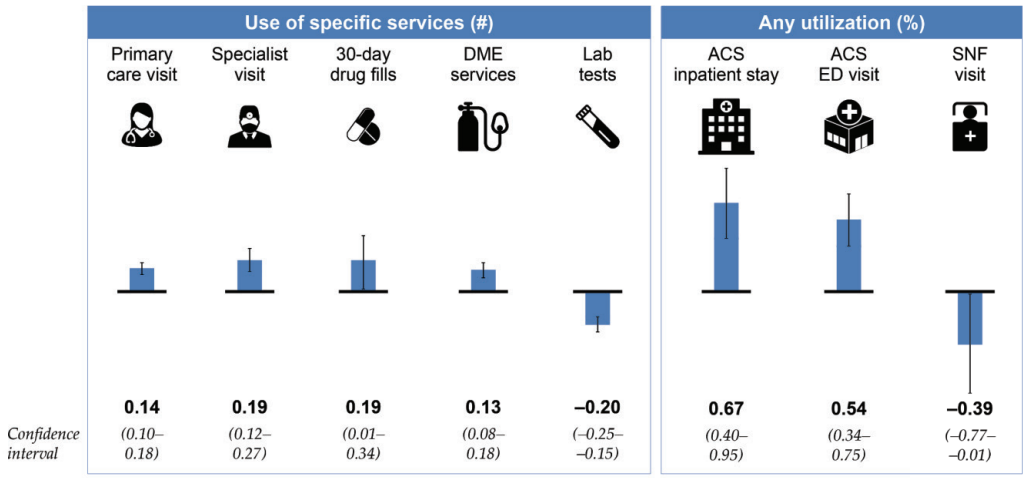
[1] For reference, 25.1 percent of VBID-eligible beneficiaries in our sample had an inpatient visit for any reason in 2017, whereas only 3.6 percent of beneficiaries had an ACS inpatient visit. Similarly, 40.2 percent of VBID-eligible beneficiaries had an ED visit for any reason in 2017, compared to 6.1 percent with an ACS ED visit. We found no change in the overall probability of having an ED or inpatient visit.

inpatient visits) could have increased because beneficiaries thought these services were eligible for reduced cost-sharing. Another potential explanation is that, by increasing interactions with PCPs and care managers, VBID contributed to beneficiaries receiving more diagnoses and referrals for additional treatments, ultimately leading to more ED and inpatient visits.

We found that VBID was associated with statistically significant changes in three additional utilization outcomes. For example, the number of laboratories and tests used declined (−0.20; CI: −0.25 to −0.15). We had no hypothesis about how VBID might affect use of laboratory tests. On the one hand, some POs aimed to increase laboratory tests with their interventions; on the other hand, better CM/DM could result in fewer duplicative or unnecessary tests. We also found increased use of durable medical equipment (DME) and a reduction in skilled nursing facility (SNF) visits. However, prior work with the CMS encounter data that we used for this report suggests potential reporting inconsistencies for DME and SNF visits, so these findings should be interpreted with caution.

Figure S.4 summarizes the statistically significant findings among the general utilization outcomes.

Figure S.4. Summary of Statistically Significant Increases in General Utilization in 2017, with Confidence Intervals



Health Care Costs per Beneficiary

We analyzed several measures of health care costs, including PBPs' spending per beneficiary on medical services and prescription drugs, premiums for MA and Part D plans, beneficiaries' OOP drug costs, PBP bids, and costs to Medicare. These measures are described in more detail in Table S.2. When measuring costs, we generally focused on the subset of PBPs that offered both MA and Part D coverage.

Table S.2. Health Care Cost Measures

Cost Measure	Definition	Data Source(s)	Model Years Included in Analyses
PBP spending			
Medical services	Spending by the PO on medical services on behalf of PBP enrollees	PBP bids	2017
Prescription drugs	Spending by the PO on prescription drugs on behalf of PBP enrollees	PDE data	2017, 2018
Beneficiary costs			
Premiums	Monthly amount paid by beneficiaries enrolled in PBPs for MA and Part D coverage	Public MA and Part D landscape files, IDR premium data	2017–2019
Part D OOP	Annual beneficiary copayments and coinsurance payments for covered Part D drugs	PDE data	2017, 2018
PBP bids			
Medical services	Projected costs of health care coverage for coming year; submitted six months before start of plan year	PBP MA bid data	2017–2019
Prescription drugs	Projected costs of prescription drug coverage for coming year; submitted six months before start of plan year	PBP Part D bid data	2017– 2019
Costs to Medicare			
Medical services	Costs paid by CMS for health care coverage. These include the portion of the MA PBP bids paid by CMS, adjusted by the risk score for each enrollee in the PBP and MA PBP rebates paid.	PBP MA bids, risk scores	2017, 2018
Prescription drugs	Costs paid by CMS for prescription drug coverage. These include the portion of the Part D PBP bids paid by CMS, adjusted by the risk score for each enrollee in the PBP, and additional CMS costs associated with Part D reinsurance and low-income subsidies.	PBP Part D bids, risk scores, publicly available Part D payment data, PDE low-income cost-sharing data	2017, 2018

NOTE: PDE = prescription drug event.

PBP spending per-beneficiary per-month on medical care and prescription drugs did not change. Spending is reported retrospectively and reflects the costs POs incurred rather than their projections.

While overall MA-PD premiums did not significantly change in VBID-participating PBPs, monthly MA premiums in VBID-participating PBPs increased from baseline relative to growth in comparison PBPs. We found increases in MA premiums from baseline for VBID PBPs in 2017 ($11; CI: $2 to $20), 2018 ($21; CI: $10 to $33), and 2019 ($18; CI: $4 to $33)

when compared to the growth from baseline in MA premiums in non-VBID PBPs. ***VBID monthly Part D premiums significantly decreased (-$9; CI: -$17 to -$1), but only in 2018.***

We found a statistically significant reduction in 2017 and 2018 Part D OOP cost-sharing at the point of service (–$21; CI: –$28 to –$14 for 2017 and –$15; CI: –$25 to –$6 for 2018). The reduction in Part D OOP cost-sharing represented a 4- and 3-percent decrease relative to an expected average of $556 and $548 for VBID beneficiaries in 2017 and 2018, respectively.

VBID was associated with a statistically significant reduction in Part D bids in 2018 and 2019, but there was no change in MA or Medicare Advantage and Part D (MA-PD) bids. We did not find a significant effect of VBID on Part D bids in 2017. Bids are the amounts that POs prospectively submit to CMS every June to indicate the expected cost of coverage per-beneficiary per-month for the upcoming year. Because bids are submitted six months before the start of the new plan year, they do not necessarily reflect actual costs for that year. We estimated the effect of VBID on both MA and Part D bids, as well as the combined MA-PD bid (the sum of MA and Part D bids for PBPs that offer both benefits). We found reductions in Part D bids for VBID PBPs in 2018 (–$10; CI: –$17 to –$3) and 2019 (–$12 CI: –$21 to –$3) relative to matched comparators. These declines in Part D bids represented statistically significant reductions of 14 percent in 2018 and 18 percent in 2019. However, these declines were offset in the combined MA-PD bids by statistically insignificant increases in MA bids.

Costs to Medicare include the portion of the MA and Part D PBP bids paid by Medicare, which are adjusted by the risk score for each beneficiary enrolled in the PBP, MA PBP rebates paid, and additional Medicare costs associated with Part D reinsurance and subsidies for low-income beneficiaries. These cost estimates do not reflect the final costs after payment reconciliation. Although Part D costs to Medicare declined in 2017 by a statistically significant $10 (CI: $1–$19), this result was not robust to sensitivity testing, and a similar change was not observed for 2018.

Beneficiary Enrollment

The number of beneficiaries eligible to participate in VBID in 2019 rose to 111,012, from 98,740 beneficiaries in 2017. In each of the years we evaluated, ***about 60 percent of all eligible beneficiaries participated in the model test***. ***Participation rates were lower (24–31 percent) among eligible beneficiaries in POs with participation requirements,*** depending on the year. Nonetheless, characteristics of participating beneficiaries resembled those of beneficiaries who did not participate.

In 2018 interviews for the evaluation, six of the ten POs participating in that year stated that beneficiary participation was lower than they had expected. This finding may be related to POs' initial difficulties in reaching out to beneficiaries, making them aware of VBID, and engaging them in CM/DM activities. We also heard from a sample of beneficiaries we interviewed that many were struggling with comorbid illnesses that were not within the focus of the VBID interventions. Additionally, many beneficiaries were skeptical of VBID in general, because they

did not feel as though they needed it or because other conditions were more important for them to deal with at the time. Although most POs reported that beneficiary participation met their expectations in 2019, participation rates did not grow between 2018 and 2019, indicating that POs may have adjusted their expectations.

Beneficiary Experiences

Beneficiary awareness of VBID interventions was low, unless prompted. In our interviews with 100 beneficiaries, almost two-thirds of respondents were aware of VBID, though few recognized the VBID interventions without further prompting, such as descriptions of the benefits offered.

Many beneficiaries who reported being enrolled in MA VBID did not seem to understand how the program works. Some beneficiaries we interviewed demonstrated confusion about what benefits were available and whether they had, in fact, received benefits for which they were eligible.

Beneficiaries we interviewed also had different perspectives on the usefulness of VBID benefits offered to them. Some beneficiaries reported that reduced copays were valuable and helped them get the care they needed, but others cited factors such as wait times and transportation as more consequential barriers to care than copays. The supplemental benefits offered by one PO garnered a mixed response from its beneficiaries, yet a telehealth benefit offered by a different PO prompted an overall favorable response. Beneficiaries were split as to the helpfulness of the CM/DM activities that POs offered.

We found no effect of VBID on overall enrollment in participating plans.

Other Outcomes

In addition to health care utilization, cost, and beneficiary enrollment, we analyzed VBID's effect on patient experience, health care quality, and health outcomes. For most of these outcomes, data were not available beyond 2017. Because of this data lag and because these types of outcomes would be unlikely to experience short-term changes, we did not expect to find statistically significant changes in these outcomes.

Patient experience of care may not have been affected by VBID. We detected a very small increase (1.44 points [CI: 0.09–2.79]) in beneficiary-reported care coordination, as measured in the MA & PDP Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey. It is important to note that the survey is not necessarily fielded to a representative sample of VBID enrollees and the change represents an increase of less than 2 percent. We also considered data from other measures on the MA & PDP CAHPS survey, including beneficiary-reported experiences with customer service, doctor communication, timely access to care, and overall rating of the health plan, but none indicated an effect of VBID on these measures of patient experiences with care.

We found no evidence that VBID affected any of the quality outcomes that we analyzed, including outcomes measured at both the contract level and the beneficiary level. An MA contract is a suite of PBPs offered by the same PO and is relevant for PBP quality because Star Ratings measures are calculated at the contract level. We analyzed four contract-level quality measures: (1) the overall Star Rating; (2) a targeted conditions quality index that considered quality outcomes across a range of VBID-targeted conditions, including hypertension, diabetes, and CHF; (3) a diabetes-specific quality index; and (4) a general medical quality index focused on quality measures not targeted by VBID but that might nevertheless be influenced by increased interactions with PCPs and care managers. VBID had no statistically significant effects on any of the contract-level quality measures that we considered. This analysis is limited by two main factors. Because of the small sample size, we included all participating PBPs in the regressions, regardless of whether their PO's VBID intervention targeted the specific quality measure considered. In addition, the types of measures considered may not be likely to experience detectable fluctuations over the short course of study in this evaluation.

We considered three condition-specific adherence outcomes at the beneficiary level: (1) adherence to diabetes medications, (2) adherence to hypertension medications, and (3) adherence to cholesterol medications. With a greater sample size at the beneficiary level, we were able to limit the treatment sample for each regression to beneficiaries in PBPs that targeted specific types of services. However, VBID had no effect on any of the condition-specific quality measures considered. We also considered two general quality measures that allowed for comparison across VBID-participating POs: The percentage of eligible medication adherence measures satisfied and adherence to recommended breast cancer screening (which may improve with increased primary care or better CM/DM). We found no effect of VBID on either of these general measures.

VBID has not had a measurable effect on the health outcomes we analyzed. We considered a range of health outcomes including empirically validated measures of self-reported health status, days in poor health, counts of activity limitations, risk scores, and mortality. We found no evidence that VBID had an influence on any of these outcomes over the duration of the follow-up period, generally one year.

Value-Based Insurance Design Model Next Steps

MA VBID is the first test of value-based health care coverage implemented in MA. In 2017, CMS granted a limited waiver of MA uniformity requirements, allowing POs a great deal of flexibility in designing their own clinically nuanced benefits based on their beneficiaries' needs, within the broad parameters of the model test. Eleven POs participated in the MA VBID model test between 2017 and 2019. Each participating PO created a unique VBID intervention, with the majority of POs requiring beneficiaries to complete certain requirements, such as participating in CM/DM activities, to receive VBID benefits.

VBID appears to be working mostly as intended for MA participants. As a necessary first step toward achieving VBID's long-term goals of improved health care quality and reduced costs, beneficiaries increased their use of high-value services. We found increases not only in the high-value services targeted directly by POs through reduced cost-sharing, but also in general utilization categories, such as primary care and specialist visits, across all POs. These changes in general utilization suggest that the care management approaches adopted by most POs may be working as intended, encouraging beneficiaries to use valuable services. The VBID model test was also associated with improvements in some cost-related outcomes, such as reductions in Part D bids and Part D OOP costs in some years. However, we also found a few unintended effects. Notably, VBID was associated with an increase in the beneficiary use of ACS ED visits and ACS inpatient stays, and MA premiums increased.

Our evaluation could only assess the effect of the model test in the short term. For most measures of VBID's effect on health care quality and outcomes, data were not available for analysis beyond 2017. Therefore, more time is needed to gauge VBID impacts on downstream outcomes such as health status, which may require sustained exposure to high-value care before meaningful effects become detectable. It is also possible that some unintended effects—such as increases in ACS ED and inpatient utilization—could subside as POs and beneficiaries gain more experience with the model.

The MA VBID designs varied widely. Although our analysis speaks to the overall effects of the MA VBID model test given the range of interventions adopted by MA VBID participants in the early years of the model (2017–2019), the results may not generalize to other VBID interventions, and it is difficult to conclude which specific components of POs' VBID designs had the biggest impact. These limitations are an important consideration as CMS embarks on VBID in the coming years (2020–2024), allowing new flexibilities and involving a different set of participating POs. As implemented by 2017–2019 participants, VBID appears to have moved utilization in the intended direction, but there is much unexplored territory going forward. Results are likely to change as the MA VBID model evolves and model participants implement new VBID benefit designs.

Acknowledgments

We are grateful to Michael Chernew, Cheryl Damberg, Mark Fendrick, Matt Maciejewski, and Brian Vegetabile for their helpful reviews of this report. We also thank staff from CMS's Office of the Actuary for their careful reviews of the cost analyses presented in this report. We appreciate the assistance of Brad Davis, Alison Pool, Robert Lang, and Marshall Bell of Wakely Actuarial for their input regarding the bids and spending analyses. We thank Ingrid Maples for excellent administrative assistance.

Abbreviations

ACS	ambulatory care sensitive
ACSC	ambulatory care sensitive condition
ADL	activities of daily living
AR-2	order 2 autoregressive
BPT	bid pricing tool
CAD	coronary artery disease
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CEM	coarsened exact matching
CHF	congestive heart failure
CI	confidence interval
CM	care management
CMMI	Center for Medicare & Medicaid Innovation
CMS	Centers for Medicare & Medicaid Services
COPD	chronic obstructive pulmonary disease
CT	computed tomography
CY	calendar year
DM	disease management
DME	durable medical equipment
ED	emergency department
ESRD	end-stage renal disease
FFS	fee-for-service
GDCA	gross drug cost above out-of-pocket threshold
GEE	generalized estimating equation
GLM	generalized linear model
HCC	hierarchal condition category
HCPCS/CPT	Healthcare Common Procedure Code System/Current Procedural Terminology

HEDIS	Healthcare Effectiveness Data and Information Set
HOS	Health Outcomes Survey
IADL	Independent activities of daily living
IDR	Integrated Data Repository
IT	information technology
LICS	low-income cost-sharing subsidy
LIS	low-income subsidy
MA	Medicare Advantage
MAO	Medicare Advantage Organization
MA-PD	Medicare Advantage and Part D
MARx	Medicare Advantage and Prescription Drug
MCS	mental component score
NDC	National Drug Code
NPI	National Provider Identifier
NPPES	National Plan and Provider Enumeration System
OACT	Centers for Medicare & Medicaid Services Office of the Actuary
OOP	out of pocket
OTC	over-the-counter
PBP	plan benefit package
PCP	primary care provider
PCS	Physical Component score
PDE	prescription drug event
PDP	prescription drug plan
PO	parent organization
PPO	preferred provider organization
ROI	return on investment
RxHCC	Prescription Hierarchal Condition Category
SE	standard error

SNF	skilled nursing facility
SNP	Special Needs Plan
VBID	value-based insurance design
VR-12	Veterans RAND 12-Item Health Survey
WHP	wellness and health care planning
WLS	weighted least squares

1. Introduction

Value-based insurance design (VBID) refers to health insurers' efforts to structure cost-sharing and other health plan design elements to encourage enrollees with chronic conditions to use the services that can benefit them the most. Examples of these incentives include lower prescription drug copayments for people with chronic conditions to increase medication adherence and reduced patient cost-sharing for visits with high-value providers to spur people to get needed care. Most of the literature on VBID focuses on interventions that have been implemented in the employer-sponsored insurance market, and studies have shown that some insurance designs, especially those that address cost-sharing for prescription drugs (Choudhry et al., 2011; Choudhry et al., 2010; Choudhry et al., 2014; Gibson et al., 2011; Maciejewski et al., 2014; Yeung et al., 2017) in concert with innovative care management and disease management (CM/DM) approaches (Chernew et al., 2008; Gibson et al., 2011; Peaslee et al., 2016), can improve disease control and reduce costly complications.

Until recently, most VBID studies were limited to relatively short periods of patient follow-up, and VBID had not been implemented or tested among individuals aged 65 and older. However, under the authority granted to the Centers for Medicare & Medicaid Services (CMS) under section 1115A of the Social Security Act, the Center for Medicare & Medicaid Innovation (CMMI) within CMS is currently conducting a Medicare Advantage (MA) VBID model test that takes place over multiple years. First implemented in 2017, the test allows Medicare Advantage Organizations (MAOs) and their parent organizations (POs) to alter specific benefit design elements within their plan benefit packages (PBPs) to encourage enrollees with targeted conditions to use high-value services and providers (see text box for key MA-related terms). Specific aspects of the test, such as the health conditions for which enhanced benefits may be applied and the states in which the test may run, have been expanded over the course of the model test. In 2019, CMS announced significant changes to the VBID model test, which

MEDICARE ADVANTAGE-RELATED TERMS

Throughout this report we refer to MA entities, all of which play a role in facilitating delivery of benefits to each beneficiary.

- A **parent organization (PO)** is a legal entity with a controlling interest in one or more MA Organizations.
- An **MA Organization (MAO)** is an insurer that offers MA plan benefit packages.
- A **contract** is a suite of PBPs offered by the same MAO and governed by the same agreement with CMS.
- A **plan benefit package (PBP)** is a specific MA insurance plan such as a specific HMO or preferred provider organization (PPO).

broadened participation to all 50 states, added a requirement to include wellness and health care planning (WHP) as part of VBID, and expanded the types of VBID designs that could be implemented. These changes came into effect in January 2020.

CMMI has contracted with the RAND Corporation to evaluate the VBID model test to determine whether adding financial incentives to use high-value care can increase the use of high-value services, improve quality of care, improve beneficiary health outcomes and experiences, and reduce overall costs of care. The first evaluation report (links here and here) focused on 2017 (Year 1 of the model test) and used findings from qualitative interviews with leaders from participating and nonparticipating PBPs, as well as analyses of characteristics of eligible and noneligible beneficiaries, to identify barriers and facilitators to VBID participation and describe early implementation experiences (CMS, 2017; Eibner et al., 2018). This report serves as the final evaluation for the first three years of the VBID model test, reporting on POs' experiences in 2018 and 2019 and quantifying the effects of VBID through a regression approach. We report on POs' experiences with the VBID evaluation and changes they have made to their VBID approaches. In addition, we present quantitative results that analyze the effect of the VBID model test on key outcomes including utilization, patient experience, quality of care, health outcomes, plan enrollment, plan bids, total medical spending, and beneficiary costs.

Value-Based Insurance Design Model Test (2017–2019)

In 2017, the first year of the model test, VBID was available to POs in seven states (Arizona, Indiana, Iowa, Massachusetts, Pennsylvania, Oregon, and Tennessee) that met a number of the following criteria (CMS, 2015):

- The plan type must have been a health maintenance organization, health maintenance organization with a point-of-service option, or local preferred provider organization (PPO).
- The plan must not have been a special needs plan (SNP), Medicare-Medicaid plan or other demonstration plan, regional PPO, cost plan, private fee-for-service (PFFS) plan, medical savings account plan, or employer group waiver plan.
- All or part of the plan's service area must have been situated within one of the model test states.
- The plan must have had at least 2,000 enrollees in a model test state.
- At least 50 percent of the plan's total enrollment must have resided in the model test states.
- The plan must have been offered in no more than two states.
- The plan must have been offered in at least three annual coordinated election (open enrollment) periods prior to the open enrollment period for calendar year (CY) 2017.
- The organization offering the plan must not have been under sanction by CMS as described in 42 C.F.R. §422.750 and 42 C.F.R. §423.750.
- The organization offering the plan must not have been an outlier in CMS's Past Performance Review.

- The plan must have had at least a three-star overall quality rating for CY 2015 (plans that are not rated because of newness or low enrollment did not qualify).
- The plan must not have been a "consistently low performing" icon on the Medicare plan finder.
- The plan's proposed intervention must have met the VBID design criteria, including targeting patients with the allowed conditions and adhering to the four permissible intervention approaches.

POs were permitted to offer VBID benefits to all enrollees within a PBP who were diagnosed with one or more of seven conditions: chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), coronary artery disease (CAD), diabetes, hypertension, mood disorders, and past stroke. Intervention designs could use one or more of the following four approaches:

- Reduced cost-sharing for high-value services
- Reduced cost-sharing for high-value providers
- Reduced cost-sharing contingent on beneficiary participation in CM/DM
- Provision of additional supplemental benefits, such as transportation services, nutritional services, or postacute care.

All four approaches reduce financial barriers to care, but the third approach also aims to engage beneficiaries in actively managing their condition(s).

Changes in 2018 and 2019

CMS changed VBID eligibility requirements in both the second and third years of the model test (2018 and 2019), and the number of POs that participated changed as well. Changes to the eligibility criteria primarily involved adding new states and new health conditions. In addition to changes in eligibility, CMS reinterpreted a key MA rule, the Uniformity Requirement, which had implications for VBID participation. Below, we describe each of these changes in more detail.

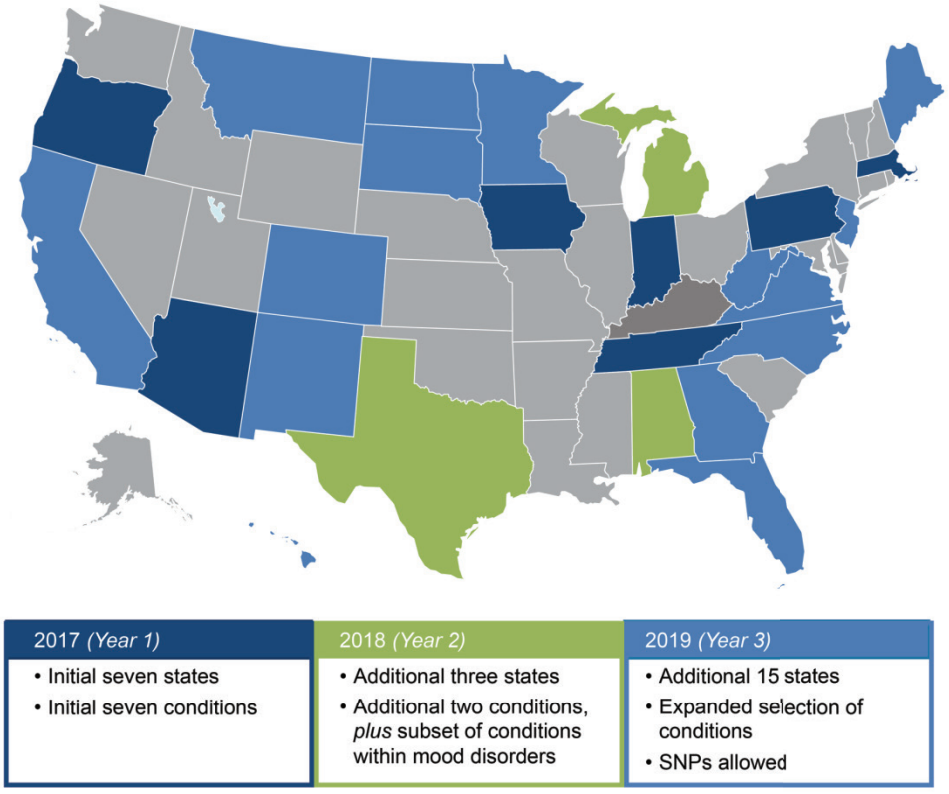
Expansion to New States

Figure 1.1 shows how eligibility for the VBID model test was expanded over time:

- Three states were added in 2018: Alabama, Michigan, and Texas.
- Fifteen states were added in 2019: California, Colorado, Florida, Georgia, Hawaii, Maine, Minnesota, Montana, New Jersey, New Mexico, North Carolina, North Dakota, South Dakota, Virginia, and West Virginia.

Although 18 states additionally became eligible to the VBID model test between 2017 and 2019, only two additional POs across two states joined the model test: A PO from Michigan joined in 2018 and a PO from Arizona joined in 2019. Arizona has been a VBID-eligible state since the first year of the model test. In addition, in 2019, a previously participating PO extended VBID to a PBP in West Virginia.

Figure 1.1. VBID Eligibility over Time

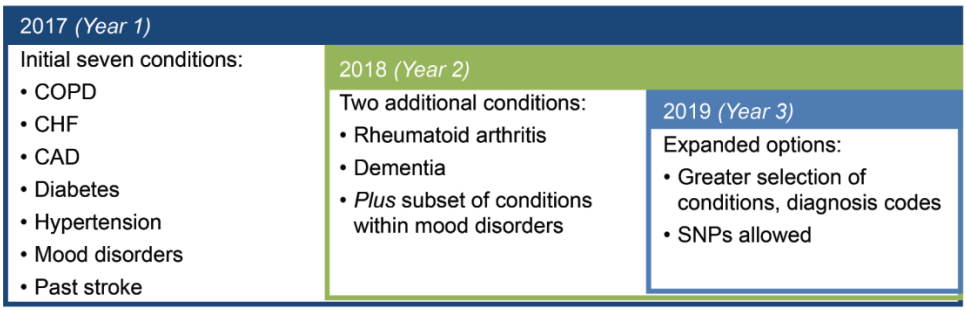


Expansion to New Conditions

The following new conditions were also added over time (Figure 1.2):

- In 2018, POs could include rheumatoid arthritis and dementia, and they were also permitted to select a subset of conditions within the mood disorder category.
- In 2019, POs were given more flexibility to identify beneficiaries based on any condition or to select their own set of diagnosis codes within the previous nine eligible conditions (e.g., depression within the mood disorders category). CMMI also allowed POs to enter SNPs into the model test.

Figure 1.2. VBID-Eligible Conditions over Time



Despite the ability to cover new conditions, no POs expanded VBID to additional conditions that were not eligible under the initial 2017 rules.

Change to the Interpretation of the Uniformity Requirement

In April 2018, CMS changed its interpretation of the long-standing MA uniformity rule, which had previously been interpreted as requiring that the same benefits be offered to all beneficiaries enrolled in a given PBP regardless of health status, conditions, or other factors. CMS also increased the flexibility in types of supplemental benefits that could be offered. This change in interpretation of the rules allows VBID-style flexibility outside the VBID model test, potentially affecting POs' willingness to participate in the model. For example, POs can now offer reduced cost-sharing for services such as specialist visits to subsets of enrollees with specific chronic conditions, without participating in the model test. The new uniformity flexibility does not apply to Part D, so POs must remain in the model test to offer VBID for drug benefits.

Participating Parent Organizations

A total of 11 POs participated in the model test at some point between 2017 and 2019: Aetna, Inc., Blue Cross and Blue Shield of Massachusetts, Inc., Blue Cross and Blue Shield of Michigan, Fallon Community Health Plan, Geisinger Health System, Highmark Health, Independence Health Group, Inc., Indiana University Health, Tufts Associated HMO, Inc., UPMC Health System (University of Pittsburgh Medical Center), and Veritage, LLC (doing business as Blue Cross Blue Shield of Arizona). Nine of these POs participated in 2017, with one new PO joining in 2018 and another joining in 2019. One of the original POs left the model test in 2019 to pursue similar VBID-style flexibilities under the new uniformity flexibility. In net, nine POs participated in 2017 and 10 participated in 2018 and 2019.

POs that participated in VBID were heavily concentrated in Pennsylvania (five participants) and Massachusetts (three participants); only one PO in Arizona, Indiana, Michigan, and West Virginia participated in VBID. Most participating POs were state-based insurance providers; there was one national participant. Of the 11 POs that participated in VBID at some point in Years 1 through 3, five were Blue Cross and/or Blue Shield affiliates.

Throughout this report, we refer to specific POs using letters (e.g., PO A, PO B) to protect their confidentiality. For continuity, we retain the same labeling approach and letter assignments as used in the first annual MA VBID evaluation report.

Recap of Findings from the First Evaluation Report

In our first evaluation report (Eibner et al., 2018), we analyzed the VBID experiences among POs based on interviews with representatives from participating and nonparticipating POs; described VBID interventions that POs implemented; presented statistics on model uptake

among POs and beneficiaries; and assessed the early effect of VBID on plan bids, revenue to plans, and enrollment. Nine POs participated in the VBID model test in 2017, and the intervention designs they implemented targeted four of the seven allowed conditions (COPD, CHF, hypertension, and diabetes). POs that participated in the first year of the model test described a desire to experiment with benefit design and less risk aversion than those that did not participate. Most participating POs lowered cost-sharing for high-value services as part of their VBID design strategies, and seven of nine participating POs required beneficiaries to complete requirements such as agreeing to participate in CM/DM to receive VBID benefits.

Benefit designs used by participating POs differed from approaches previously tested in the literature. Only two POs reduced cost-sharing for drugs—the only VBID approach for which the literature provides robust support. Two POs used rebates to reimburse beneficiaries for cost-sharing, rather than reducing copays at the point of service. Rebates previously have not been tested in VBID, and because beneficiaries must pay full cost-sharing up front, some beneficiaries may be deterred from seeking high-value care. However, POs reported that rebates may be more rewarding for beneficiaries and were easier to administer because they did not require coordination with providers.

The first year of the evaluation revealed both implementation facilitators and barriers. Barriers to implementing VBID included challenges associated with setting up different benefit designs for participating and nonparticipating beneficiaries, CMS restrictions on communications about VBID to current and prospective enrollees (which CMS gradually changed over the course of the current model test), the need to educate staff about the program, and poor health literacy among beneficiaries. Facilitators included keeping VBID designs simple, naming a dedicated staff lead, enabling communication among staff members involved in VBID implementation, and leveraging prior experiences, such as from VBID implemented in commercial plans.

Most POs (7 of 9) required beneficiaries to complete requirements to receive VBID benefits, and not all eligible beneficiaries participated. Across all POs, 61 percent of eligible beneficiaries participated in VBID, though this rate varied considerably among POs, from a low of 7 percent in one PO to a high of 100 percent in POs that did not have participation requirements. Participation requirements may have deterred some beneficiaries from joining the model test. In addition, low participation may have been related to low awareness of the model: Less than 10 percent of eligible beneficiaries who responded to the Medicare Advantage and Prescription Drug Plan Consumer Assessment of Healthcare Providers and Systems (MA and PDP CAHPS) survey reported having been offered reduced cost-sharing or extra benefits because of a health condition. Some POs reported that marketing restrictions reduced beneficiary awareness.

Enrollment in VBID-participating plans did not change in the first year, nor did VBID have a measurable effect on 2017 PBP bids. Data on other outcomes, including utilization, prescription drug fills, health plan quality, and beneficiary health status, were not available at the time of the first annual report writing.

Highlights of Relevant Findings from the First Evaluation Report

- MA VBID approaches differed from previously tested models.
- Barriers to implementation included managing two sets of benefits, navigating CMS marketing, educating staff, and addressing poor health literacy among beneficiaries.
- Facilitators included keeping VBID designs simple, enabling communication across departments, and leveraging prior experiences with VBID.
- Low awareness of the model test and PO participation requirements may have reduced participation among eligible beneficiaries.

Value-Based Insurance Design Model Test Moving Forward (2020 and Future Years)

In January 2019, CMS announced a set of updates that added programmatic flexibilities to the VBID model test effective January 2020. These flexibilities will allow POs to customize the additional benefits based on chronic condition or socioeconomic status (e.g., identified based on receipt of the low-income Part D subsidy or dual eligibility with Medicaid); allow POs to offer higher-value rewards or incentives (including rewards and incentives connected to the Part D benefit); and allow POs to offer some services remotely, so long as the telehealth option does not supplant in-person options. Starting in 2020, model participants will be required to include WHP as part of the model. In addition, POs will be permitted to offer the Medicare hospice benefit in the MA benefits package through VBID beginning in 2021 (CMS, 2019a). This is CMS's first ever test of carving the Medicare hospice benefit into the MA program.

This report does not focus on changes beginning in 2020 of the model test.

Evaluation Approach

Our evaluation of the VBID model test takes a mixed-methods approach by integrating primary qualitative data with secondary quantitative data to assess the model test's effects on key outcomes. This approach allows us to observe, from multiple angles, the experiences of POs, beneficiaries, and providers with the model test and develop a more complete picture of the potential benefits and drawbacks of VBID in the Medicare population. Below, we briefly describe how we gathered and analyzed data for this evaluation. A more detailed discussion of the methods can be found in the appendixes.

Interviews with Participating Parent Organizations

We conducted a series of annual semistructured interviews with the representatives of participating POs. These interviews were conducted either in person or by phone between June and September 2018 and then between June and September 2019. On average, we interviewed four representatives from each PO; the number of interviewees, however, ranged

from one to nine individuals. The PO representatives we interviewed held a variety of positions, including Medicare product specialists, Medicare compliance officers, actuarial directors, directors of regulatory affairs, CM directors and staff, informatics specialists, and medical directors of government programs. Although the interview protocol for the two new POs was similar to the one we used for the first evaluation report, the interview protocols for returning plans focused primarily on such topics as changes made to VBID intervention designs, beneficiary participation requirements and experiences, approaches used to enroll and communicate with beneficiaries, implementation experiences, intervention uptake, expected intervention outcomes, and thoughts about changes to the VBID model test. POs that decided not to continue their VBID participation were also asked about the reasons for leaving the model test. More detailed information about these interviews, including our approach to qualitative data analysis, is presented in Appendix A.

Interviews with Value-Based Insurance Design-Eligible Beneficiaries

We conducted semistructured telephone interviews with 100 VBID-eligible beneficiaries across all ten POs participating in 2018. Our sample included all eligible beneficiaries, regardless of whether they participated in the VBID model test (i.e., our sample included participating beneficiaries as well as eligible beneficiaries who opted out of VBID or who had not completed participation requirements). We interviewed 75 eligible beneficiaries who were reported by POs as having participated in the model test and 25 who were reported by POs as not having participated. The main purpose of these interviews was to learn about beneficiary awareness of and thoughts about VBID benefits. We created two versions of the interview protocol: One for those beneficiaries who participated in VBID and another for those who did not. Both versions included general questions about Medicare and MA, clinical conditions that pertained to the interview participant, barriers to care, and VBID awareness. A version of the protocol for VBID-participating beneficiaries also included questions about various VBID benefits, participation requirements, attitudes toward VBID, and expected outcomes. The version for nonparticipants included questions about reasons for not joining VBID and things that could motivate them to participate. See Appendix B for methodological details of these interviews.

Interviews with Providers

To learn about providers' experiences with the VBID model test, we conducted semistructured phone interviews with 12 clinicians (five primary care providers [PCPs] and seven cardiologists) whose patients participated in the PO I VBID intervention. We chose to focus on PO I and conduct a case study of its VBID intervention because this PO was the only model participant that designed its VBID intervention to include telehealth and hence directly affected how providers deliver care to their patients. Conducted between March and April 2019, these provider interviews focused on topics such as understanding underlying unmet needs and

barriers to accessing care; patient motivation to engage in CM/DM; familiarity with the telehealth intervention and with VBID more broadly; perceived advantages, limitations, and burdens of participating in the telehealth program; and perspectives on future telehealth interventions for Medicare beneficiaries. See Appendix C for details about the PO I case study and provider interviews.

Descriptive Analysis of Centers for Medicare & Medicaid Services Data

POs that offer VBID through the model test were required to submit information on beneficiary participation to the Medicare Advantage and Prescription Drug (MARx) reporting system. We used these data to calculate the number of VBID-eligible beneficiaries in participating POs, the share of VBID-eligible beneficiaries who participated in the model test (versus opting out or not completing participation requirements), and changes over time in participation rates.

We also used descriptive analysis to provide background information on characteristics of VBID-participating POs and beneficiaries, and to describe trends on outcomes that could be affected by the VBID model, such as utilization, plan bids, health care quality outcomes, and costs. These descriptive analyses cannot be interpreted as demonstrating the causal effect of VBID on the outcomes of interest, because they do not adjust for trends over time or other characteristics. However, they provide context for understanding our regression analyses, described next.

Difference-in-Differences Regression Analyses

We used difference-in-differences regression models to estimate whether POs that participated in VBID and their eligible beneficiaries experienced changes in outcomes relative to a matched comparison group. Our analyses estimate how POs' participation in the VBID model test affected outcomes. For most analyses, we pool all VBID-participating POs and beneficiaries (and their matched comparators) into a single regression. As a result, the "treatment" effect is generally exposure to any VBID intervention implemented by a participating PO, rather than exposure to a specific VBID design.

To select matched comparators, we used a tiered approach, in which we first selected matched comparison PBPs from ten states that were not exposed to (i.e., eligible to participate in) VBID during the model test. We then identified matched comparison beneficiaries from all individuals within these comparison PBPs. The average absolute standardized difference in characteristics, a measure of the quality of the match, was 0.18 for the PBP-level match and 0.03 for the beneficiary-level match (lower values indicate a closer match).

After identifying matches, we ran difference-in-differences regression models, which compared trends in key outcomes for the treatment group relative to matched comparators. An underlying assumption of the difference-in-differences methodology is that time trends in the outcome would have been the same, or "parallel," for the treatment and comparison groups

in the absence of the VBID intervention. If trends diverged after the model was implemented in 2017, and if the parallel trends assumption was satisfied, we can infer that VBID may have had a causal effect on the outcome under consideration. We tested the parallel trends assumption using data from 2014 to 2016, the period before VBID started. We analyzed many outcomes, and the parallel trends assumption held in most cases, though not all. When the parallel trends assumption did not hold, we reweighted the comparison group to achieve parallel trends (see Appendix D).

The outcomes that we considered included enrollment, utilization, health care quality, health outcomes, beneficiary OOP costs for Part D, MA and Part D premiums, PBP bids, and costs to Medicare. Some of these outcomes are reported at the PBP level (e.g., enrollment, PBP bids), and others are reported at the beneficiary level (e.g., utilization, health outcomes). Our analysis of beneficiary-level outcomes took an intent-to-treat approach, where we considered effects for all VBID-eligible individuals regardless of whether they opted to participate in the model test.[1]

Our data came from a variety of sources, including MA encounter data, the prescription drug event (PDE) data, Medicare Star Ratings, and others. It takes months or often years to finalize most of these data, owing to issues such as reporting delays, quality checks, and claims adjudication. For example, the encounter data for 2017, which report utilization of health care services, were not finalized until June 2019. Because of the data reporting schedule, many of the quantitative results presented in this report focus on 2017. However, we have extended the analysis to include additional years when more recent data were available, focusing on MA PBP bids, VBID beneficiary participation as reported in MARx, outcomes from the MA and PDP CAHPS survey, and MA enrollment. Figure 1.3 shows the timeline during which data for 2018 became available. As illustrated here, final risk scores and MA encounter data for 2018 were not available for analysis at the time this report was written.

Table 1.1 provides an overview of the outcomes that we analyzed using regression analyses. Our regressions covered six outcome domains—beneficiary experiences, enrollment, utilization, health care quality, health outcomes, and cost outcomes (e.g., PBP bids and costs to Medicare). In most cases, we pooled observations from all VBID-participating POs to conduct the analyses. However, when considering utilization of VBID-targeted services, we run separate regression for each PO.

[1] The intent-to-treat approach in this case focuses on the entire eligible population, regardless of whether they received VBID benefits. This approach provides an estimate of VBID's overall impact on the targeted population, accounting for the fact that some eligible people may not complete participation requirements, may be unaware of the model test, or may opt out.

Figure 1.3. Timeline for VBID Model Test and Release of 2018 Data

[a] Plan bids are submitted before the experience year but are the basis for plan payments during that year.
[b] Data are submitted by POs to CMS on an ongoing basis, but submission timelines can vary across POs and the data are not considered final until a certain period of time after the close of the experience year.
[c] Quality measure data sets are used to construct the Star Ratings and include Healthcare Effectiveness Data and Information Set (HEDIS), Part D adherence measures, and the Medicare Health Outcomes Survey (HOS).
[d] Medicare Star Ratings are used to assist beneficiaries in making enrollment choices for the upcoming plan year; therefore, Star Ratings for a given year are released the fall before that year. Star Ratings for a given year are based on experience year data from two years before.
[e] Risk score data from the Hierarchal Condition Category (HCC) and Prescription Hierarchal Condition Category (RxHCC) data sets are released using midyear utilization data; those data are available earlier than the final data. This figure shows the release dates for final risk score data.

Table 1.1. Outcomes and Data Sources for Regressions

Outcome	Level of Analysis	Comparison Group	Data Source	Post-2016 Years Included
Beneficiary experiences				
Patient experience measures	Beneficiary	All beneficiaries in comparison plans who responded to survey	MA & PDP CAHPS	2017, 2018
Enrollment				
Total enrollment	PBP	Matched PBPs	CMS enrollment files	2017–2019
New enrollment	PBP	Matched PBPs	CMS enrollment files	2017–2019
Enrollment among those with chronic conditions	PBP	Matched PBPs	CMS enrollment files	2017–2019
Utilization				
VBID-targeted services	Beneficiary	Matched beneficiaries	MA encounter and PDE data	2017
General utilization	Beneficiary	Matched beneficiaries	MA encounter and PDE data	2017

Outcome	Level of Analysis	Comparison Group	Data Source	Post-2016 Years Included
Quality outcomes				
Star Ratings and quality indexes	Contract	Matched PBPs	Star Ratings data	2017
Quality measures from HEDIS	Beneficiary	Matched beneficiaries	Star Ratings data	2017
Quality measures from Part D, medication adherence	Beneficiary	Matched beneficiaries	Star Ratings data	2017, 2018
Health outcomes				
Mortality	Beneficiary	Matched beneficiaries	Beneficiary death data	2017
Risk scores	Beneficiary	Matched beneficiaries	HCC/RxHCC data	2017, midyear 2018
Health status	Beneficiary	Matched beneficiaries	HOS	2017
Cost outcomes				
MA/Part D bids	PBP	Matched PBPs	PBP bid information from OACT	2017–2019
Total realized costs to Medicare	PBP	Matched PBPs	PBP bid information from OACT	2017, 2018
Total plan spending on health care services	PBP	Matched PBPs	PBP bid information from OACT	2017
Total plan spending on prescription drugs	PBP	Matched PBPs	PDE data	2017, 2018
MA and Part D premiums	PBP	Matched PBPs	PBP premium data	2017–2019
Part D OOP costs	Beneficiary	Matched beneficiaries	PDE data	2017, 2018

NOTE: OACT = Centers for Medicare & Medicaid Services Office of the Actuary.

Organization of Report

In this report, we build on the first-year findings to describe how POs' experiences with VBID have changed over time; the implementation experiences of POs that have newly joined the model test; and the effect of VBID on outcomes including utilization, quality of care, health outcomes, patient experiences, enrollment, plan bids, and costs. Table 1.2 describes the goals of each of the subsequent report chapters. Chapters 2 and 3 describe POs' VBID interventions, their implementation experiences to date, and reasons for leaving the model test. Chapters 4 and 5 consider beneficiaries' participation in the intervention, their perceptions of and experiences with the VBID model, and the effects of VBID on enrollment in participating PBPs. Chapters 6–8 consider VBID's effects on key outcomes including utilization, health care quality, beneficiary health status, and health care costs. Finally, Chapter 9 summarizes our findings and offers concluding thoughts.

Table 1.2. Organizational Structure and Research Questions Addressed in This Report

Report Chapter	Goals	Methods	Research Questions
2. MA VBID interventions	To understand the VBID approaches POs implemented in 2018 and 2019	Analysis of VBID applications; interviews with participating POs	• What VBID approaches did POs implement and why? • What changes did they make over time?
3. VBID implementation experiences	To understand implementation experiences of new and returning POs and to explore why some POs left the model test	Interviews with participating POs	• What barriers and facilitators to implementation did POs experience? • How were these different from those in 2017? • Why did some POs leave the model test?
4. Beneficiary uptake of interventions	To understand the extent to which beneficiaries participated in VBID interventions	Descriptive comparison of participating and nonparticipating beneficiaries; interviews with POs and beneficiaries	• How many eligible beneficiaries opted not to participate or did not complete requirements? • Were nonparticipating beneficiaries different from participating beneficiaries?
5. Beneficiary experiences	To assess eligible beneficiaries' awareness of VBID benefits available to them	Analysis of MA & PDP CAHPS data; interviews with beneficiaries	• Were eligible beneficiaries more likely than matched comparators to report that their health plans offered VBID benefits? • What did beneficiaries think about VBID? • Did VBID affect plan enrollment among participating PBPs?
6. Effects of VBID on health care utilization	To determine how VBID affected use of health care services	Analysis of encounter data; PO interviews	• Did utilization of health care services change among participants? • How did POs perceive VBID's effect on utilization?
7. Effects of VBID on health care quality and health outcomes	To assess the effect of VBID interventions on health care quality and health outcomes	Analyses of quality metrics and data on beneficiaries' health status; PO interviews	• Did VBID affect health care quality? • Did VBID affect beneficiaries' health outcomes?
8. Effects of VBID on health care costs	To determine whether VBID affected health care costs for beneficiaries, PBPs, and/or Medicare	Analyses of beneficiary costs, plan bids, Medicare costs; PO interviews	• In what ways did VBID affect health care costs?
9. Conclusions	To highlight key implications	Synthesis of previous chapters	• How can the model be strengthened in future years?

In addition, several appendixes provide more detailed information on the methods used in this evaluation:

- Appendixes A and B describe the methods for interviews with POs and beneficiaries, respectively.
- Appendix C offers information on the telehealth case study, based on the experiences of PO I, and includes methodological details of our interviews with PO I providers.
- Appendix D covers the quantitative methods for plan and beneficiary matching, as well as parallel trends.
- Appendixes E–I describe the methods and detailed results for the effects of VBID on plan enrollment, beneficiary experiences, utilization, health care quality and health outcomes, and health care costs, respectively.
- Appendix J takes a more detailed look at the effects of utilization on medication use and adherence; we focus on POs offering Part D interventions.
- Appendix K considers whether effects on utilization outcomes differed between POs that had participation requirements for beneficiaries and POs that did not have such requirements.
- Appendix L estimates the price elasticity of demand for primary care in our sample.

2. Medicare Advantage Value-Based Insurance Design Interventions

VBID-participating POs implemented a variety of intervention designs, ranging from simple interventions that reduced copayments for a single service to multifaceted interventions that combined reduced cost-sharing, CM/DM, and supplemental benefits (Table 2.1). In this chapter, we describe the intervention designs of the 11 POs that participated in the model test between 2017 and 2019, highlighting changes they made to their designs and noting when they joined and left (if applicable) the model test. We also describe how POs communicated VBID-related information to eligible beneficiaries and providers. Data come from model test applications that POs submitted to CMS in 2018 and 2019 and interviews with POs that RAND conducted between June and September in both 2018 and 2019.

Key Takeaway Points:

- Participating POs did not make major changes to their VBID designs since the start of the model test.
- CHF and diabetes were the most commonly targeted VBID conditions, and specialist visits were the most frequently chosen high-value service.
- Seven of the 11 POs made the receipt of VBID benefits conditional on beneficiary completion of participation requirements.
- Six POs required eligible beneficiaries to enroll in VBID by either calling the PO or responding to an outreach call.
- Although all 11 POs communicated with beneficiaries about VBID via mail, only six POs conducted additional telephone outreach to increase VBID awareness. Six POs also issued new beneficiary ID cards indicating VBID participation, and two POs mentioned VBID benefits on member-only portals accessible to VBID-eligible beneficiaries.
- Although most POs included VBID-related information in their provider newsletters, only six POs conducted in-person outreach activities in either provider offices or regional medical director meetings.

Table 2.1. 2018 VBID Approach Components

POs	A	B	C	D	E	F	G	H	I	J	K	Total
Condition												
Diabetes	X	X	X					X			X	5
CHF			X			X	X	X	X			5
Hypertension				X								1
COPD		X	X		X	X						4
CAD										X		1
Intervention components												
PCP visits	X	X[a]				X	X					4
Specialist visits	X	X			X	X	X	X			X	7
Drugs				X			X			X		3
Diagnostics/ DME		X			X							2
High-value providers		X										1
Supplemental benefits		X					X[b]		X			3
Cost-sharing rebates	X		X[c]									2
Beneficiary participation requirements	Score-card[d]	CM/ DM	CM/ DM		CM/ DM	CM/ DM	CM/ DM		CM/ DM			7
Years of participation	2017–2019	2017–2019	2017–2019	2017–2019	2017–2018	2017–2019	2017–2019	2017–2019	2017–2019	2018–2019	2019	

[a] This benefit was removed for 2019.
[b] This benefit was added for 2019.
[c] PO C offered rebates for any incurred MA cost-sharing. The rebate was converted to an over-the-counter benefits debit card in 2019.
[d] "Scorecard" refers to completion of four preventive services in 2017 and 2018 and three services in 2019.

Value-Based Insurance Design Intervention Designs

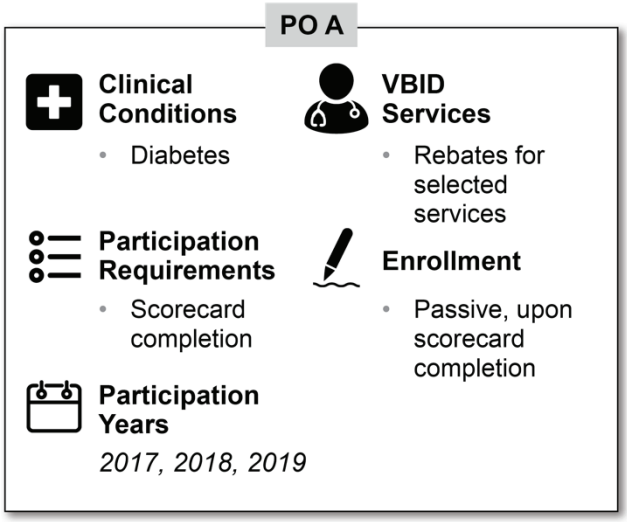
For the most part, returning POs did not change the design of their VBID interventions between 2017 and 2018. Because the 2018 applications were due in January 2017 (the same month the model test started), most returning POs stated that they did not have enough experience with VBID implementation to modify their designs. Nonetheless, two returning POs made changes to their VBID interventions in 2018. PO C combined two CM/DM activities (a self-administered personal health assessment survey and an initial call with a care manager) into one activity and added additional preventive care activities (including a flu shot) that counted toward fulfilling VBID participation requirements. PO E removed the pulmonary rehabilitation component from its VBID design and made this benefit available to all beneficiaries at no cost in 2018. It also made small changes to the procedure codes and specialist types eligible for reduced cost-sharing.

More POs made changes to their intervention designs in 2019. PO A removed the lipid test from its scorecard because it was no longer part of the HEDIS measures. PO B added cardiologists to the list of VBID-eligible specialists and removed reduced cost-sharing for PCP visits from VBID by waiving PCP copays for all beneficiaries enrolled in its plans. Because PO B designated all VBID-eligible specialists as high-value providers and waived PCP copays for all beneficiaries in its plans, its 2019 VBID benefits include reduced copays to specialists and supplemental benefits. PO C switched to an over-the-counter (OTC) benefit card instead of issuing cost-sharing rebates. Finally, PO G expanded its VBID intervention to additional PBPs and added meal and transportation supplemental benefits.

Design Snapshots

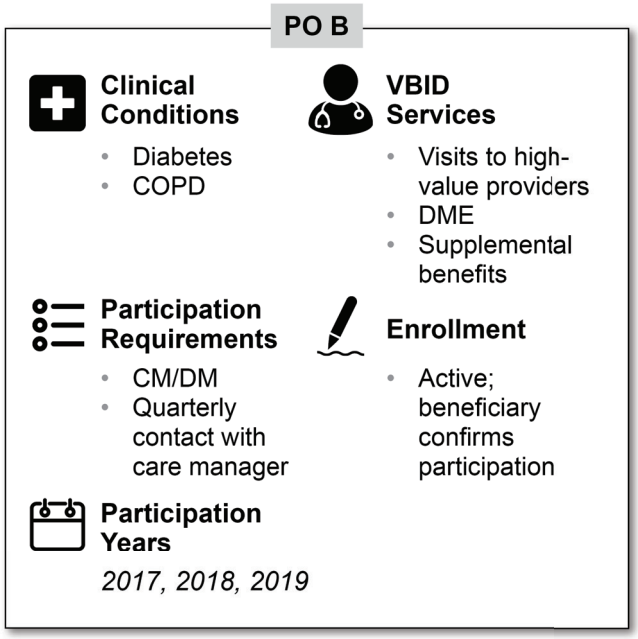
Across the intervention designs for 2018 and 2019, ***CHF and diabetes were the most commonly targeted conditions, and specialist visits were the most frequently chosen high-value service.*** Only three POs offered reduced cost-sharing for medications, the type of VBID design most consistent with the published evidence. Most model participants required beneficiaries to meet certain participation requirements before receiving VBID benefits. The addition of new model participants in 2018 and 2019 increased the overall number of participants that focused on Part D benefits, made diabetes as frequently targeted VBID condition as CHF, and added CAD to the list of clinical conditions chosen by VBID participants. Below, we describe the VBID designs of all 11 VBID-participating POs.

PO A participated in VBID during the first three years of the model test. It offered rebates (up to $200 annually) to beneficiaries with diabetes for incurred MA cost-sharing for visits to PCPs, endocrinologists, and foot care providers, as well as certain eye examinations with an ophthalmologist or optometrist. To receive reduced cost-sharing through rebates, beneficiaries were required to complete a scorecard with specific diabetes monitoring services: (1) a hemoglobin test, (2) a lipid profile, (3) an eye examination, and (4) a urine test. Although PO A did not require VBID-eligible beneficiaries to participate in CM/DM activities and all beneficiaries with a diabetes diagnosis were eligible for the program, they did not receive reduced cost-sharing unless they



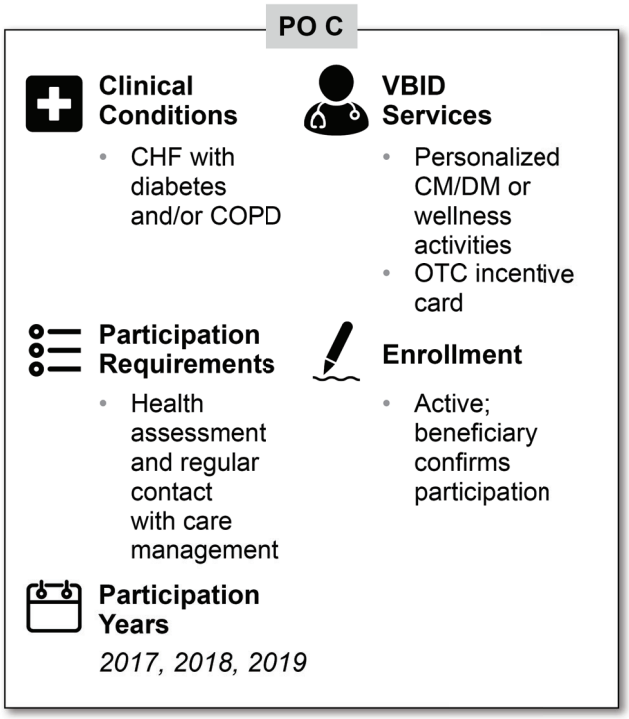
completed the required scorecard screenings and incurred eligible MA costs. Once the scorecard was complete, beneficiaries received rebates for eligible cost-sharing three times per year. Starting in 2019, PO A no longer included a lipid profile on the scorecard.

PO B participated in VBID during the first three years of the model test. In 2017 and 2018, it combined three VBID approaches: reduced cost-sharing for office visits with high-value providers, reduced cost-sharing conditional on participation in CM/DM activities, and the provision of additional supplemental benefits for beneficiaries with either diabetes or COPD. Reduced cost-sharing on services included $0 copayments for up to four office visits per year to PCPs and $10 copayments for up to four office visits per year to specialty care providers designated as high-value providers. In addition, PO B offered new supplemental benefits at no cost to VBID-eligible beneficiaries, including one diabetic retinal photograph per year, one periodontal maintenance procedure per year, periodontal scaling and root planing, and four periodontal surgical procedures during a lifetime. PO B also reduced copays for several existing supplemental benefits, such as $5 copays for transportation for up to 48 one-way trips to medical appointments per year and reduced coinsurance for diabetic testing supplies to 5 percent. Beneficiaries were required to maintain quarterly contact with a care manager to be eligible for VBID benefits. The CM/DM team determined the right level of CM/DM for each beneficiary. The CM/DM activities could include wellness coaching (e.g., for weight loss or smoking cessation); disease education programs that promote self-care; or more intensive case management that would involve a higher level of service, such as initiating regular communication between the beneficiary and his or her providers or in-home assessments to manage social and economic needs.



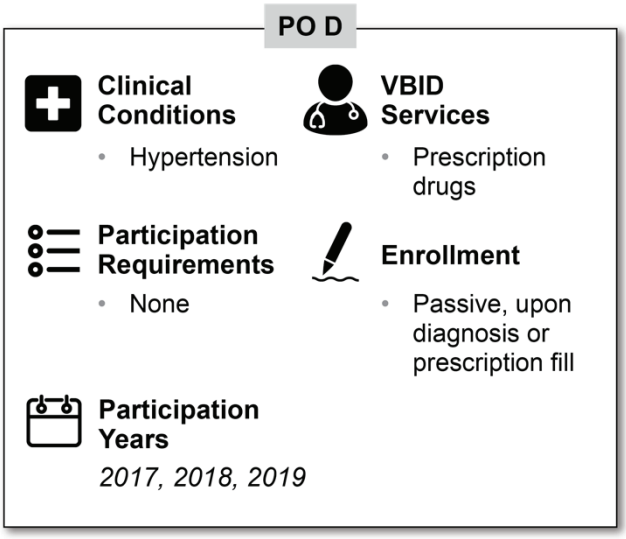
In 2019, PO B added cardiologists to the list of VBID-eligible providers and reduced PCP copays to $0 for all beneficiaries in its VBID-participating plans, thereby removing the high-value PCP benefit from its VBID intervention. PO B considered all VBID-eligible specialists to be high-value providers owing to concerns about access and availability of data to determine whether specialists were high value. Its VBID benefits in 2019 effectively included supplemental benefits and reduced copays to specialists that were offered to beneficiaries who engaged in CM/DM.

PO C participated in VBID during the first three years of the model test. In 2017 and 2018, it chose to offer quarterly rebates for incurred MA cost-sharing to beneficiaries with CHF who also had diabetes, COPD, or both if they completed up to six CM/DM activities. All beneficiaries were required to complete a health assessment survey and a personal health review with a care manager. In 2017, this counted as two separate activities, but in 2018 and 2019, the PO combined these activities into one call for which the beneficiary would earn $50. Beneficiaries could then choose up to four quarterly personalized CM/DM or wellness activities, including weight loss counseling, smoking cessation, disease education, and a flu shot (which was added in 2018). For each completed activity, beneficiaries earned $25 toward any incurred MA cost-sharing. Total rebates were capped at $150 annually. Rebate checks were mailed to beneficiaries on a quarterly basis, but only if the beneficiary incurred cost-sharing during that quarter.



In 2019, PO C replaced rebate checks for incurred cost-sharing with debit cards that could be used for purchasing OTC health-related items at participating retailers. The PO automatically reloaded the debit card when a beneficiary completed an activity. Although these funds could only be used to pay for health items, beneficiaries no longer needed to incur MA cost-sharing to receive the financial benefit.

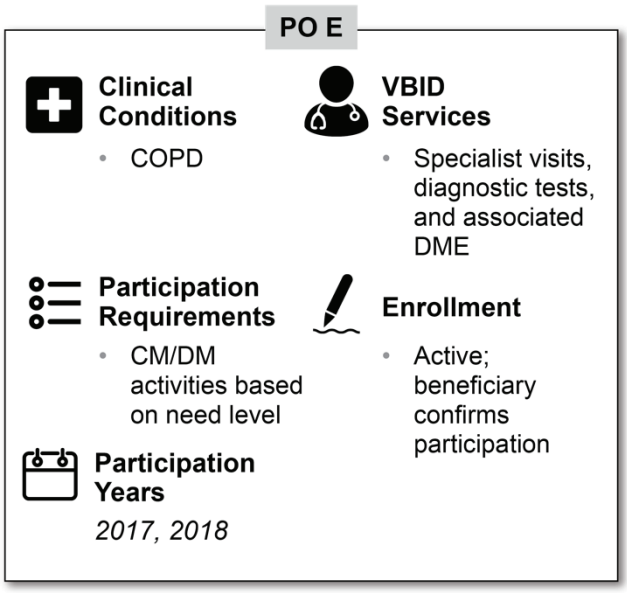
PO D participated in VBID during the first three years of the model test. It eliminated cost-sharing for hypertension drugs on tiers 1 through 3, which included waiving the deductible and any cost-sharing incurred in the prescription drug coverage-gap or catastrophic-benefit phases.[1] All beneficiaries with a



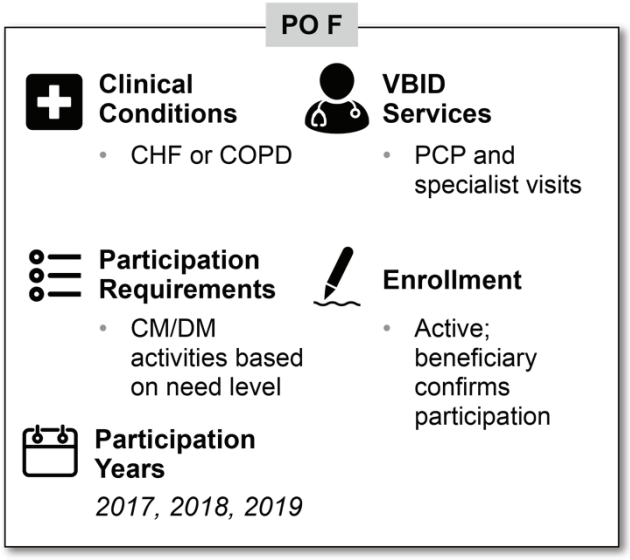
[1] Medicare Part D drug coverage uses a standard benefit design. For 2018, the standard design included a $405 deductible and cost-sharing up to $3,750, at which point the beneficiary enters the coverage gap (also known as the "donut hole"). Within the coverage gap, beneficiaries incur cost-sharing levels of 35 percent for brand-name drugs and 44 percent for generic drugs until they have paid $5,000 in OOP costs, after which the beneficiary enters the catastrophic phase and pays 5 percent of drug costs.

diagnosis of hypertension or those who filled a prescription for one of the eligible hypertension medications were automatically enrolled into VBID and received these medications at no cost.

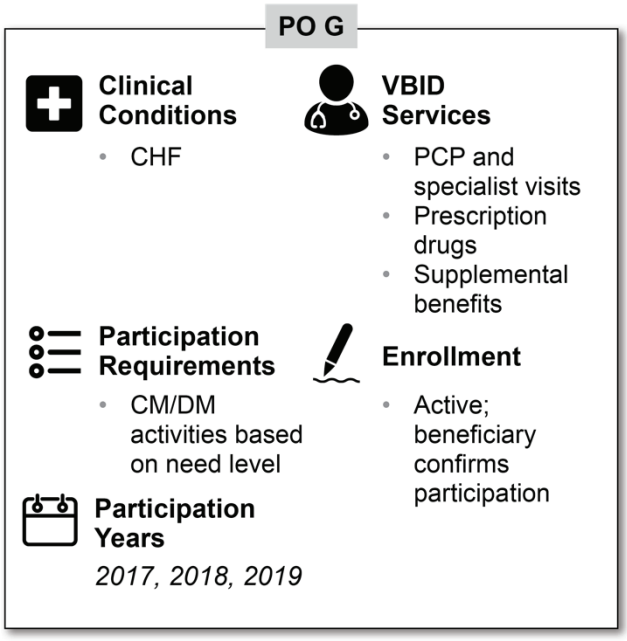
PO E participated in VBID during the first two years of the model test. It eliminated copayments for certain high-value services for beneficiaries with COPD who participated in CM/DM. PO E waived copayments for any visits to pulmonologists or cardiologists and for sleep medicine and palliative care visits, and it eli minated cost-sharing for pulmonary rehabilitation visits. PO E also eliminated cost-sharing for some tests and durable medical equipment (DME), including pulmonary function tests, sleep studies, computed tomography (CT) scans for the chest, and oxygen supplies. Beneficiaries were required to participate in CM/DM activities to receive VBID benefits. Reduced copays for pulmonary rehabilitation visits were included in 2017 but were removed from the list of 2018 VBID benefits because PO E decided to reduce cost-sharing for this service for all beneficiaries. Required activities and the level of interaction with CM/DM staff depended on the beneficiary's need level. Activities included care plan development, medication reconciliation, and disease-specific education.



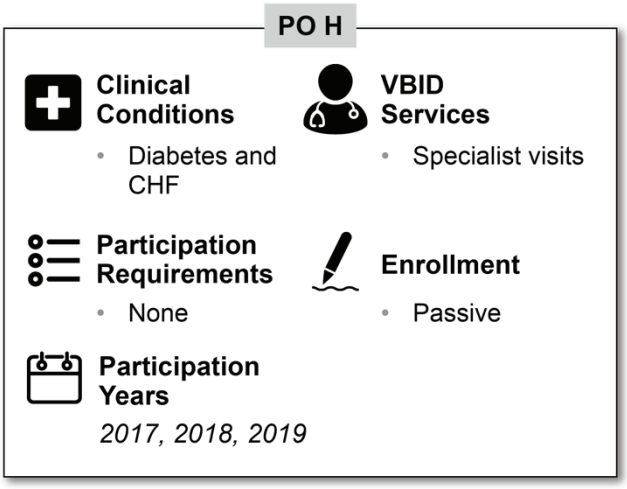
PO F participated in VBID during the first three years of the model test. It reduced or waived cost-sharing for certain high-value services for beneficiaries with CHF or COPD if they participated in CM/DM. PO F eliminated copayments for primary care visits and reduced copayments to $10 or $20 for visits to cardiologists and pulmonologists, depending on the PBP. Beneficiaries were required to participate in CM/DM to receive VBID benefits. The frequency and type of CM/DM activities were tailored to the beneficiary's need level. Activities included in-home assessments, regular calls with CM/DM staff, and disease-specific education.



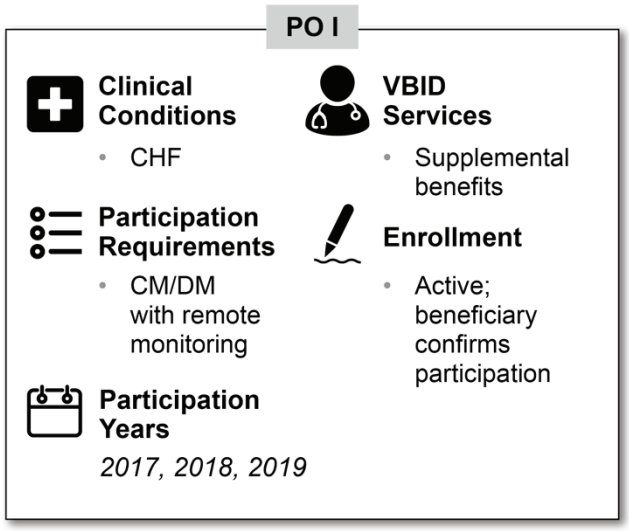
PO G participated in VBID during the first three years of the model test. It eliminated copayments for visits to PCPs and cardiologists and for select generic prescription drugs to treat CHF (for tier 1 drugs, copayments decreased to $0 from $7). Beneficiaries were required to participate in CM/DM developed specifically for the VBID model test (using components of some preexisting CM/DM programs) to receive reduced cost-sharing. The CM/DM staff created a care plan tailored to each beneficiary's level of need. CM/DM activities may have included communication with CM/DM staff, in-home assessments, an annual medication review, medication adherence, quarterly visits to a PCP, and an annual visit to a cardiologist. To enroll in VBID, eligible beneficiaries were required to confirm their willingness to participate in CM/DM in their initial conversation with a care manager. For 2019, PO G expanded its VBID intervention to three new PBPs in Pennsylvania and West Virginia, as well as added several supplemental benefits, including a meal benefit (two meals per day for 14 days, up to three times a year) and 24 one-way trips to medical appointments for its Pennsylvania plans only.



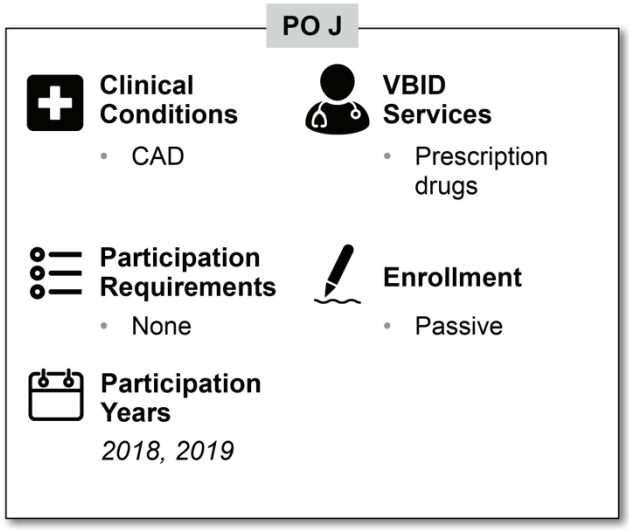
PO H participated in VBID during the first three years of the model test. It focused its VBID intervention on beneficiaries with both CHF and diabetes. This PO reduced copayments for any visit to certain specialists: cardiologists and endocrinologists (from $45/$40, depending on the plan, to $10) and podiatrists (from $45/$40, depending on the plan, to $5). There were no participation requirements to receive VBID benefits.



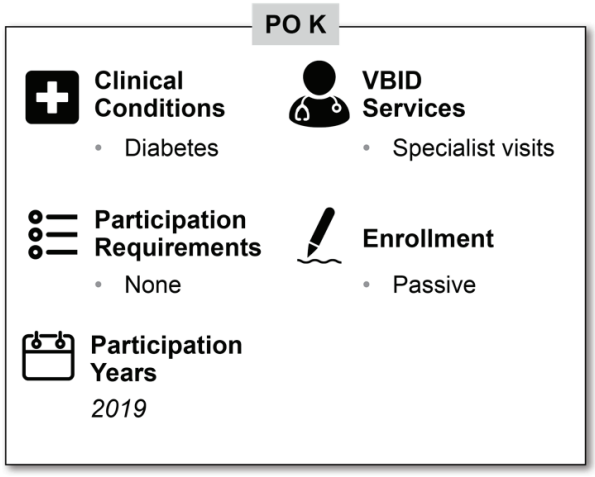
PO I participated in VBID during the first three years of the model test. It provided beneficiaries with CHF supplemental benefits such as free weight scales, blood pressure cuffs, and pulse oximeters that were remotely monitored by CM/DM staff. Telehealth nurses executed a provider-approved diuresis protocol if vital signs readings were abnormal (see Appendix C for a case study of PO I's VBID intervention).



PO J participated in VBID during 2018 and 2019. It eliminated all cost-sharing (including in the coverage gap and the catastrophic phases of Part D) for four classes of drugs used to treat CAD: antiplatelets, statins, angiotensin-converting enzymes/angiotensin II receptor blockers, and beta-blockers. There were no participation requirements for this PO.



PO K joined VBID in 2019. It eliminated cost-sharing for endocrinologist visits for beneficiaries with diabetes. VBID-participating beneficiaries were offered an opportunity to join an enhanced CM/DM program, but participation in this program was not a requirement for receiving $0 cost-sharing for endocrinology visits.



Participation Requirements and Enrollment Processes

Seven POs made the receipt of VBID benefits conditional on beneficiary completion of participation requirements. Six of these POs (B, C, E, F, G, and I) required beneficiaries to engage with a care manager throughout the year. The seventh (PO A) required beneficiaries to complete a scorecard before they could receive VBID benefits. Five POs with participation requirements (C, E, F, G, and I) also required eligible beneficiaries to enroll in VBID via a telephone conversation with their health plan. The remaining four POs (D, H, J, and K) allowed eligible beneficiaries to receive VBID benefits without enrolling or completing any participation requirements. Table 2.2 summarizes the participation requirements and enrollment procedures across all POs.

Table 2.2. Participation Requirements, Enrollment Procedures, and CM/DM Characteristics

PO	Beneficiary Participation Requirement	Enrollment Procedure	CM/DM Delivery Mode
A	Receipt of preventive screenings listed on a scorecard	Occurs on scorecard completion	N/A
B	Quarterly engagement with a care manager	PO calls beneficiary to confirm enrollment	Telephone and/or clinic
C	Fulfillment of up to six CM/DM activities	Beneficiary calls the PO to initiate	Telephone
D	N/A	None (passive)	N/A
E	CM/DM engagement frequency depends on the acuity of the beneficiary	Beneficiary calls the PO, or PO calls the beneficiary, to initiate	Telephone and/or clinic
F	Quarterly engagement with a care manager	Beneficiary calls the PO to initiate, or passive enrollment for those already in CM/DM	Telephone and/or clinic
G	CM/DM engagement frequency depends on the acuity of the beneficiary	Beneficiary calls the PO, or PO calls the beneficiary, to initiate	Telephone
H	N/A	None (passive)	N/A
I	Daily reading of blood pressure, oxygen levels, and weight using remotely monitored equipment	Beneficiary calls the PO to initiate	Telephone
J	N/A	None (passive)	N/A
K	N/A	None (passive)	N/A

These various participation and enrollment requirements may affect beneficiaries' willingness to enroll in VBID. Hypothetically, these requirements could create a situation where some healthier, lower-risk beneficiaries are willing and able to engage with care managers, and therefore enroll in VBID, whereas other lower-risk, newly diagnosed beneficiaries may not see the benefit of VBID and CM, and thus decide not to participate. Similarly, some high-risk beneficiaries in later stages of disease progression could find these requirements too burdensome and not enroll into VBID, whereas other beneficiaries with complex health conditions may think that VBID benefits are worth the cost of participating in CM.

Enrollment and participation requirements may also affect both health and cost outcomes. If beneficiaries are engaged in CM/DM over a long period of time, care managers could help slow

the course of their disease progression and save POs money in the long term. The potential for short-term savings and gains in health outcomes may be relatively low for CM/DM participation by low-risk beneficiaries. In contrast, high-risk beneficiaries who use a substantial amount of health care services could benefit a great deal from CM/DM, if it is sufficiently intensive to address complex needs. Nonetheless, POs may not realize cost savings if the majority of their participating beneficiaries are high-risk patients with high utilization rates. All POs with CM/DM requirements stressed that their ***CM/DM programs are tailored to beneficiary needs*** and noted that beneficiaries work with a care manager to determine the activities or services in which they will participate.

> *"CM/DM calls are usually conducted by a non-clinician . . . . They have a brief script that says, 'Has anything changed in your condition over the last three months? Any new medications?' So they kind of review those things. And if there have been changes or concerns and the person is expressing some need for more support, then that person could be referred to a nurse care manager for evaluation and potential enrollment in a higher-level program. If they don't seem to give any indication that they're having any problems and everything seems to be stable given the responses to the questionnaire, then they say, "Okay, I'll check in with you in three more months." And if something were to happen, like an ER visit or a hospitalization, the care management area will know about it and we'll do transition management and potential enrollment in a higher-level program if it's needed at that time."*
>
> *PO F representative on the process of customizing CM/DM activities based on beneficiary needs*

Moreover, some POs ***adjusted the frequency of contact with CM/DM staff to ensure that these requirements were not burdensome*** to beneficiaries and would not lead to beneficiaries opting out of the VBID program. For example, in 2018, PO B reduced interaction with beneficiaries whose conditions were well managed. Instead of requiring calls, the PO began sending quarterly letters informing beneficiaries about their continuous participation in VBID, stating that there was evidence of their conditions being well managed because there had been no adverse events such as inpatient or emergency department (ED) utilization. Not only was this approach designed to reduce participation burden, but it also allowed the plan to use its resources strategically. A representative from PO B stated: "We want to contact the members who need us most. So using these 'well-managed letters' is a way for us to explore having the right conversation with the right members and still contacting others to make sure that they stay on track. I think a little nudge and gold star can go a long way."

To make it easier for beneficiaries to participate, ***POs delivered CM/DM activities either by telephone or in person during a clinic visit or a home visit.*** Telephone was the most frequently used mode of engaging with beneficiaries in CM/DM activities (see Table 2.2); all six POs with a CM/DM participation requirement relied on telephone outreach. POs B, E, and F also reported counting CM/DM activities provided in-person in provider offices toward VBID participation requirements.

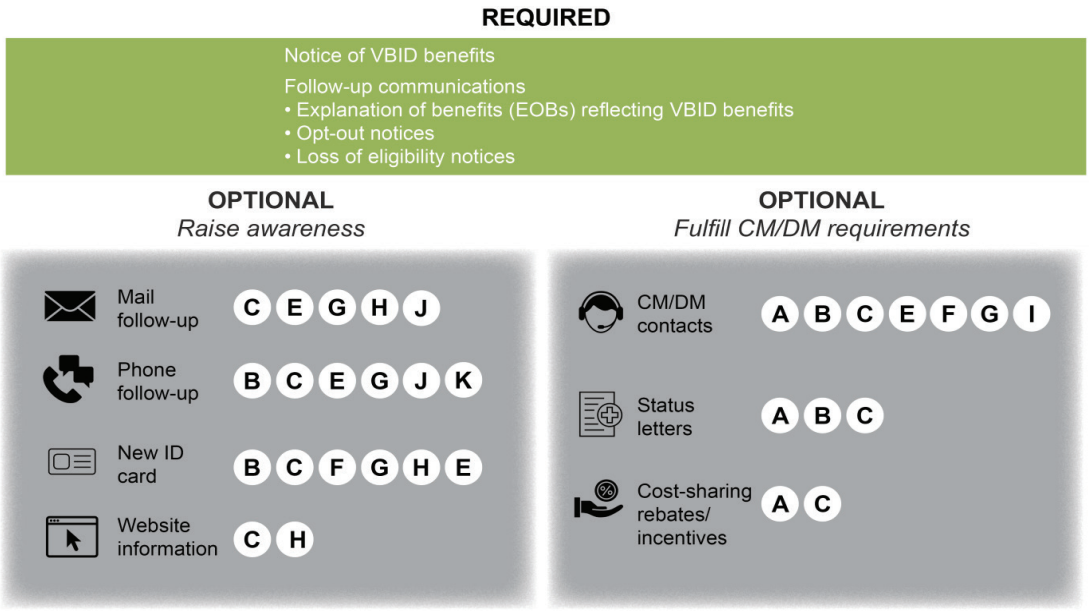
Care managers could be employees of the PO or providers' offices. For example, PO B decided that a diabetes education program offered by one of the local provider offices could count toward fulfilling participation requirements. Many POs conducted ***home visits for high-need beneficiaries as part of their regular CM/DM programs*** and also counted these visits toward VBID participation requirements. Although the majority of VBID care managers were nurses, POs also reported relying on social workers, community health workers, respiratory therapists, pharmacists, and others to deliver comprehensive CM/DM. These activities typically began with a personal health review and/or needs assessment.

Some POs, like PO E, offered a comprehensive CM/DM program that developed a care plan for each of its VBID beneficiaries, many of whom were in CM/DM before VBID. Such programs tended to resemble intensive case management designed for complex patients and often relied on in-person contact with beneficiaries. Other POs designed wellness-oriented programs intended to learn about beneficiary needs and goals (PO F), to facilitate the establishment of personal relationships between beneficiaries and care managers who could advise beneficiaries on how best to address their care needs (PO C), to encourage beneficiaries to engage in preventive activities such as flu shots (PO C), and to educate beneficiaries about their conditions and refer beneficiaries for additional care if needed (POs B, F, G, and I).

Communication with Beneficiaries

POs used different approaches to communicate with beneficiaries about VBID to meet the model test requirements, increase beneficiary awareness of the program, and help beneficiaries complete VBID participation requirements (see Figure 2.1).

Figure 2.1. Summary of PO Communication Methods with Beneficiaries



Required Communications

All POs communicated with beneficiaries about VBID via mail to satisfy the model test requirements (CMMI, 2019b). All POs were required to send notices of VBID benefits either at the beginning of the year or when a beneficiary receives a VBID-eligible diagnosis and therefore becomes eligible for VBID. The notices included the program name (if any), information about VBID benefits and CM/DM participation requirements, and opt-out procedures.

CMS also required POs to send certain follow-up communications. For beneficiaries wishing to opt out of VBID, POs had to mail a letter confirming the opt-out, describing the benefits they would no longer receive, and explaining how to opt in again. POs also were required to send a letter to the beneficiaries who asked to reenroll in VBID. If beneficiaries lost VBID eligibility because they stopped engaging with a care manager or switched to another plan, POs were required to notify beneficiaries by sending a letter. Finally, POs needed to send EOB forms that reflected the VBID benefits that had been accrued.

Optional Communication to Raise Value-Based Insurance Design Awareness

Some POs sent additional mailings to increase VBID awareness among eligible beneficiaries or to inform participating beneficiaries not meeting CM/DM requirements that their VBID benefits might end soon. PO C, for example, sent a follow-up reminder notice in January 2018 to inform beneficiaries already in VBID about their benefits. POs C, E, G, H, and J sent reminders to eligible but not yet participating beneficiaries. Six POs (B, C, E, F, G, and H) ***issued new beneficiary ID cards indicating VBID participation,*** which served as another way of raising beneficiary awareness of the program. Although these mailings may have reminded beneficiaries about VBID, PO B reported ***consolidating all of its benefits-related mailings into one quarterly packet***. Its representatives noted that too much mail could be counterproductive: "Our members complain about [excessive] mailings and phone calls."

Although mail is the main mode of communication about VBID, six POs (B, C, E, G, J, and K) also ***placed calls to beneficiaries to encourage them to enroll in VBID or raise program awareness.*** These POs felt that it was important to follow-up with beneficiaries by telephone because many of them do not open their mail. As a PO J representative explained, "For the population that isn't going to open their mail, maybe they'll respond to the people on the phone." Telephone outreach required POs to dedicate staff and resources; some participating POs, including POs F and I, cited this as a reason they did not take this approach.

Only two POs (C and H) reported describing their VBID benefits on a member-only portal accessible to VBID-eligible beneficiaries as a way of raising beneficiary awareness.

Optional Communication Related to Case Management/Disease Management and Rebates

Six POs with CM/DM participation requirements periodically reached out to beneficiaries to help them fulfill these requirements. Conducted mostly by phone but sometimes in person, these

contacts were necessary for beneficiaries to receive reduced cost-sharing or rebates. Many POs with these requirements sent ***additional mailings to inform beneficiaries of the CM/DM activities they had fulfilled thus far*** or how much they had earned in rebates, with the goal of nudging beneficiaries to complete additional requirements.

The two ***POs offering rebates as part of their intervention designs had additional communications with beneficiaries to increase saliency of VBID benefits.*** PO A sent rebate checks three times a year. VBID-eligible beneficiaries also received quarterly scorecard completion letters listing completed preventive care activities. PO C sent a summary letter with each rebate check informing beneficiaries about the services for which they received rebates. A representative from PO C stated that they thought the reminder letters helped keep beneficiaries "engaged throughout the entire year." These letters were simplified in 2018 to exclude detailed claims information, which some beneficiaries found confusing in 2017. A representative explained: "So first quarter, you might have earned $75 because you did the [personal health review] and then the $25 activity. However, you only have one claim listed and then your out of pocket is $20, so then they're confused why they've only gotten a $20 check and why then the next quarter they may get that additional money." PO C also used an automatically generated ***"robocall" to encourage beneficiaries to open their mail and cash their rebate checks.*** Beneficiaries who had not cashed their rebate checks after 30 days also received a follow-up call.

Communication with Providers

Compared to beneficiary outreach, VBID-related communication with providers was less intensive because most VBID designs did not require active provider engagement. Most POs reported informing providers about VBID via regular ***newsletters*** (POs A, B, C, D, E, F, H, and I). PO A reported sending ***reports to providers*** that listed all beneficiaries on their patient panels who had not yet received all preventive services on the scorecard. Finally, PO H sent additional mailings to providers whose patients participated in VBID, asking providers to encourage VBID beneficiaries to use the program benefits.

In addition to mailing information about VBID, six POs (B, E, G, I, J, K) conducted outreach to physician groups. POs G, E, and K ***regularly met with and provided education about VBID to provider groups*** with high numbers of VBID beneficiaries in their practices. PO J gave ***talks about VBID at regional medical director meetings,*** which were attended by clinicians who worked with provider groups contracted with the PO. PO B, which included high-value providers in its intervention design, conducted in-person outreach to providers to let them know whether they were considered high value and created a process allowing providers to dispute their high-value status. PO B also noted the need for ***regular communication with office managers*** to ensure that they collected the correct copays for VBID beneficiaries.

PO I, which implemented a telehealth intervention, sent monthly reports to providers about telemonitoring results of their VBID patients and asked providers whose patients joined the

VBID model to sign a diuretic protocol. Representatives from PO I also traveled around the state and shared information about the program with their physician groups to make sure that physicians were aware of their patients who participated in VBID.

Summary

Although CMS made changes to the VBID model test for both 2018 and 2019, few returning POs changed their intervention designs or conditions of focus in the second or third year of participation in the model. In general, returning POs stated that the application deadline for the 2018 and 2019 model test years occurred before there was enough data on which to inform changes to their designs.

Most POs required eligible beneficiaries to enroll in VBID by either calling the PO or responding to an outreach call. Most POs also continued to require beneficiaries to engage in CM/DM activities, though many demonstrated flexibility in those requirements and actively tried to reduce the burden on beneficiaries.

Some POs conducted outreach to beneficiaries and providers beyond what was required of them by CMS. POs typically reached beneficiaries using mail or telephone. Additional outreach to beneficiaries included phone calls to follow-up on mailings regarding CM/DM requirements or enrollment in VBID. Outreach to providers included newsletters, reports on which patients of the provider are participating in VBID, and discussions or educational presentations with physician groups.

3. Value-Based Insurance Design Implementation Experiences

Implementing VBID was a significant undertaking for many of the participating POs, requiring design of a new benefit, additional communications with beneficiaries to describe the model, modifications to information technology (IT) to identify eligible beneficiaries, new processes to track enrollees' participation status, and cross-departmental collaboration to ensure that beneficiaries were provided the correct set of benefits. Participating POs generally faced a steep learning curve in the first year of implementation. Although processes tended to become smoother over time, some challenges remained even for seasoned participants. This chapter explores POs' VBID implementation experiences and includes interview data from all 11 POs participating in the model test at some point between 2017 and 2019. We conclude this chapter with a brief discussion of POs' plans for continuous participation in the model test and a review of the reasons POs offered for leaving the model in 2020.

Key Takeaway Points:

- POs considered 2017–2018 VBID marketing restrictions to be implementation barriers that negatively affected beneficiary awareness of VBID benefits.
- As during the first year of the VBID model test, beneficiary identification and enrollment in VBID created implementation challenges in later years. These processes were slowed down by the time it takes to process claims and reliance on diagnosis codes that may be an imprecise way to identify beneficiaries with eligible conditions.
- The need to continuously engage beneficiaries in CM/DM was a challenge for POs with participation requirements.
- Seven of the 11 POs that participated in the first three years of the model test discontinued their participation by 2020.
- POs' reasons for leaving the model included the high administrative burden associated with VBID, the lack of a clear positive return on investment (ROI), and perceived inability to implement new model requirements before 2020 applications were due.

Communication and Marketing Restrictions

Initially, CMS prohibited POs from marketing VBID benefits to potential enrollees. During the 2017 open enrollment period (which occurred in late 2016), many POs interpreted this guidance as preventing them from communicating with eligible beneficiaries before the start of the model test. In 2017, CMMI updated its marketing guidance to allow POs to communicate with VBID-eligible beneficiaries in September, enabling POs to review VBID benefits with eligible beneficiaries prior to the 2018 open enrollment period (CMMI, 2019b). However, POs were still prohibited from marketing VBID to prospective enrollees and from discussing VBID

with noneligible beneficiaries. For the 2019 open enrollment period, CMMI further changed the model's marketing restrictions, allowing POs to market their VBID benefits to prospective enrollees and beneficiaries without VBID conditions, although they needed to be clear that VBID was only available to those with specific conditions (CMMI, 2019c).

POs considered the marketing restrictions that were in place in 2017 and 2018 to be implementation barriers that (1) limited the outreach activities of insurance brokers and agents and impeded beneficiaries' abilities to make an informed decision about plan changes during the annual enrollment period, (2) reduced general awareness of the VBID benefits among beneficiaries, and (3) increased beneficiary skepticism about VBID benefits when POs reached out to them. Concerns about marketing restrictions were particularly acute during our 2018 interviews, when seven POs described these restrictions as implementation barriers.

POs B, F, and J noted that ***marketing restrictions posed challenges to insurance brokers and agents who could not mention VBID benefits to potential new enrollees.*** POs interpreted the marketing restrictions as preventing them from telling beneficiaries with a VBID-targeted condition about the reduced cost-sharing or supplemental benefits they would receive if they were to switch from a non-VBID plan. According to a PO B representative, these restrictions "created a little bit of hesitancy to maybe jump in on some of the VBID pieces for fear that we could overstep. . . . I'm afraid that it might make us walk a little bit slower, tighter."

Moreover, limits imposed on the brokers also required some POs to conduct direct outreach to current VBID-eligible beneficiaries during open enrollment to make sure they fully understood their choices and could make informed decisions about switching plans.

> *"When AEP [annual enrollment period] occurred . . . a number of people who were in plans . . . with a VBID benefit saw a lower premium plan and switched to that without realizing that that lower premium plan did not have the [VBID] benefit. So then we conducted outreach to make sure that they understood that they were losing that benefit by switching to the other plan, so they could make an informed decision."*
>
> *PO H representative on the need to conduct additional outreach to VBID-eligible beneficiaries during open enrollment period*

Representatives from some POs argued that ***marketing restrictions may have negatively affected awareness of VBID benefits even among VBID-eligible beneficiaries,*** particularly for POs with participation requirements. These POs relied on beneficiaries' willingness to open letters and pick up the phone to enroll in the VBID program. A representative of PO H said, "I think [marketing restrictions] hamper the general awareness of VBID because it does then become incumbent on the person to open that letter or answer the phone call when we call." Moreover, many POs interpreted the marketing restrictions as preventing them from being able to include information about VBID benefits in the materials distributed right after the open enrollment period, which they said reduced beneficiary engagement.

> *"We know that members are signing up for this benefit during open enrollment. Open enrollment ends, but we still can't mail them until January 1. If we were able to do that . . . I could start talking to them about it, but we have to wait for the mailing to go out [in January] because we've found that they needed to have that information in their hands . . . even giving me a month would help."*
>
> *PO B representative on perceived communication restrictions with VBID-eligible beneficiaries*

Finally, POs expressed concern that ***marketing restrictions may increase beneficiaries' skepticism about their benefits and health plans.*** Medicare beneficiaries are often skeptical about getting something for free. As a representative from PO E said, "You can't go out and tell anybody about [VBID benefits]. Then all of a sudden, you're making cold calls to people offering this great benefit. And, of course, senior citizens are skeptical when they're getting a phone call from somebody because of scam artists, you know, they're always leery." A representative from PO A also noted that beneficiaries do not have an easy way of validating the legitimacy of VBID benefits because the information about this program is not available on any public-facing websites that POs maintain. Even with the relaxation of marketing restrictions in 2019, PO I representatives said they chose not to market the benefits broadly, worrying that doing so could have a negative effect on their risk scores: "Based on what we chose and what we have been doing, it appears that if we were to market that, we could bring on some anti-selection." Although we have not heard from any POs that they market their VBID benefits widely, PO B representatives, for example, stated that they are looking for ways to market their VBID offerings more broadly.

Beneficiary Identification and Enrollment in Value-Based Insurance Design

Beneficiary identification and enrollment in VBID also created some implementation challenges, which often had to be resolved with manual overrides in IT systems. POs E and J noted that ***there was a lag between the time beneficiaries get a claim that makes them VBID eligible and the time they receive VBID benefits.*** Sometimes, POs had to manually process refunds for copays paid after an eligible diagnosis was confirmed, but before the internal systems were updated to reflect a new VBID status.

> *"The doctor might not submit a claim for 30 days. The claim gets processed, and the member is flagged. Once that member is flagged as VBID and goes through the system, at the pharmacy team, we go through and we get that list. We then do a back look for those members to see if they had any drugs that were filled that were eligible, and then we'll process the refunds to those members [manually]."*
>
> *PO J representative on the need to manually issue copay refunds*

> *"CMS generally does everything by the first of the month. When [we are] speaking to members and [they confirm their willingness to participate in CM], we want to offer [the VBID benefits] at that moment that we're talking to them. So if they make an appointment the next day, they're going to get the reduced cost-sharing. And that isn't in sync with the way CMS does things, for somebody to be into a program in the middle of a month. So that was a challenge for us to try to manage it that way."*
>
> *PO E representative on the delays in offering VBID benefits to newly-eligible beneficiaries*

Moreover, ***relying on diagnosis codes to identify VBID-eligible beneficiaries created a situation in which POs had to offer VBID benefits to those who did not actually have the targeted condition.*** It is possible that not all beneficiaries identified using claims data were eligible for VBID, because some people whose records had a relevant diagnosis code may not have actually had the condition. A PO J representative explained: "From a claims perspective, a doctor might just want to have a test. So a beneficiary would end up with a coronary artery disease diagnosis. . . . For that particular population, because they have the CAD diagnosis, we don't remove them out of the program." In such a situation, POs continued offering benefits to those who did not really need them.

At the same time, however, some POs allowed patients to self-attest to VBID eligibility or let providers refer their patients to VBID to ensure that their patients could get VBID benefits sooner.

> *"Sometimes members [referred by their providers] would call in and say: 'I would like to be enrolled in the VBID program. I have heart failure or COPD.' And so when they did that, we just enrolled them. And what we found out is that many of those, if not most, after the care manager made an outreach, were really not eligible because they probably didn't have the diagnosis. That became pretty problematic."*
>
> *PO F representative on the problems with beneficiary identification via provider referrals*

Continuous Engagement of Value-Based Insurance Design-Participating Beneficiaries

Four out of seven POs with conditional participation requirements reported challenges related to keeping VBID-participating beneficiaries engaged, which highlights the importance of aligning and continuously adjusting VBID intervention designs to the needs of enrolled beneficiaries. Continuous engagement was challenging not only for POs that required quarterly contact with a care manager (PO C) but also for POs that tailored the frequency and intensity of required CM/DM sessions to the needs of each beneficiary (POs E, F, and G). According to PO F representatives, ***development of an ongoing relationship with a care manager turned out to be emotionally difficult and burdensome for high-risk beneficiaries.*** One said, "So probably the really sick people, they have doctors' appointments, they have visiting nurses, they have lots of

different people calling them, and it's another voice on the other side of the phone and it's . . . I think it is challenging for them to add one more thing."

One reason why engagement in CM/DM activities might have been challenging is that ***some POs made assumptions about the required number of CM/DM activities based on their prior experience with short-term CM programs.*** A representative from PO C remarked: "What we are expecting is that somebody is going to keep engaging with me, time after time, after time, after time throughout the year. Not only one year, but multiple years."

Others, however, noted that their PO's ***previous CM/DM programs focused on chronically ill patients who understood the benefits of ongoing support. Engaging with beneficiaries early in the disease progression stages was a new challenge for POs*** whose VBID-eligible beneficiaries would not have been included in other CM/DM programs.

> *"So it's those individuals that we were looking at how can we outreach differently to help them understand the value of being in the VBID program for the longer term. Even though they may be newly diagnosed, we want them to understand the value of the program, work with case management, see their primary care physicians, take the medications."*
>
> *PO G representative on the challenges of engaging newly diagnosed beneficiaries*

Although challenging, ***engaging this new group of beneficiaries who otherwise would not have participated in CM/DM was a positive unintended consequence of the VBID model test.*** A representative from PO F said: "We weren't getting at some of the earlier diagnosed people, so we created a new wellness category in CM/DM for CHF and COPD for [our VBID program]. This gave us an opportunity to maintain contact with that group. It's kind of giving us a key into some early rising risk members."

The use of rebates delays financial rewards, which may negatively affect the saliency of VBID benefits and hamper continuous engagement of beneficiaries. According to representatives, PO C had some beneficiaries who completed one or two CM/DM activities and earned $25 or $50 but could not receive their rebates until they visited the doctor, so keeping "them engaged from January to October can be challenging." This was the primary reason why this PO switched to using an OTC debit card in 2019. In addition, PO A stated that paying copayments on services that make beneficiaries eligible for MA rebates may have confused beneficiaries and negatively affected their VBID participation. For example, beneficiaries had to first pay a copay to complete an eye examination or lipid test before they could be eligible to receive rebates.

Decisions About Future Model Participation

Our interviews revealed diminishing interest among several POs in the VBID model test: PO E left the model test in 2019; POs A, D, F, H, I, and K left the model test in 2020. Although

most POs that left the model test have participated for at least two years, PO K left the model test after only one year. Only four 2019 model participants participated in the MA VBID model test in 2020. PO representatives gave three main reasons for leaving the model test.

The administrative burden of participating in the model test was too high, according to four POs (A, E, F, and I). Representatives from these POs raised concerns about challenges related to keeping track of which beneficiaries were in VBID and which were not, reporting this information to CMS, having their VBID-related communications materials be reviewed first by the model team and then by the regional CMS office, and processing claims manually for beneficiaries in VBID. Three of these POs—E, F, and I—reported planning to offer their VBID benefits outside the model test to reduce their administrative costs, thereby benefiting from the uniformity rule changes. A representative from PO I, which offered a telehealth intervention as part of its VBID design, said, "We can better manage the care, achieve the outcomes that we're looking to do with the VBID program without the administrative burden." It is worth noting that several of the remaining POs are also offering, or are in the process of designing, VBID-style benefits under the uniformity flexibility while staying in the model test. None of them, however, shared the details of their VBID-like benefits offered outside the model test.

Four POs (A, F, I, and K) reported not seeing desired outcomes as one of the reasons for leaving the model test. As detailed in Chapter 6, PO F, which wanted to see if copays were indeed a barrier to receiving care, determined that lowering copays did not increase the number of PCP or specialist visits. PO I did not see positive ROI: "It was more the cost of the program [including the cost of telemonitoring and equipment] to the plan, to be able to deliver it in a cost-effective manner."

Finally, two POs (A and H) considered the addition of certain requirements to the MA VBID model test for 2020 to be burdensome. They stated that designing an implementation strategy for the requirement to include an advanced care planning strategy as part of their WHP by the time 2020 applications were due in the spring of 2019 was not feasible. PO A stated that the implementation of this requirement would require provider engagement, whereas PO H was not confident it would be "able to implement it." PO H representatives also stated that "if the program continues, we may look to reenter in the future." In contrast, the four POs that are continuing participation in the model test either already offer advance care planning discussions in their regular CM workflow or found a vendor that could scale up these discussions with beneficiaries.

Summary

By 2019, most VBID participants reported no longer experiencing any major implementation challenges. This is not surprising, because most POs did not make major changes to their intervention designs between 2017 and 2019. Some POs felt that marketing restrictions that were in effect during the 2017 and 2018 open enrollment periods may have negatively affected model

implementation by making it more difficult to engage beneficiaries. POs also expressed a concern that internal data processing issues made it difficult to identify VBID-eligible beneficiaries in a timely way, in some cases requiring cumbersome work-arounds (e.g., manual data entry). Finally, some POs with CM/DM participation requirements reported challenges with keeping beneficiaries engaged in the model test. These findings suggest that successful VBID implementation requires integration of VBID-related processes with standard business operation procedures and calls for periodic fine-tuning of beneficiary communication, benefit tracking, and CM/DM activities. Finally, the majority of VBID-participating POs left the model test in 2019 or 2020. The main reasons for discontinuing their model test participation were the administrative burden, a lack of expected outcomes, and the burden of implementing new model test participation requirement for 2020.

4. Beneficiary Participation

Beneficiaries' willingness to participate in VBID is critical to the success of the model. In this chapter, we describe beneficiary participation in VBID interventions using data that POs provided to CMS through the MARx reporting system. We supplement these findings with responses from our interviews with POs and beneficiaries.

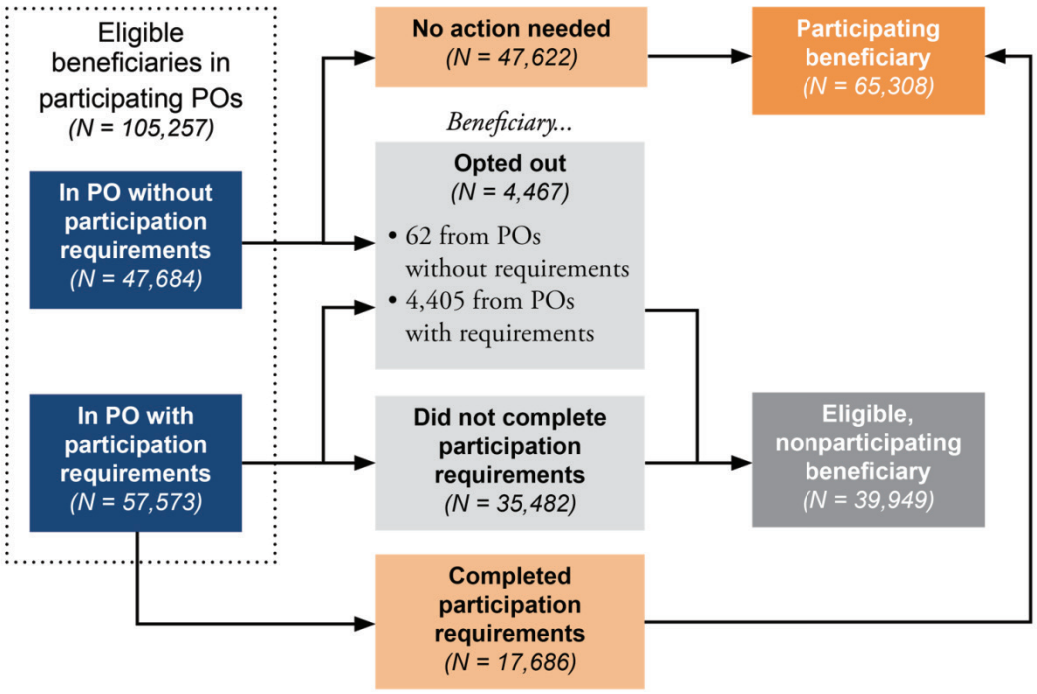
Key Takeaway Points:

- Roughly 62 percent of 105,257 eligible beneficiaries in 2018 and of 111,012 eligible beneficiaries in 2019 participated in the model test. Participation rates were lower, between 24 and 31 percent (depending on the year), among eligible beneficiaries in POs with participation requirements.
- Beneficiaries who participated in the model test were generally similar to those who did not participate.
- In 2018, six of ten POs felt that the intervention uptake among eligible beneficiaries was lower than expected. Five of these six POs had participation requirements or asked beneficiaries to proactively sign up for VBID. Three POs felt that intervention uptake fully met their expectations. Two of these three POs did not have participation requirements. Only one PO reported higher-than-expected participation rates. This PO had a participation requirement.
- The number of POs that were satisfied with beneficiary uptake increased between 2018 and 2019, although beneficiary participation rates remained the same across all POs and declined among POs with participation requirements.

Intervention Uptake Using MARx Data

In 2018, there were 105,257 VBID-eligible beneficiaries enrolled across the ten POs that offered VBID benefits. Seven of these POs had participation requirements, and some VBID-eligible beneficiaries did not participate in the model test. Figure 4.1 illustrates VBID participation status among the 105,257 eligible beneficiaries in 2018. The data indicate that ***65,299 beneficiaries (62 percent of those eligible) participated in the model, while 39,958 (38 percent) did not participate.*** Nearly 100 percent of beneficiaries in POs without participation requirements participated in the model test. Of 57,573 eligible beneficiaries enrolled in POs with participation requirements, 17,686 (31 percent) completed those requirements.

Figure 4.1. 2018 Participation Status Among VBID-Eligible Beneficiaries (*N* = 105,257)



NOTE: Data are from the MARx reporting system. The population includes all beneficiaries reported to be eligible for VBID at some point during 2018. All beneficiaries who were alive at some point in 2018 are included; eligible beneficiaries who died during 2018 are assigned participation status based on their most recent status before death.

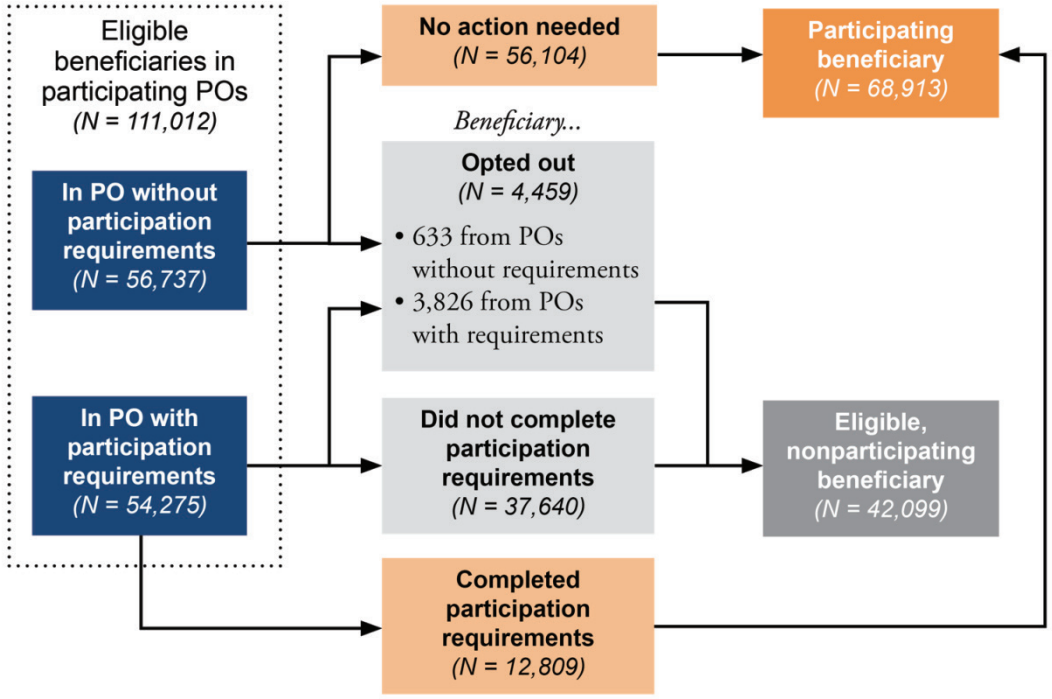
Slightly more people, 111,012, were VBID eligible in 2019 (Figure 4.2). ***Among those eligible in 2019, 68,913 (62 percent of those eligible) participated in the model test,*** the same rate as in 2018.[1] Among beneficiaries in POs without participation requirements, 99 percent joined the model test, compared to 23.6 percent of beneficiaries in POs with participation requirements. Across the ten POs that participated in 2019, six had participation requirements and four did not have participation requirements.

In aggregate, about 12,000 additional people became eligible for VBID between 2017 and 2019 (see Table 4.1). Over the same period of time, MA enrollment across the United States grew from 21.2 to 24.5 million, and MA enrollment in states in which the VBID model test was available grew from 3.4 to 16.3 million (the large increase reflects that VBID was offered in more states each year).[2]

[1] For all POs except PO H, we assess beneficiary eligibility base on eligibility at any point in the CY and record participation status based on their most recent status within the year. For PO H, we assess eligibility and participation status only through September 2019, due to reporting anomalies in subsequent MARx data updates.

[2] These numbers were calculated using IDR enrollment data and reflect the number of unique beneficiaries observed as being enrolled in MA in each year.

Figure 4.2. 2019 Participation Status Among VBID-Eligible Beneficiaries (*N* = 111,012)



NOTE: Data are from the MARx reporting system. For all POs but PO H, the population includes all beneficiaries reported to be eligible for VBID at some point during 2019. All beneficiaries who were alive at some point in 2019 are included; eligible beneficiaries who died during 2019 are assigned participation status based on their most recent status before death. For PO H, the population includes all beneficiaries reported to be eligible for VBID through September 2019; further data updates for PO H were excluded due to reporting anomalies.

Among VBID-eligible beneficiaries, participation rates increased slightly—by about 2-percentage points—between 2017 and 2019 (Table 4.1). However, because the largest new model entrant (PO K) did not have participation requirements, the steady overall participation rate masks a decline in participation rates in POs with requirements. When we limit the analysis to POs with requirements, participation rates fell from 31 to 24 percent between 2018 and 2019. This decline in part reflects the departure of PO E, which reported relatively high participation in 2018. In addition, PO C—which had a relatively large eligible population—had a decline in participation between 2018 and 2019. Notably, although most POs commented that beneficiary participation was below expectations in 2018, PO C was the only PO that continued to report lower-than-expected beneficiary participation in 2019 (we discuss this issue in more depth later in this chapter, under "Perceptions of Intervention Uptake").

The relatively low rate of participation among POs with participation requirements raises the concern that participating beneficiaries may be systematically different from eligible beneficiaries who did not participate. Table 4.2 compares characteristics of participants and eligible nonparticipants for the seven POs that had participation requirements in 2018 and 2019.

Table 4.1. VBID Participation Rates Among Eligible Beneficiaries in 2017–2019

	2017	2018	2019
Number of eligible beneficiaries	98,740[a]	105,257	111,012
Number of eligible beneficiaries in PBPs with participation requirements	53,816	57,573	54,275
Share of all eligible beneficiaries who participated	59.8 percent	62.0 percent***	62.1 percent***
Share of eligible beneficiaries in PBPs with participation requirements who participated	26.4 percent	30.7 percent***	23.6 percent***

NOTE: Tests of statistical significance indicate whether the estimate was statistically different from the corresponding estimate in 2017.
* indicates 10 percent significance ($p > 0.05$, $p \leq 0.10$),
** indicates 5 percent significance ($p > 0.01$, $p \leq 0.05$), and
*** indicates 1 percent significance ($p \leq 0.01$). The sample in each year reflects all eligible beneficiaries in POs that participated in the model test in that year.
[a] The number of eligible and participating beneficiaries for 2017 changed slightly from the 95,063 reported in our first evaluation report because POs updated their data.

Table 4.2. Characteristics of Participating and Nonparticipating Beneficiaries in POs with Participation Requirements, 2018 and 2019

	Completed Requirements 2018	Did Not Complete Requirements 2018	Opted Out 2018	Completed Requirements 2019	Did Not Complete Requirements 2019	Opted Out 2019
Age	76.4	77.1***	77.1***	77.7	77.3***	77.9
Risk score (HCC)	1.8	1.8	1.7***	1.6	1.7	1.6
Dual eligible (percent)	12.6	12.8	9.8***	11.0	12.8***	10.1
Male (percent)	47	47.6	45.3**	45.1	47.9***	45.0
White (percent)	93.8	92.7***	94.8***	93.3	92.8**	94.9***
Black (percent)	2.7	3.3***	2.1***	3.0	3.2	2.0***
Asian/Pacific Islander (percent)	0.6	0.9***	0.3**	0.7	0.9**	0.3***
American Indian/Alaska Native (percent)	0.2	0.2	0.2	0.2	0.2	0.2
Hispanic (percent)	1.2	1.4***	1.1	1.3	1.3	1.1***
Multiple races (percent)	1.6	1.6**	1.5***	1.6	1.6***	1.5**

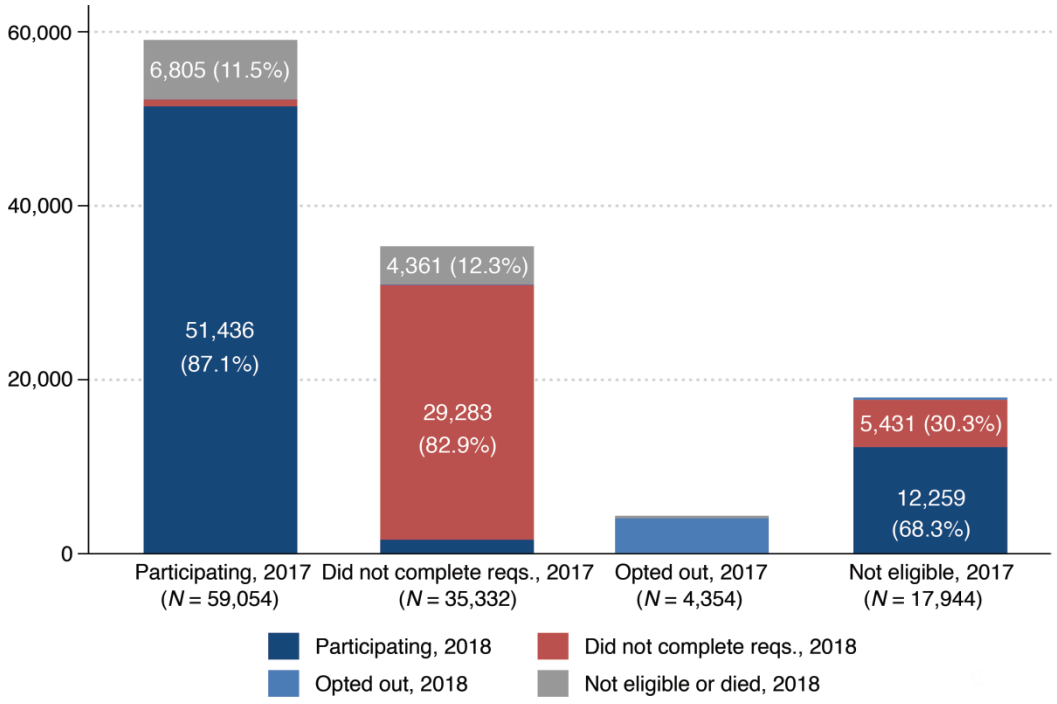
NOTE: Tests of statistical significance indicate whether beneficiaries who did not complete requirements or who opted out of the VBID model test in a given year were statistically different from those beneficiaries who completed requirements in the same year. ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.

Although some of the demographic differences are statistically significant, they are small in absolute magnitude. For example, in 2019 participants were a statistically significant 0.4 years older than eligible individuals who did not complete requirements, and those who did not complete requirements were 0.5 percentage points less likely to be white. Because the

differences are quite small, our conclusion is that ***those who participated in the model test were demographically similar to those who did not participate.*** This result mirrors the findings from our first evaluation report, in which we found few meaningful differences between participating and nonparticipating beneficiaries in 2017.

Figure 4.3 shows how VBID participation status changed between 2017 and 2018, with 2017 participation status shown in the x-axis and 2018 participation status represented on the bars. The figure indicates that participation status was relatively stable over time. For example, among the 59,054 beneficiaries who participated in VBID in 2017, 51,436 also participated in 2018—implying a retention rate of 87 percent. The finding that participation was relatively stable corresponds to what we heard in our interviews with POs. For example, representatives from PO B remarked that one of their achievements as part of the model test was their ability to retain participants over time.

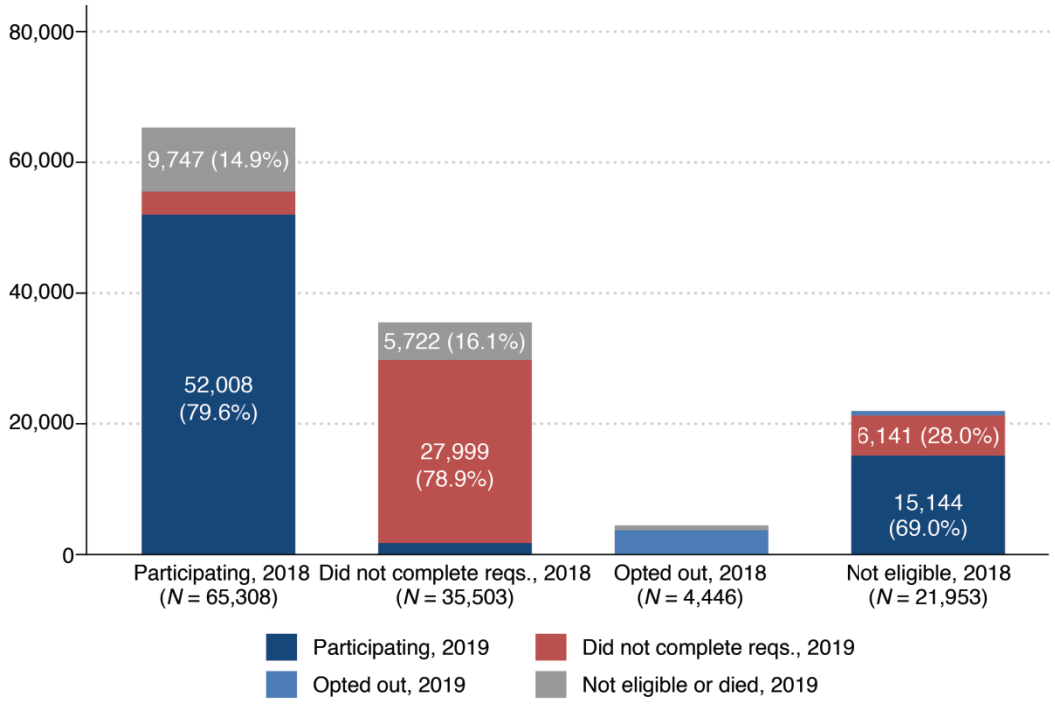
Figure 4.3. Change in VBID Participation Status Among Eligible Beneficiaries, 2017–2018



NOTE: The bars illustrate the relationship between 2017 participation status and 2018 participation status. For example, among 59,054 eligible beneficiaries who participated in VBID in 2017, 51,436 were still participating in 2018, 6,805 had become ineligible for VBID or died, and the remainder either opted out or did not complete participation requirements in 2018.

Figure 4.4 shows changes in participation status between 2018 and 2019. Although most 2018 participants continued to participate in 2019, retention rates declined between 2018 and 2019. Out of 65,308 people who participated in 2018, only 52,008 (80 percent) participated in 2019, lower than the 87-percent retention rate found for 2018.

Figure 4.4. Change in VBID Participation Status Among Eligible Beneficiaries, 2018–2019



NOTE: The bars illustrate the relationship between 2018 participation status and 2019 participation status. For example, among 65,308 eligible beneficiaries who participated in VBID in 2018, 52,008 were still participating in 2019, 9,747 had become ineligible for VBID or died, and the remainder either opted out or did not complete participation requirements in 2019.

A total of 17,944 individuals became newly eligible for VBID in 2018, and 21,953 individuals became newly eligible for VBID in 2019. People can become newly eligible for VBID for a variety of reasons, including becoming newly eligible for Medicare, switching from a nonparticipating to a participating plan, and being newly diagnosed with a VBID-eligible condition (Table 4.3). ***Most newly eligible 2018 beneficiaries (11,389, or 63 percent) were previously enrolled in a VBID-participating PO and newly met eligibility criteria*** in 2018 (e.g., due to a new diagnosis). In contrast, ***nearly half of newly eligible 2019 beneficiaries (10,299, or 47 percent) were enrolled in PO K, which joined the model test in 2019.***

Table 4.3. Reasons Beneficiaries Became Eligible for VBID in 2018 and 2019

Reason for Eligibility	Newly Eligible in 2018	Newly Eligible in 2019
Newly eligible for Medicare	1,547	958
Current plan is a new VBID entrant	270	10,299
Switched from nonparticipating PBP or FFS to participating plan	4,738	3,830
Previously enrolled in VBID-participating plan, newly eligible for VBID	11,389	6,866
Total	17,944	21,953

Perceptions of Intervention Uptake

During our interviews with POs, we asked whether the intervention uptake was lower than expected, as expected, or higher than expected.

In 2018, six of ten POs felt that the intervention uptake among eligible beneficiaries was lower than expected (POs C, E, F, G, I, and J). Five of these six POs required beneficiaries to participate in CM/DM or to proactively sign up for VBID before receiving lower cost-sharing for high-value services. The POs offered explanations for the relatively low levels of beneficiary participation, many of which were confirmed in our interviews with beneficiaries. Below, we summarize these explanations, which can be grouped into six broad themes.

1. **Health status of beneficiaries:** Many POs commented that VBID-eligible beneficiaries, especially those with comorbid conditions, were much sicker than anticipated. Interviews with beneficiaries support this perception: At least one-third of the VBID-eligible beneficiaries we interviewed reported dealing with comorbidities, many of which were outside the VBID intervention focus. Several beneficiaries explained that dealing with VBID-eligible conditions was not as much of a priority to them as addressing their other comorbid conditions. Moreover, according to POs C and F, many of the sicker beneficiaries were either unable or unwilling to proactively sign up for VBID and complete CM/DM participation requirements, which suggests that reducing barriers to receiving VBID benefits as part of the model test may help engage high-risk beneficiaries.
2. **Engagement of healthier beneficiaries:** Although some POs noted challenges in engaging sick beneficiaries, others (notably E, F, and G) found it more difficult to engage healthier or newly diagnosed beneficiaries. For example, a representative of PO E noted challenges in attracting this "less-sick population" because "for those [beneficiaries] who hadn't talked to us before . . . not only were we selling [them] on case management, now we were selling [them] on these free services that we had never previously offered." A beneficiary from PO I who received a VBID notification letter but denied having a heart condition echoed this challenge, recalling: "When I read [the letter], I said I don't need this." This finding suggests that educating lower-risk beneficiaries about the value of VBID benefits may be an important engagement technique.
3. **Difficulties in contacting members:** Several POs (E, F, and G) noted difficulties in reaching beneficiaries because of inaccurate contact information, unwillingness to answer the phone, or unwillingness to open mail from insurance companies. As a representative from PO E put it, "Not only phone numbers, but addresses even to send people letters [are a challenge], because [beneficiaries' contact information] changes, and we don't necessarily know what that is." A representative from PO H added that beneficiaries receive a lot of communications in general: "Once [beneficiaries] found out about it, [they would say] 'I think this is great, we love it.' So I don't think that it's the structure of the communications that's the difficulty, it's breaking through all the other clutter." Our interviews with beneficiaries support this finding: Only 65 percent of beneficiaries reported receiving and reviewing VBID-related information. Finding ways to encourage beneficiaries to regularly update their contact information may facilitate beneficiary engagement.

4. **Burden of participation:** POs F and G perceived that some of their beneficiaries felt that the requirement to participate in CM/DM activities was too burdensome. A representative from PO F said that the burden is often emotional, in that beneficiaries "would have to actually think about their disease and respond to questions. And maybe they don't want to think about it." Although some beneficiaries who did not join VBID may have perceived participation requirements to be too burdensome, our interviews with beneficiaries who participated in CM/DM indicated that they were comfortable with participation requirements.

> *"Oh, I'm okay with it. They asked me is it okay? Can we call every three months? And, sure, that's fine. And they ask me questions. They're very nice. In fact, they offer . . . a ride to the doctors or to get my prescriptions. . . . So, no, I have no problem."*
>
> *A beneficiary from PO B on the low participation burden in CM/DM activities*

Of course, those who participated may be less likely to view the intervention as burdensome than those who did not join. To address lack of participation, POs may need to "sell" the value of CM/DM activities to beneficiaries so that they can see the benefit of ongoing engagement.

5. **Beneficiary skepticism and unwillingness to receive lower copays:** Several POs stated that the Medicare population is often suspicious about receiving anything for free. A representative from PO E remarked, "The older population would say: 'Well, there's nothing free. That can't be real.'. . . It was really not as easy as we thought it would be to get them engaged." In our interviews with beneficiaries, some expressed concern about increased medical benefits detracting from other needed benefits, such as housing subsidies. To determine federal rent subsidies, the amount beneficiaries paid in medical copays is deducted from their gross incomes. If the copays are substantially lower than in previous years, the gross income becomes higher and the rent subsidies could potentially decrease. This was of concern to several beneficiaries, like the one from PO H: "I'm on this . . . special help for paying [for housing]. I can't do anything to disturb that. I don't want to cause any problems, I don't want them to re-look at something, and all of a sudden give me a letter that I'm not eligible anymore. So, I guess I live in a little bit of fear of that." To address beneficiary skepticism, POs may need to educate beneficiaries on the value of VBID participation so that beneficiaries could make a more informed decision about VBID benefits.
6. **Challenges in program implementation:** One representative from PO I stated that leadership turnover slowed down beneficiary outreach activities during the first year of the model test. The representative said: "Because of the turnover with leadership and also the people involved within the program . . . and the doctors, I think we got a delayed start to really get this thing off." Some beneficiaries from PO I, as well as providers, also reported experiencing implementation difficulties, especially with installation of the remote monitoring equipment. One beneficiary reported that the PO could not install scales on carpeted floor, and a provider stated that some of his patients live in very small apartments with little room for the required equipment. This finding further reinforces the importance of designing VBID interventions with implementation in mind and the role of leadership support as key implementation facilitators highlighted in our Year 1 evaluation report.

Despite the challenges described above, not all POs agreed that participation in their VBID intervention in 2018 was lower than expected. ***Three POs, A, D, and H, felt that intervention uptake among eligible beneficiaries fully met their expectations.*** Two of these three POs, D and H, did not have participation requirements, which may have made participation easier for some beneficiaries. PO A, however, required beneficiaries to complete four preventive screening activities to receive VBID benefits.

Only PO B, which had a participation requirement, reported higher-than-expected intervention uptake among eligible beneficiaries in 2018. A representative of PO B said: "Historically, we've seen engagement with care management in the 20 to 30 percent range, so having a program that requires a quarterly contact to still be above 50 percent[3] at the end of Year 1 was something that we were tremendously proud of." PO B offered one of the more comprehensive packages of VBID benefits during the first two years of the model test; its package included high-value providers, DME, and several supplemental benefits, which may have been a factor in increasing uptake. PO B also called all eligible beneficiaries to make them aware of VBID and its benefits. It also allowed beneficiaries whose conditions were well managed to participate in fewer CM activities (see Chapter 2).

In 2019, however, only PO C continued reporting lower-than-expected intervention uptake. Although PO B still reported higher-than-expected uptake, it was joined by PO G, which reported exceeding its recruitment target of 1,000 beneficiaries by 100. PO G attributed this increase to the expansion of VBID benefits to additional PBPs and the introduction of meal and transportation benefits. Other POs we interviewed in 2019 reported that the uptake met their expectations or that they had no expectations because they made the benefits available to all eligible beneficiaries. It is worth noting, however, that our quantitative analyses show that overall beneficiary uptake of the VBID intervention did not change between 2018 and 2019, and that uptake in POs with participation requirements declined. These results suggest that participating POs might have modified their expectations. Indeed, three POs in 2019 stated that they had no expectations about the uptake in 2019 (in contrast, no POs reported that they had no expectations about uptake in 2018).

Summary

Between 2017 and 2019, around 60 percent of all eligible beneficiaries participated in the model test. Among eligible beneficiaries in POs with participation requirements, however, participation rates were lower, between 26 and 31 percent. Characteristics of participating beneficiaries were generally similar to those of beneficiaries who did not participate. The number of eligible beneficiaries increased by about 12,500 individuals over the first three

[3] Data from MARx indicate that 48.6 percent of eligible beneficiaries in PO B participated in 2018.

years of the model test. Almost 90 percent of beneficiaries who participated in 2017 continued their participation in 2018 and 73 percent of beneficiaries who participated in 2018 continued their participation in 2019.

Interviews with representatives from POs conducted in 2018 generally supported the finding about relatively low participation rates, with six POs stating that participation was lower than expected. Five of these six POs had CM/DM participation requirements. Lower-than-expected participation may have resulted in part from difficulties in reaching out to beneficiaries and engaging them in CM/DM. We also heard from both POs and beneficiaries that many were struggling with comorbid illnesses that were not within the VBID intervention focus, that the CM/DM activities were burdensome, and that beneficiaries were skeptical of VBID in general.

In contrast, four POs felt that beneficiaries' enrollment met or exceeded their expectations in 2018. Two of these four POs did not have participation requirements, making it easier for beneficiaries to receive benefits. Beneficiary enrollment in VBID exceeded projections at the one PO that combined several VBID approaches into its intervention design, proactively reached out to eligible beneficiaries to encourage participation, and modified required CM/DM activities based on the needs of the beneficiary.

The number of POs that were satisfied with beneficiary uptake increased between 2018 and 2019, although participation rates remained the same across all POs and declined among POs with participation requirements. These results could suggest that POs modified their expectations about beneficiary uptake over time.

5. Beneficiary Experiences

In our interviews with beneficiaries (see Appendix B for methodological details), we asked them to share their experience with the benefits they received, and we compared our findings with results from the MA & PDP CAHPS survey. Using both types of data allowed us to contextualize instances of low awareness of VBID, such as when beneficiaries did not understand that VBID benefits they may receive were different from the benefits offered to everyone in their PBP. In addition, the interview data provided insight into the value of VBID benefits to beneficiaries, both from a financial perspective and in terms of the supplemental benefits they received through the intervention. In addition to describing beneficiaries' experiences with VBID, in this chapter, we also discuss beneficiaries' insights regarding how to improve VBID interventions.

Key Takeaway Points:

- Beneficiary awareness of VBID was very low in the MA & PDP CAHPS survey in 2017 and 2018, with less than 12 percent of VBID-eligible beneficiaries reporting being offered lower copays or extra benefits because of a health condition.
- Beneficiary interviews showed higher awareness, with 63 of 100 VBID-eligible beneficiaries reporting being aware of VBID or its benefits. Most of these 63 beneficiaries, however, recognized VBID only after hearing a description of VBID benefits.
- There was widespread confusion among VBID-participating beneficiaries about how VBID worked, what its eligibility criteria were, and what benefits were included in the model test.
- There was no consensus among beneficiaries on the value of VBID benefits, including reduced copays, additional supplemental benefits, and CM/DM activities.
- Of 11 MA & PDP CAHPS measures of experiences of care, only perception of care coordination was affected by VBID, with small increases in this outcome.

Awareness and Self-Reported Utilization of Value-Based Insurance Design Benefits

To notify eligible beneficiaries about VBID benefits, all POs were required to mail information about the program to beneficiaries. Although all POs reported mailing information about VBID, ***only 65 percent of beneficiaries we interviewed reported that they remember receiving some information about VBID.*** Receiving VBID-related information, however, did not always mean that beneficiaries read the information. Indeed, several beneficiaries reported relying on their family members to read their correspondence because they have vision problems or do not understand how insurance works. Others stated that they did not understand the

information presented in materials they received in the mail or that they did not pay attention to it because they receive too many letters: "All of these letters come and . . . you think, well now should I do this or what are they talking about here? I've got an entire drawer filled with papers that really don't account for much" (PO B beneficiary).

Although some beneficiaries did not remember receiving anything by mail, 35 of 100 recalled having a telephone conversation about VBID. In some cases, the PO called the beneficiary; in others, the beneficiaries called the PO. A beneficiary from PO H stated: "[I called a toll-free number] because I wasn't sure what the heck it was. They answered all my questions. They were good with it." As with the mailings, some beneficiaries also reported receiving too many calls, which may have led them to ignore the calls or be suspicious of them. A PO G beneficiary said: "We get so many calls, you don't know who you're talking to, what you're talking about."

Given that not all beneficiaries remembered receiving information about VBID, it was not surprising that figures on the awareness of VBID benefits and the subsequent receipt of these benefits were low. In addition to interviewing beneficiaries, we looked at the results of the MA & PDP CAHPS survey. CMS added two questions to the MA & PDP CAHPS survey to assess awareness of the VBID model benefits starting with the 2017 survey administration. The two questions were worded as follows:

- **Question 1:** A copay is the amount of money you pay at the time of a visit to a doctor's office or clinic. In the last six months, did your health plan offer to lower the amount of your copay because you have a health condition (like high blood pressure)?
  - **Answer options:** Yes; no; I am not sure; I do not have a copay; I do not have a health condition; I was offered a lower copay for another reason.
- **Question 2:** Your health plan benefits are the types of health care and services you can get under the plan. In the last six months, did your health plan offer you extra benefits because you have a health condition (like high blood pressure)?
  - **Answer options:** Yes; no; I am not sure; I do not have a health condition; I was offered extra benefits for another reason.

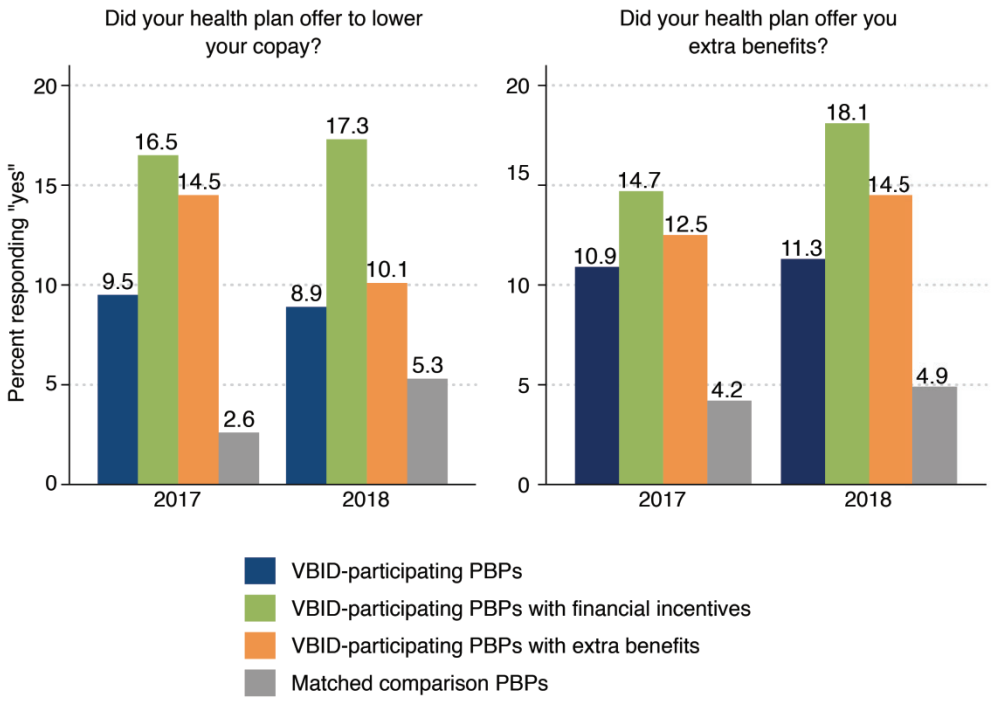
Figure 5.1 compares responses of VBID-eligible beneficiaries in VBID-participating PBPs with beneficiaries in matched comparison PBPs that would have met the VBID eligibility criteria if it were applied to them. Compared to beneficiaries in matched comparison PBPs, beneficiaries in VBID-participating PBPs had higher awareness of lower cost-sharing and extra benefits that may have been offered to them.

Even though the share of VBID-eligible beneficiaries in VBID-participating PBPs who reported VBID-related benefit changes was higher than in matched comparison PBPs (as we would expect), the overall share was quite low—8.9 percent of eligible beneficiaries reported being offered reduced cost-sharing in 2018 and 11.3 percent reported being offered extra health

benefits. The wording of the questions—which were written before POs submitted their VBID applications—do not capture the design of the VBID model accurately for some POs. For example, most VBID interventions incentivized participation in CM/DM, which is not reflected in the wording of the questions.

Figure 5.1 also shows responses among VBID-eligible beneficiaries separately for PBPs offering financial incentives, such as a lower copayment for a specialist visit, and PBPs offering extra benefits, such as weight scales or transportation. Awareness was higher in these groups when compared to all VBID-participating PBPs, but still remained under 20 percent for both measures in both years of the survey administration. None of the beneficiaries in the matched comparison PBPs should have been offered a reduced copay or extra benefits due to a health condition, yet approximately five percent of these beneficiaries responded "yes" to each of these measures in 2018. This indicates some confusion over their benefits and coverage or confusion due to the wording of the questions.

Figure 5.1. Awareness of VBID Benefits Among VBID-Eligible Beneficiaries from MA & PDP CAHPS



NOTE: Data are from the 2017 and 2018 MA & PDP CAHPS survey. Eligibility is based on RAND's implementation of the beneficiary eligibility algorithms (see the first annual report for details). The total number of respondents in VBID PBPs was 1,512 in 2017 (and 1,025 in 2018, denoted subsequently in parentheses); 395 (313) of these were in PBPs with financial incentives, and 354 (296) were in PBPs offering extra benefits. There were 3,255 matched comparators in 2017 (2,263 in 2018).

Results of our interviews with beneficiaries, which included both eligible participants and eligible nonparticipants, showed higher VBID awareness than the MA & PDP CAHPS survey findings. The majority of beneficiaries we interviewed (63 percent) reported being aware of VBID. Of those familiar with VBID, only 15 (24 percent) recognized VBID and described the program before we asked questions about it during the interview. ***Most beneficiaries reporting being aware of VBID (43 percent) recognized it only after hearing a description of VBID benefits offered by their PO.*** Roughly one-third ($N = 20$) confirmed their awareness of VBID after hearing the name of their PO's VBID program. One person (2 percent) recognized the program after the interviewer spelled out the VBID acronym.

Not being aware of VBID was the most frequently mentioned reason for not participating in the program. Indeed, ***after participating in our interview, several beneficiaries who were eligible for VBID but were not participating told us about their intention to sign up for VBID***. An additional reason for not participating in the model test was that ***not all beneficiaries felt that they needed VBID benefits.*** For instance, a beneficiary from PO E stated that he or she did not feel that a reduced copay for the pulmonologist would be beneficial for him or her at this time: "It isn't that I'm refusing [the offer], it's just that I don't feel that I'm there yet. . . . It's a wonderful thing, but at this moment, I don't feel I do." Other beneficiaries ***did not accept the potential severity of their chronic conditions;*** some mentioned feeling as though regular examinations are "a waste of money," or people getting a "freebie" (PO E beneficiary), seeing it as an imposition on a provider or care manager to "be telling me what to do and what not to do" (PO J beneficiary).

Many of those reporting being enrolled in VBID did not seem to understand how the program worked, including its beneficiary eligibility and selection criteria. Beneficiaries often expressed the thought that everyone in a given plan was getting VBID benefits. Some thought that VBID was "keyed to be more directed for fixed income people" (PO J beneficiary). Others felt that program participants were chosen randomly, that eligibility was determined by a provider, or that the program was for "anybody that's old enough" (PO C beneficiary). Finally, some erroneously thought they had to meet certain criteria (not related to any actual VBID criteria) before becoming eligible for VBID benefits. For example, one PO E beneficiary with COPD thought that his oxygen levels would have to drop below a certain percentage before he could receive oxygen. ***Only 2 of the 63 beneficiaries (3 percent) who reported being aware of VBID realized that the VBID benefits were offered to those with certain diagnoses.*** As one PO H beneficiary put it: "I think it had a lot to do with congestive heart failure being in the records and the diabetes being in the records. . . . I would imagine that's how the insurance company felt about making sure certain people knew."

Benefits covered under VBID were another point of confusion. Some beneficiaries in PO J, for example, did not understand why some drugs were not included for lower cost-sharing as part of VBID. Several PO E beneficiaries did not understand why a service performed for a diagnosis that was not part of VBID would cost more (e.g., a chest CT is eligible for reduced cost-sharing

for COPD under VBID, but the CT scan would not be eligible if it was performed for another condition such as pneumonia). Some PO B beneficiaries who met CM/DM participation requirements reported not receiving reduced copays, most likely because their provider was not deemed "high value." As one PO B beneficiary explained, however, sometimes keeping one's current provider may be more important than receiving a lower copay, which requires switching to a new provider: "My specialist, he does not participate in the 10 dollar program. If I want to keep the doctor, I have to give him the regular price, which right now is 40 dollars. But my eye doctor, which I'm happy with, I pay 10 dollars instead of 40."

Finally, ***not all beneficiaries in our interview sample who were eligible for reduced cost-sharing reported receiving this benefit.*** Of the 30 beneficiaries in our interview sample eligible to receive reduced cost-sharing for prescription drugs, 17 (roughly three-fifths) reported getting this VBID benefit. Similarly of the 30 beneficiaries in our sample eligible to receive reduced cost-sharing for primary care visits, 14 (almost half) reported receiving them. Of the 50 beneficiaries eligible to receive reduced cost-sharing for specialist visits, 21 (a little over two-fifths) reported seeing lower copays for this service. Finally, nine of ten PO A and C beneficiaries reported receiving rebates. We note, however, that the information provided by beneficiaries did not always correspond to POs' data submissions. For example, some beneficiaries who reported receiving VBID benefits were classified by POs as having not met participation requirements.

The discrepancies between the interview and survey results about VBID awareness and the receipt of VBID benefits, however, are to be expected. Not only were the samples and questions different, the interviewers were instructed to probe about different aspects of VBID interventions implemented by a particular PO, to tailor the question wording to the nuances of the specific VBID intervention that beneficiaries were exposed to, and to ask clarification questions if something was not clear.

Intervention Satisfaction and Perceived Value of Value-Based Insurance Design Benefits

Parent Organization Perspective

Only two POs reported tracking beneficiary satisfaction with the VBID program, both through surveys of VBID participants. PO A, for example, conducted a survey of 30 to 50 beneficiaries (of roughly 1,300 participating beneficiaries) who received cost-sharing rebates as part of the VBID program. Among this small sample, results showed that beneficiaries were 100 percent satisfied with the program. Results from the survey fielded by PO C indicated that VBID-participating beneficiaries were generally satisfied with the program, but the PO noted that they received responses from only 10 to 15 percent of contacted beneficiaries.

Other POs reported anecdotal evidence of positive beneficiary experiences. For example, a representative of PO J stated: "I think there are a few members who are quite delighted that they

don't have any copays." PO G also relayed hearing positive feedback on reduced copayments and CM, noting that the beneficiaries are "very appreciative of the reduction in cost share and copays for their Part C and Part D benefits" and "the outreach for the case management team." According to representatives of PO I, some VBID beneficiaries and their relatives reported spending less time in the ED.

> *"We had a member that was . . . not checking their sugars, they were drinking two 2-liters of pop a day, they were eating doughnuts and not exercising, and by the time the case manager was done with him they cut out—they had lost 30 pounds, they cut out the soda, they were eating a healthier diet, and they were walking like two miles a day and stuff like that."*
>
> *PO B representative on positive beneficiary experiences with their VBID intervention*

Beneficiary Perspective

Beneficiaries expressed a range of opinions about the value and expected effect of VBID benefits, including reduction in copays, receipt of supplemental benefits, and participation in CM/DM programs.

Reduced Copays

Of the 63 VBID-participating beneficiaries we interviewed who reported being aware of VBID benefits, ***roughly half (N = 30) felt that reduced copays were valuable.*** Most notably, beneficiaries felt that reduced copays for medications and provider visits provided financial relief, particularly for those on a fixed income. As one PO H beneficiary put it, "If I save $100 a month, that's $100 I can use for something else." Others remarked that the financial incentives helped them afford the care they need: "The decrease in my heart medication [cost] was a lifesaver," said a PO J beneficiary. Some also discussed how the financial incentive motivated them to seek proper care and treatment. One PO C beneficiary said, "It gives me a push to do things the way I should be doing them. Even though I plan on doing it, it gives me the incentive to do it." For other beneficiaries, the financial incentive demonstrated that insurance companies are equally invested in beneficiaries' health. A PO D beneficiary stated that he was "very happy about [the program's] concern for [his] health and welfare."

At the same time, among the beneficiaries who reported being aware of VBID, ***one-third (N = 21) felt that reduction in copays was inconsequential.*** Some felt that the financial benefits of VBID were not strong enough for them to change their behaviors. One PO F beneficiary said, "We don't go to the doctors that often. So saving $25 a year is not going to motivate me." A PO D beneficiary stated that, for him, it was more important to stay with the same doctor, rather than save "a few bucks [on prescription drugs] a month, which doesn't make a difference." Other beneficiaries believed that, while the monetary gains were nice, they would have gone to the doctor and taken their medications regardless of an incentive: "I'm going to take my medicine

because I've been through too much. If I don't take care of myself, I'm not going to be here. My grandchildren won't let me," stated a PO H beneficiary.

These results are consistent with beneficiary responses to our questions about barriers to care, which showed that, although very important, ***financial concerns were not the only barriers to getting the care they need.*** For example, while some beneficiaries considered drug copays ($N = 23$) and provider copays ($N = 15$) to be barriers to care, others cited difficulties in scheduling appointments with specialists and PCPs ($N = 10$). Some beneficiaries were particularly concerned about the time it takes to get an appointment with a specialist: "The heart doctors, you would have to wait a month or two to see them" (PO J beneficiary). Others reported switching providers to be able to secure appointments in a timely manner: "I'm switching doctors now for the second time this year because . . . when I make an appointment, it's a month away, two months away and you can't get in on a problem situation" (PO J beneficiary). Ten beneficiaries cited transportation to and from medical appointments and pharmacies as a barrier. Some noted that they rely on family members and other caregivers to address their transportation needs: "I have to depend on somebody to get me [to my appointments]. Fortunately, I have a daughter that's willing to do that, and she just lives next door. But when she goes on vacation . . . I have to find somebody else to take me" (PO J beneficiary).

Supplemental Benefits

In 2018, only two POs provided supplemental benefits as part of their VBID designs (see Chapter 2 for more details). In providing feedback on the supplemental benefits they received, beneficiaries from PO B and PO I shared differing experiences. Many PO B beneficiaries did not seem to recall dental benefits, transportation, and diabetic retinopathy screening among the supplemental benefits offered under VBID. One beneficiary who was reported by PO B as having not yet completed VBID participation requirements described receiving "prediabetic training" but not diabetic retinopathy screening. None of the PO B beneficiaries mentioned receiving periodontal benefits. Others reported never being offered assistance with transportation, did not feel it was necessary, or did not want to "look into" this benefit. As one PO B beneficiary noted, "I think I can get some kind of transportation. But you got to run around and get some kind of cards or something for it. I usually can get a ride wherever I have to go. I don't have too much of a problem there." Another PO B beneficiary was frustrated with the amount of paperwork and communication required to get the transportation benefit.

In contrast, the supplemental benefits offered through PO I's telehealth program were largely met with positive feedback, in that participants did not find the equipment overly invasive or burdensome. Some even felt "safer" knowing that their metrics were being monitored daily, and others felt that the daily monitoring was a reminder to engage in healthier behaviors, such as quitting smoking. Appendix C provides a more detailed case study of the telehealth intervention offered by PO I.

Case Management/Disease Management Activities

The number of beneficiaries who reported that they found participating in CM/DM activities to be valuable (N = 11) was roughly the same number as those who found it to be inconsequential (N = 13). Beneficiaries participating in CM/DM programs requiring their ongoing engagement felt that periodic check-ins from a care manager were noninvasive, convenient, and helpful in terms of keeping them accountable for their health. Some spoke specifically about the relationship and trust they had built with their care manager and said that it "helps tremendously" to know that, if they had any questions, a helpful person was a phone call away.

> *"I like the idea that there's somebody there . . . just to have the benefit of them being there so that if you do have a problem . . . there's somebody you can [talk] to. . . . It's nice that they have a nurse that will work with you, that will coordinate if you're having problems with your healthcare providers and . . . will assist you in getting service and education."*
>
> *A beneficiary from PO F on the value of CM/DM activities*

Beneficiaries who did not find the CM/DM activities helpful, however, felt that they were already managing their own health themselves or with the help of their caretakers, and therefore did not need to participate in CM/DM. Others felt that, while it may be helpful to have a care manager to check in with, modifying health behaviors was too difficult at their age: "There isn't much I can change about my health behavior," said one PO F beneficiary. Similarly, some beneficiaries were very pessimistic about care managers' abilities to slow the course of their diseases, stating that "the damage is done, so there's nothing they can do. . . . You learn to work with it yourself" (PO E beneficiary).

In addition, ***several beneficiaries felt that the requirements to participate in CM/DM activities were too complicated, unnecessary, or unhelpful.*** For example, a PO J beneficiary said: "All these things require you to go to so many different places, have all the information, have all this stuff with me, and they basically know what's wrong and there isn't that much they can really do at this point." In rare cases, beneficiaries described how comorbidities, such as multiple sclerosis or cancer, limited their ability to engage in CM/DM activities.

Beneficiary Experiences of Care

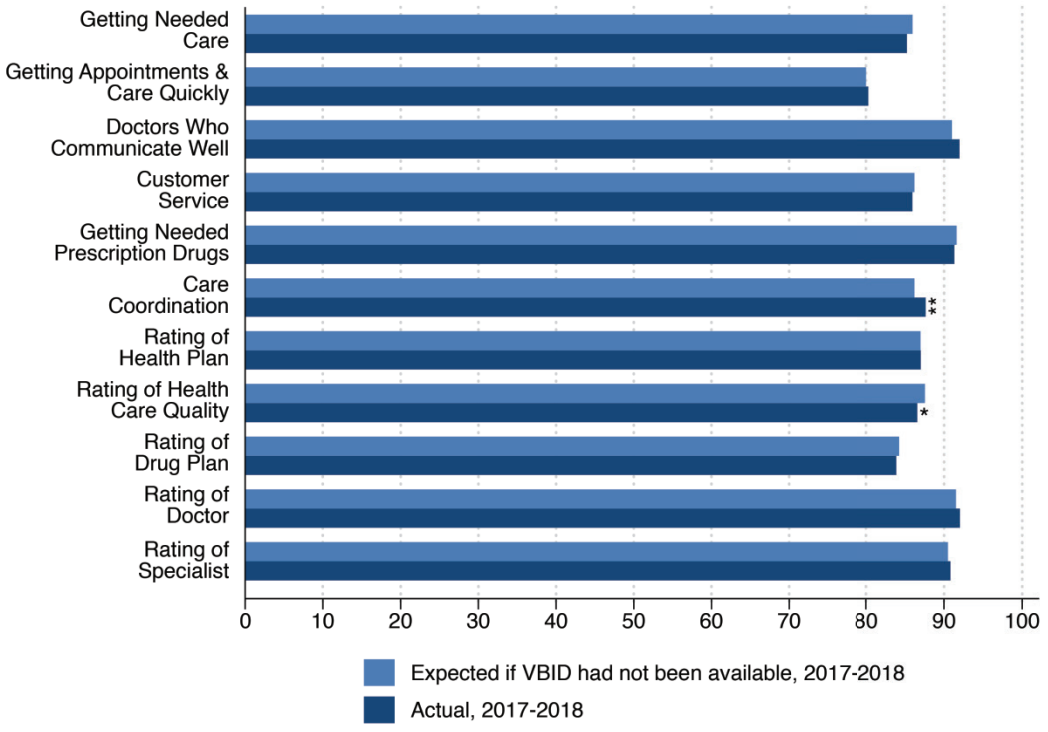
The MA & PDP CAHPS survey collects responses regarding patient experience with MA plans and summarizes these experiences with six composite measures and five overall rating measures. We used a case-mix adjusted difference-in-differences model to compare survey responses, scored 0–100, among VBID-eligible beneficiaries from 2014 to 2018 in VBID-participating PBPs with those of beneficiaries in matched comparison PBPs. A detailed

description of the models and results are found in Appendix E. Figure 5.2 provides the estimated effect of VBID by comparing the combined survey responses from 2017 and 2018 (in dark blue) with the expected values had VBID not been available (in light blue). Of the 11 measures analyzed, only care coordination has an estimated effect larger than 1 point, which also corresponded to the only statistically significant result (at the 5-percent level).

These results must be interpreted with caution; for 6 of the 11 measures, there was evidence that the pre-VBID-period trends were not parallel. This departure was characterized by a spike in scores from 2014 to 2015 among VBID-eligible beneficiaries in VBID-participating PBPs, followed by a drop between 2015 and 2016. More information on this issue can be found in Appendix E. Despite these limitations, the results of these analyses of beneficiary experiences of care suggest that if VBID had an effect, it was limited in scope and small in magnitude.

Although the positive effect of VBID on care coordination should be interpreted cautiously owing to methodological issues, the finding is consistent with the POs' perspectives on the importance of CM/DM and using financial incentives to engage beneficiaries in their own care. Indeed, several POs considered CM/DM, rather than financial incentives, to be a key component of the VBID interventions.

Figure 5.2. Comparison of MA & PDP CAHPS Survey Responses, Actual Versus Expected If VBID Had Not Been Available, 2017 and 2018



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. Sample sizes ranged from 7,757 to 19,289. See Appendix E for details.

Effects of Value-Based Insurance Design on Plan Benefit Package-Level Enrollment

Beneficiaries' experiences with care in the VBID model could have spillover effects on enrollment in participating PBPs. For example, enrollment might increase overall or among beneficiaries eligible for VBID if the benefits are perceived as attractive. Because POs were not allowed to market their VBID participation to new beneficiaries during the first two open enrollment periods following the announcement of the VBID model test, any effect of enrollment would have to have come through word of mouth. Our beneficiary interviews gave some indication that such word-of-mouth effects could have occurred.

> *"My wife and I talk to our circle of friends. We've had eight more people join this program that live within a five-mile radius to me. When they found out I was paying zero premiums and I showed them what my coverages were, what they're worth, and how much I was liable for, they couldn't believe it. They called, and eight more people, four couples, joined [this plan], including my sister and brother-in-law."*
>
> *A beneficiary from PO D on spreading the word about VBID and its benefits*

Because of VBID's potential to affect beneficiary enrollment, we used CMS administrative data to analyze several enrollment outcomes, including total enrollment, new enrollment, and enrollment among beneficiaries with select chronic conditions (COPD, CHF, diabetes, and hypertension). Our analyses focused on PBP-level enrollment from 2014 through 2019 and used a difference-in-differences approach to assess whether enrollment in VBID-participating PBPs trended differently from enrollment in matched comparison PBPs after the VBID-model test took effect. We found no statistically significant effects on enrollment in any year, a finding that is consistent with results reported in our first annual evaluation report (Eibner et al., 2018). A detailed summary of the enrollment analysis can be found in Appendix F.

Suggested Improvements to Value-Based Insurance Design: Beneficiary Perspectives

Given beneficiaries' experiences with VBID participation—from initial awareness to the actual use of VBID benefits—it is not surprising that some beneficiaries offered a few suggestions for improving their POs' intervention designs. Many beneficiaries, across multiple POs, mentioned the need to improve communication strategies, including offering more frequent, more explicit, and clearer information on the purpose of the intervention, enrollment procedures, and CM/DM participation requirements (if applicable). One PO F beneficiary noted, "I bet 90% of seniors don't look at it or read through [printed materials]," while a PO C beneficiary added, "Maybe they could give out a little bit more information that is a little more clear and simple.

Because if I don't understand it, and I'm a nurse, how many other people understand it even less?" A PO D beneficiary stated plainly, "If they send something and just say that, in bold print, 'Your copay[1] will remain the same. We are starting a program to include hypertension $0 copay.' I mean, I probably would have read it." Suggestions were also put forth to incentivize providers to spread the word and educate their patients on the program.

Beneficiaries we interviewed also suggested offering additional VBID-eligible services such as reduced prescription copays for POs that do not currently offer it as a VBID benefit, increasing the financial incentives for enrolling in CM/DM programs, offering stress-reduction therapy to help reduce hypertension, expanding transportation benefits, and offering reduced-cost gym memberships, home nurse visits, and state-of-the-art options for DME (such as smaller, more easily portable oxygen tanks).

Beneficiaries in VBID interventions with a CM/DM participation requirement also suggested ways to improve the CM/DM component by offering more options for interacting with care managers, such as adding in-person appointments, expanding care managers' availability to take telephone calls, and ensuring that—when a care manager reaches out by phone—caller identification clearly states that the incoming call is from the PO. Additional suggestions included providing beneficiaries with detailed paper reports of each patient–care manager encounter, noting that it was difficult to recall what was discussed with a care manager over the phone.

Summary

Beneficiary awareness of VBID was low in the MA & PDP CAHPS survey, with less than 12 percent of VBID-eligible beneficiaries reporting being offered lower copays or extra benefits because of a health condition. These low rates of awareness may have been related to the wording of the CAHPS questions. Almost two-thirds of the beneficiaries we interviewed were aware of VBID, though few recognized the VBID interventions without further prompting, such as descriptions of the VBID benefits offered. Among those familiar with VBID, however, many were confused about how VBID worked, the benefits available, and whether they had in fact received benefits for which they were eligible. Although two POs administered their own surveys to participating beneficiaries that indicated high satisfaction with their interventions, our interviews with beneficiaries underscored a more varied experience. Some reported that reduced copays were valuable and helped them get the care they needed, but others cited the need to address other barriers to care, including difficulties in scheduling appointments when needed and lack of transportation. The supplemental benefits offered by one PO, including transportation, retinal eye examinations, and certain dental benefits, garnered a mixed response from its

[1] Although this beneficiary used a term "copay," he or she most likely thought of premiums.

beneficiaries, yet a telehealth benefit offered by a different PO prompted an overall favorable response. Likewise, beneficiaries were split as to the helpfulness of the CM/DM activities that POs offered.

Our analysis of the MA & PDP CAHPS data indicated that most measures of beneficiaries' experiences of care were unaffected by VBID. One important exception to this finding was that VBID was associated with a small, statistically significant increase in beneficiaries' experiences with care coordination. Although these results should be interpreted with caution because of issues with the parallel trends for this outcome, the finding is consistent with the CM/DM approach adopted by most participating POs. We found no effect of VBID on enrollment.

Beneficiaries we interviewed suggested improvements that POs may want to make to their interventions, such as improving the clarity of communications on the benefits of participation, offering additional VBID-eligible services to further encourage participation, and expanding the opportunities to engage with CM/DM staff in participating PBPs.

6. Effects of Value-Based Insurance Design on Health Care Utilization

One of the goals of the VBID model is to encourage beneficiaries with chronic conditions to use high-value services. In many cases, POs eliminated or reduced cost-sharing for such services—for example, reduced copayments for endocrinologist visits for beneficiaries with diabetes. Two POs offered rebates for enrollees who took specific actions, such as completing preventive care screenings. In addition, many POs paired VBID with CM/DM programs, which could encourage the use of high-value services by providing information about care recommendations or by helping beneficiaries develop CM/DM plans. Although the VBID model did not allow POs to increase cost-sharing for low-value services, the model could affect the use of low-value or avoidable care through indirect channels. For example, beneficiaries who are more adherent to medication regimens may be at lower risk for ED visits or avoidable hospitalizations than nonadherent beneficiaries.

In this chapter, we assess how VBID influenced utilization of a range of health care services, including services eligible for reduced cost-sharing (which we call the VBID-targeted services) and general utilization, which included inpatient and ED visits.

Key Takeaway Points:

- VBID had a larger effect on utilization than any other outcome domain considered in this report, with statistically significant changes for 18 of 27 outcomes analyzed.
- VBID was associated with statistically significant increases in utilization for ten services targeted by VBID-participating POs.
- VBID was associated with statistically significant increases in primary care visits, specialist visits, prescription drug fills, and use of DME.
- VBID-eligible beneficiaries had fewer laboratory tests than matched comparators.
- Counter to expectations, there were small, statistically significant increases in the probability that eligible beneficiaries in VBID-participating POs had ambulatory care sensitive (ACS) ED or ACS inpatient admissions.

Value-Based Insurance Design-Targeted Services

POs that participated in the model test targeted a variety of services with their interventions (Table 6.1). In most cases, POs reduced copayments to encourage utilization of high-value services such as PCP or specialist visits. However, in a few cases, POs' strategies were more complicated. For example, PO A provided rebates for primary care and specialist visits for beneficiaries who completed a scorecard that involved several preventive screenings.

Table 6.1. Services Targeted by POs' VBID Interventions

PO	Category	Service
PO A	Laboratory tests	HbA1c test, fasting lipid profile, urine test, diabetic eye examination
PO A	Specialist visits	Endocrinology, podiatry, ophthalmology
PO A	PCP visits	Primary care
PO B	Supplemental	Transportation
PO B	PCP visits	High-value primary care visits
PO B	Specialist visits	Endocrinology, ophthalmology, nephrology, pulmonology, and podiatry (up to four)
PO B	DME	Diabetic supplies
PO D	Drugs	Selected tier 1–3 drugs
PO E	Specialist visits	Cardiology, pulmonology, palliative care, sleep specialists
PO E	Laboratory tests	Pulmonary function test
PO E	Therapy/rehab	Oxygen therapy treatment and pulmonary rehab
PO E	Laboratory tests	CT chest scans and sleep studies
PO F	PCP visits	Primary care
PO F	Specialist visits	Cardiology and pulmonology
PO G	PCP visits	Primary care
PO G	Specialist visits	Cardiology
PO G	Drugs	Selected tier 1 drug
PO H	Specialist visits	Endocrinology, cardiology, podiatry

NOTE: The full list of services is in the appendix. We did not analyze all targeted services due to sample size issues. PO C does not appear in this list because it offered rebates for any incurred cost-sharing, conditional on participation in CM/DM; PO I does not appear because it did not have benefits that could be tracked in the encounter data.

In this section of the report, we use CMS encounter data to estimate VBID's association with eligible beneficiaries' use of the targeted services shown in Table 6.1. As mentioned in earlier chapters, most POs included CM/DM participation requirements as part of their VBID interventions. However, CM/DM programs are not tracked in encounter data, so we were not able to analyze changes in the use of these programs. Similarly, we did not analyze data for PO I, which offered supplemental benefits that were not tracked in the encounter data, nor did we analyze targeted services for PO C, since it did not target any specific services (and instead used rebates for any incurred nonpharmacy cost-sharing). In addition, we were unable to analyze dental encounters for PO B because not enough encounters were reported in the data.

Regression Results

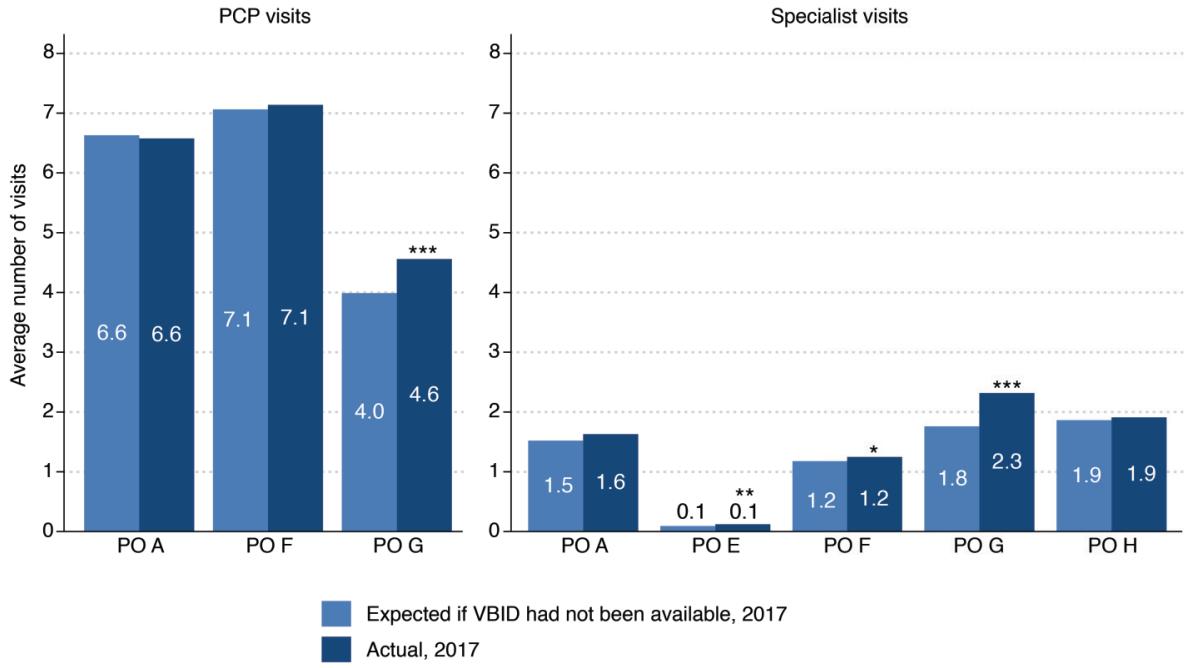
We report descriptive results in Appendix G for the VBID-targeted services. ***To estimate the causal effect of VBID on utilization of targeted services, we used a difference-in-differences regression approach*** (Appendix D contains the full methodological description of this approach). This methodology involves comparing utilization trends among eligible beneficiaries in VBID-participating POs with utilization trends among a matched comparison group. VBID PBPs were matched to comparison PBPs on characteristics such as whether the plan offers Part D, PBP size, and service area characteristics. Within the matched set of PBPs, beneficiaries were matched on characteristics such as risk score, gender, and age, using a propensity score approach. For each

VBID-targeted service, we ran a difference-in-differences regression that analyzed whether utilization trends shifted for treated beneficiaries relative to comparison beneficiaries in 2017, the year that the model test took effect.

The assumption that trends in outcomes would have been parallel for treatment and comparison beneficiaries in the absence of VBID is critical for the difference-in-differences methodology. Although it is impossible to know whether trends after 2017 would have been parallel, we can determine the reasonableness of the parallel trends assumption by analyzing whether trends in outcomes were parallel during the pre-VBID period (2014–2016). Some outcomes were not parallel in the pre-VBID period. With weighting, the trends often improved, in which case we present weighted difference-in-differences estimates. We imputed several covariates that were missing data and would have otherwise dropped beneficiary-year observations from the model. For two outcomes for PO B (DME and transportation), there was not enough utilization of these services in the comparison group, so we do not report results for these outcomes. In total, we analyzed 16 VBID-targeted services across seven POs. A more detailed summary of our analyses, including tests of parallel trends and full regression results, can be found in Appendix G.

Figure 6.1 summarizes the changes in PCP and specialist visits targeted by POs A, E, G, F, and H. The graphs show actual utilization of the targeted services compared to the regression-

Figure 6.1. Comparison of PCP and Specialist Visits for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017

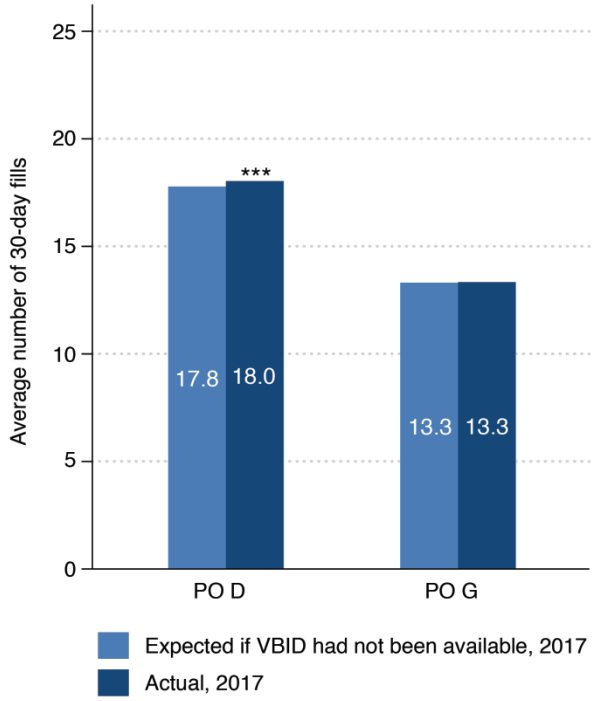


NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

based prediction of what utilization would have been in the absence of the VBID-model test; we use the comparison group to predict utilization in the absence of VBID. PO G experienced statistically significant increases of approximately 0.57 PCP visits (from 3.99 to 4.56 visits) and 0.56 specialist visits (from 1.76 to 2.32 visits) per-beneficiary per-year. Beneficiaries in PO E and F also experienced small, but statistically significant increases in specialist visits.

POs D and G targeted prescription drugs in their interventions. Figure 6.2 shows that there were no statistically significant changes in targeted prescription drug utilization for PO G, but there was a statistically significant increase of 0.25 30-day fills (from 17.78 to 18.03 fills) for PO D. However, pretrends for prescription drug use among beneficiaries in PO D were not parallel with their matched comparators (see Appendix G), so these results should be interpreted with caution.

Figure 6.2. Comparison of Prescription Drug Utilization for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017

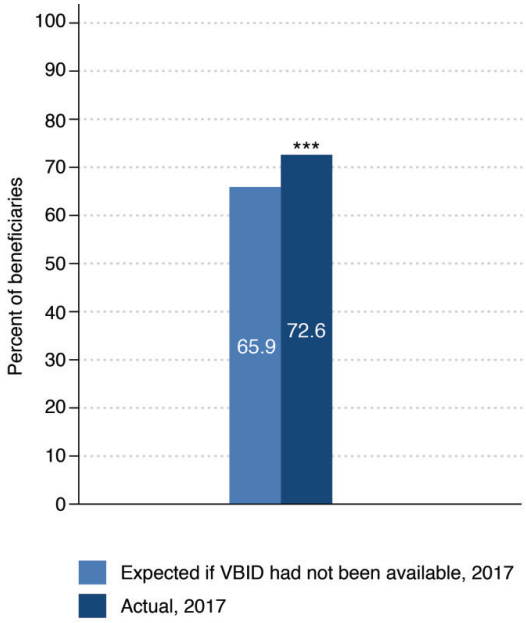


NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

The proportion of VBID-eligible beneficiaries in PO A who completed all four diabetes-related preventive services increased approximately 6.7-percentage points (Figure 6.3) and was statistically significant at a 1-percent significance level.

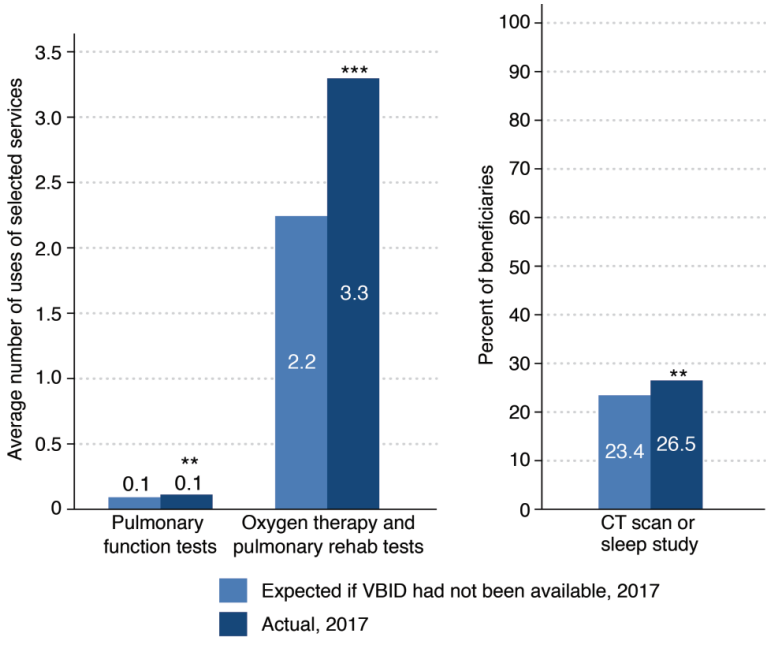
VBID-eligible beneficiaries in PO E utilized more oxygen therapy and pulmonary rehabilitation services than matched comparators (Figure 6.4)—a change of one claim

Figure 6.3. Comparison of Preventive Services for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017, PO A Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results. The four diabetes-related preventative services are an HbA1c test, lipid profile, eye examination, and foot check.

Figure 6.4. Comparison of Other Services for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017, PO E Only

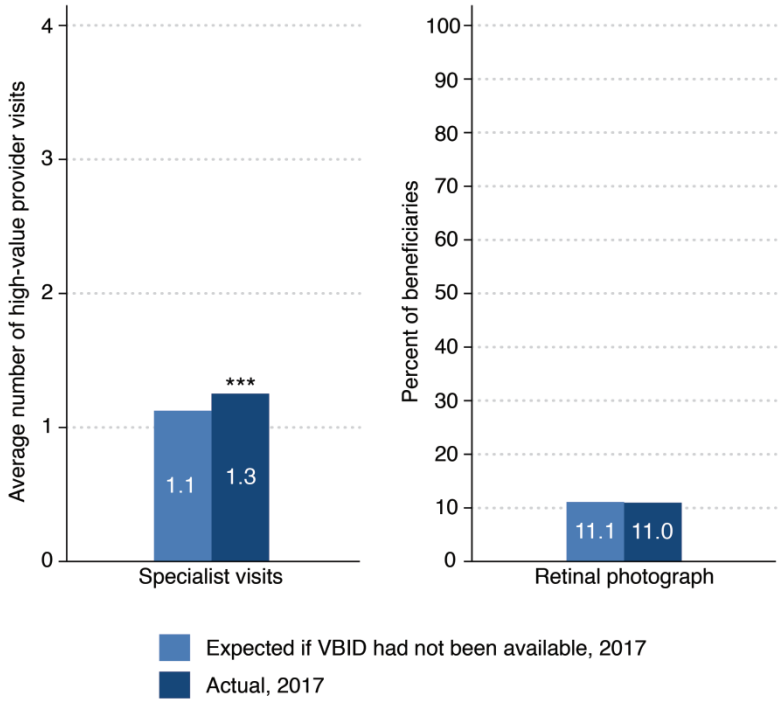


NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

per-beneficiary per-year, which was statistically significant. Beneficiaries in PO E also used significantly more treatments for COPD (oxygen therapy and pulmonary rehab) and CT scans and sleep studies.

For PO B, we found a statistically significant increase in the number of high-value specialist visits of approximately 0.13 visits per-beneficiary per-year (1.12–1.25). There were no changes in retinal photography use (see Figure 6.5).

Figure 6.5. Comparison of Other Services for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017, PO B Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

Overall, we found statistically significant increases for the following 10 of 16 VBID-targeted services that we were able to assess using regression analyses:

- Proportion of beneficiaries completing scorecards in PO A (6.7-percentage point; confidence interval [CI]: 2.9–10.5)
- High-value provider visits in PO B (0.13 visit; CI: 0.07–0.18)
- Selected prescription drugs in PO D (0.25 fills; CI: 0.13–0.38)
- Specialist visits in PO E (0.03 visits; CI: 0.00–0.05)
- Proportion of beneficiaries accessing CT scans and sleep studies in PO E (3.1-percentage points; CI: 0.5–5.6)
- COPD treatments in PO E (1.05 services; CI: 0.86–1.25)

- Pulmonary function tests in PO E (0.02 tests; CI: 0.00–0.04)
- Specialist visits for PO F (0.07; CI: 0.02–0.12)
- PCP visits in PO G (0.57 visits; CI: 0.35–0.79)
- Specialist visits in PO G (0.56; CI: 0.32–0.80).

Parent Organization Assessment of Value-Based Insurance Design-Targeted Services

Although we found increases in utilization for most VBID-targeted services, ***POs were uncertain that VBID was having a meaningful effect on outcomes.*** For example, a representative from PO F told us that they "found that copay did not appear to be a significant barrier to people getting those [PCP and specialist] visits. So when we removed the copay and reduced the copay to specialists, we did not see that the enrolled population had an increase in utilization of the PCP and specialist visits." Representatives from PO C mentioned that the multiple occurring programs to encourage utilization of certain services within their plans made it difficult to separate the effect from VBID from other programs. It may also have been that CM/DM was taking the place of some physician or other visits, as a representative of PO F explained: "So my hypothesis is that they're more in care management, meaning there is less of a need for them to see a doctor, right?"

PO G speculated that its intervention did not lead to significant changes in utilization because it was not able to engage beneficiaries who were not adherent to care protocols: "For the engaged members, there wasn't a whole lot of change because a lot of the time, those who opted into the program are those people who were [already] going for the specialist and using their cardiologist." This PO also speculated that it did not see changes in drug utilization because many of its beneficiaries were either already taking the targeted medications or taking medications not affected by the copayment change and did not want to switch. Other POs were still waiting for additional data to determine the directions of effects: "Overall it doesn't look like there are significant positive results . . . it's inconclusive, just to be honest" (PO J).

It is possible that POs' impressions differ from our quantitative findings because our estimates are extremely small. For example, the changes in utilization that we estimated typically represented average increases of less than one visit or service use per year. From the PO perspective, these small increases may not be meaningful or even detectable. It is also possible that our findings differ from the results of POs' internal evaluations because of methodological differences. Although we used out-of-state comparators, during the interviews, POs told us that they used different comparison groups, such as VBID eligible but not participating beneficiaries or beneficiaries in another PBP not eligible for VBID.

Many POs also stated that their participation in the model test was valuable to test benefit design changes, regardless of whether there were changes in utilization. "[W]e would probably think it's a success either way: either we learn something or we are generating the savings," said PO H representatives. As PO C stated: "We are using the VBID as a playground to learn. So if we start seeing things that we need to incorporate into the rest of our population we say, okay,

let's make it a standard practice." Finally, PO F representatives stated that their participation will be useful in evaluating whether to pursue these benefit designs in the future.

> *"Even if it didn't achieve the desired results, we wouldn't look at it as a failure so much as it really helps to inform the decisions that we make about benefit design in that group with the different teams in our area to help us make those decisions for [PBPs in] future states."*
>
> *PO F representative on the benefits of participating in the MA VBID model test*

General Utilization

In addition to VBID-targeted services, we analyzed whether VBID affected utilization of several general categories of health care services, such as primary care visits, specialist visits, ED use, and prescription drugs. VBID could have affected nontargeted outcomes for a variety of reasons. Some services may be complementary; for example, a VBID intervention that encourages more interactions with PCPs could lead to greater use of recommended drugs, even if drugs are not directly targeted by the intervention. Greater use of recommended services could also reduce the use of some types of care, such as avoidable hospital inpatient stays and ED visits.

The importance of CM/DM in the MA VBID model test makes it different from many traditional VBID models implemented in the commercial market that primarily focus on financial incentives (reduced copays for drugs, for example). POs argued that financial incentives were a way to increase beneficiary engagement in CM/DM activities, which, in turn, could directly affect the use of both high- and low-value services: "The financial part is the carrot, but the long-term behavior change comes from the care management," said a PO B representative. CM/DM may have had an effect that is separate from or magnifies the effect of financial incentives on service utilization.

Table 6.2 shows the 11 general services we considered, as well as the hypothesized direction of the VBID effect. In contrast to the VBID-targeted services, which we hypothesized should all increase under VBID, hypothesized effects on general utilization depend on the outcome considered. Because most VBID programs aimed to increase utilization of primary care, specialty care, and prescription drugs, we hypothesized that utilization of these services might increase among VBID-eligible beneficiaries. In contrast, ED visits and inpatient care could decrease among VBID beneficiaries—for example, if better CM/DM and medication adherence reduced complications that could lead to emergencies and hospitalizations. There are categories of services where we did not have hypotheses about how and whether VBID should lead to changes, namely skilled nursing facility (SNF) use, home health visits, and laboratory tests. Going to the doctor more frequently, for example, may result in more tests or procedures, and it is not clear that this is a desirable outcome.

Table 6.2. Summary of General Utilization Measures and Hypothesized Direction of Effect

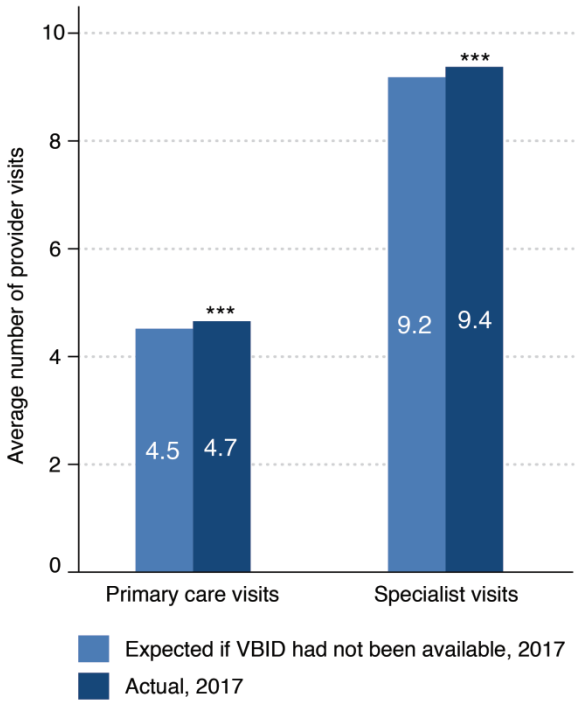
Outcome Number	Utilization Category	Definition (per Beneficiary per Year)	Hypothesized Direction of Effect
1	Inpatient, all-cause	Proportion of beneficiaries with at least one inpatient stay for any reason	Decrease
2	ACS inpatient	Proportion of beneficiaries with at least one inpatient stay for ACS conditions	Decrease
3	ED, all-cause	Proportion of beneficiaries with at least one ED visit for any reason	Decrease
4	ACS ED	Proportion of beneficiaries with at least one ED visit for ACS conditions	Decrease
5	Office visits, primary care	Number of primary care visits	Increase
6	Office visits, specialists	Number of specialist visits	Increase
7	SNF	Proportion of beneficiaries with at least one SNF visit	Uncertain
8	Home health	Proportion of beneficiaries with at least one visit	Uncertain
9	Laboratory tests	Number of claims for laboratory tests	Uncertain
10	DME	Number of DME claims	Increase
11	Prescription drug	Number of 30-day fills, all active ingredients combined	Increase

Regression Results

We estimated the causal effect of VBID on general service utilization using difference-in-differences methods as described in Appendix G (where the descriptive analyses of these outcomes can also be found). As with the VBID-targeted services analyses, this approach involved comparing trends in service utilization among VBID-eligible beneficiaries with trends in service utilization among matched comparators. However, for the general services analysis, we ran a single regression for each outcome, combining beneficiaries from all participating POs (and their matched comparators) into a single analysis. The sample size for the general utilization regressions is 134,397 (VBID and comparison beneficiaries), except the Part D analyses which only include PBPs offering Part D, so the sample size for these regressions is reduced to 128,887.

VBID-eligible beneficiaries experienced a small but statistically significant increase in the number of PCP visits relative to beneficiaries in the matched comparison group (Figure 6.6). The magnitude of the increase, 0.14 visits, corresponds with a 3-percent increase in service utilization among eligible beneficiaries compared to the value that was expected in the absence of VBID. ***Specialty care visits also increased for VBID-eligible beneficiaries.*** The magnitude of the increase was also small but statistically significant—0.19 of a visit per-beneficiary per-year, which corresponds to about a 2-percent increase in specialty care visits.

Figure 6.6. Comparison of Primary Care Provider (PCP) and Specialist Visits for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017



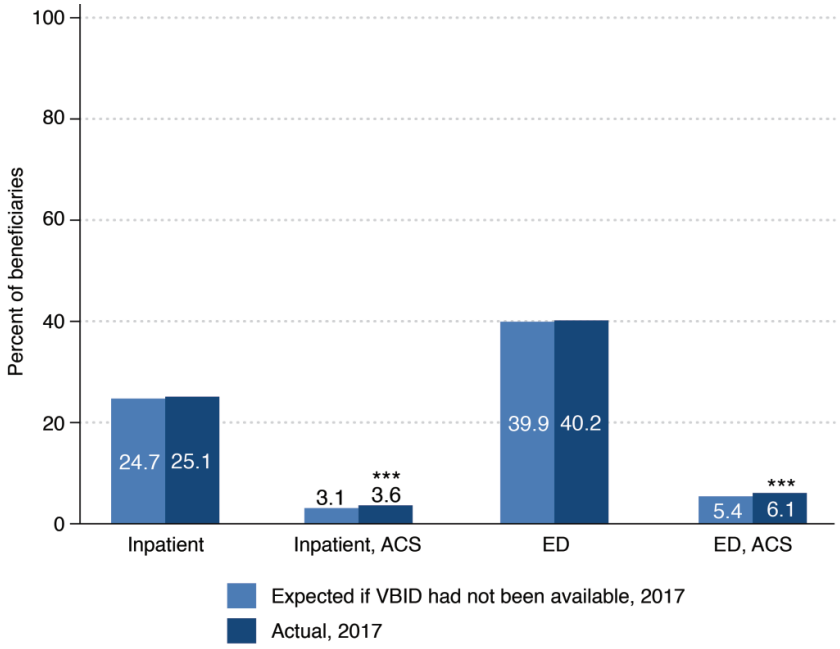
NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

There were no statistically significant changes in the probability of having any ED or inpatient visit (Figure 6.7). However, VBID-eligible beneficiaries experienced a small but statistically significant increase in the probability of having an ACS inpatient stay or ACS ED encounter (Figure 6.7). ACS visits are defined as visits that might be reduced or avoided if patients received more timely care in outpatient settings and represent a small share of overall ED and inpatient visits.[1] The probability of having an ACS inpatient visit increased by 0.55-percentage points, and the probability of an ACS ED visit increased by 0.67-percentage points (CMS, 2015).

The proportion of beneficiaries with at least one SNF visit decreased 0.4-percentage points. The proportion of beneficiaries with any home health visits was unchanged (Figure 6.8). However, previous work suggests that the encounter data for SNF and home health service categories may be less completely reported by POs than other service categories (Medicare Payment Advisory Commission, 2019).

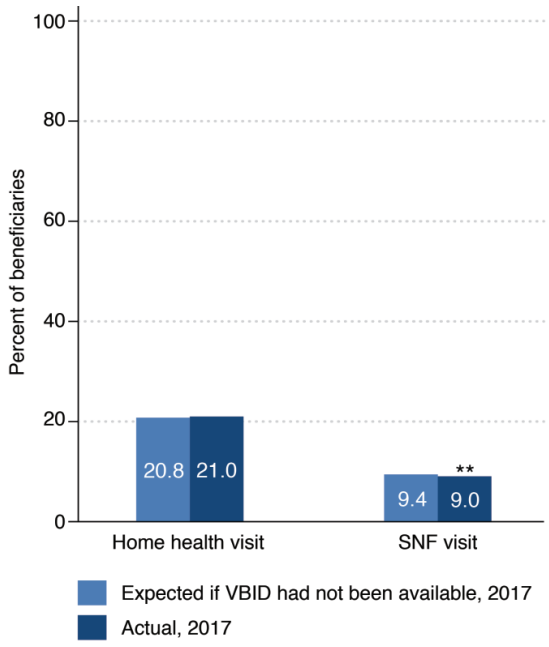
[1] We define ACS ED and ACS inpatient visits as visits for ACS conditions measured using the Acute Conditions Composite measure (CMS, 2015). This composite measure focused on bacterial pneumonia, urinary tract infection, and dehydration. More detailed information on this measure can be found in Appendix G.

Figure 6.7. Comparison of Proportion of Beneficiaries with an ED or Inpatient Visit for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

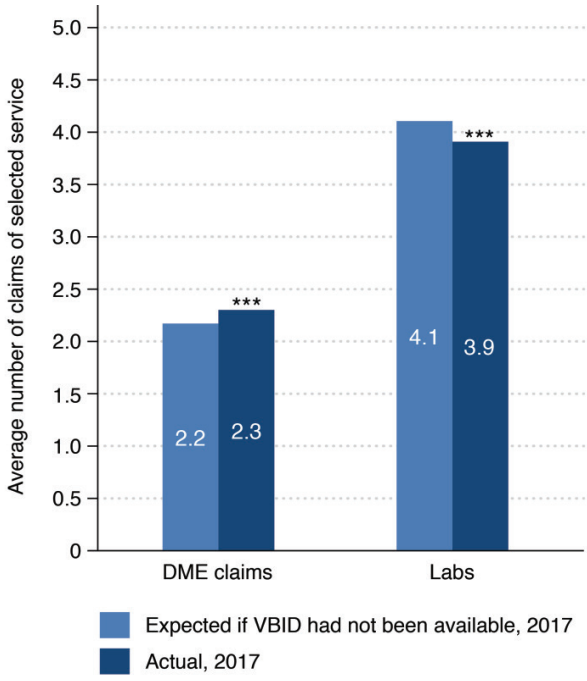
Figure 6.8. Comparison of Home Health and SNF Utilization for VBID-Eligible Beneficiaries with at Least One Visit, Actual Versus Expected If VBID Had Not Been Available, 2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

The number of claims for DME increased from 2.2 to 2.3 claims per-beneficiary per-year, whereas the number of laboratory tests decreased from 4.1 to 3.9 laboratory tests per-beneficiary per-year (Figure 6.9). Both changes were statistically significant. DME claims may also be less reliably reported than other service categories such as inpatient, outpatient, and professional visits (Medicare Payment Advisory Commission, 2019; Mulcahy et al., 2019).

Figure 6.9. Comparison of DME and Laboratory Test Utilization for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017



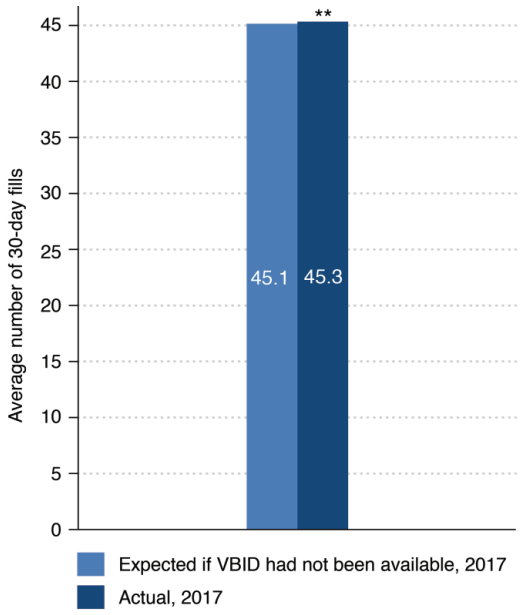
NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

The use of prescription drugs (30-day fills of all drugs) for VBID-eligible beneficiaries increased from 45.14 to 45.33 fills per-beneficiary per-year (a 0.19 of a fill increase, Figure 6.10).

Sensitivity Analyses

In Appendixes J and K, we explored whether VBID-eligible beneficiaries in POs with specific types of interventions had different utilization outcomes compared to other VBID-eligible beneficiaries. We also reran some of the analyses presented above for subsets of POs with specific types of VBID interventions. Appendix J focuses on the two POs with Part D interventions (POs G and D), and Appendix K focuses on the seven POs with participation requirements (POs A, B, C, E, F, G, and I). To improve the comparability of beneficiaries in POs with different VBID interventions, some of the analyses in Appendixes J and K focus on a

Figure 6.10. Comparison of Number of 30-Day Fills for VBID-Eligible Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. See Appendix G for full regression results.

more limited sample than used in this chapter. Based on this restricted sample, the analyses suggest that the increases in 30-day drug fills were driven primarily by POs with Part D interventions, while increases in PCP and specialist visits were driven by POs with participation requirements (typically requirements to participate in CM/DM).

POs with Part D Interventions: Relative to VBID-eligible beneficiaries in other POs, we found that eligible beneficiaries in POs with Part D interventions had increases in 30-day drug fills and reductions in ED and inpatient utilization after the VBID model test took effect. Relative to matched comparators in non-VBID POs, VBID-eligible beneficiaries in POs with Part D interventions had increases in 30-day drug fills after VBID took effect.

POs with Participation Requirements: Eligible beneficiaries in POs with participation requirements experienced reductions in 30-day drug fills, increases in PCP visits, and increases in inpatient stays after VBID took effect relative to VBID-eligible beneficiaries in other POs. Relative to matched comparators in non-VBID POs, VBID eligible beneficiaries in POs with participation requirements experienced increases in ED utilization, inpatient utilization, specialist visits, and PCP visits after VBID took effect.

Overall, the effects for POs with Part D interventions were consistent with the prior literature, with increases in medication utilization and declines or no changes in hospitalization and ED use. In contrast, POs with participation requirements had mixed effects, with statistically significant increases in primary care and specialty care utilization, statistically significant increases in inpatient utilization, and (borderline) statistically significant increases in ED use relative to matched comparators.

In Appendix L, we used data from both treatment and comparison beneficiaries to estimate the price elasticity of demand for primary care in our sample. We found that the elasticity was much lower than prior estimates, consistent with our finding that VBID beneficiaries' responses to copay changes were small. These relatively limited responses to price changes may reflect that, for the chronically ill beneficiaries in our sample, medical care was viewed as a necessity.

Summary

Of 27 utilization outcomes analyzed, we found statistically significant changes for 18. We found increases in utilization for most services directly targeted by POs' VBID interventions (10 statistically significant results of 16 analyses for which we were able to run regressions). These increases were consistent with expectations and align with prior literature showing that reductions in cost-sharing are associated with higher utilization. However, the increases were generally small, and POs did not typically report increases in utilization among their beneficiaries.

Of the 11 general utilization outcomes that we considered, eight had statistically significant changes. As hypothesized, we found statistically significant increases in PCP visits, specialist visits, prescription drug use, and DME use. Although the absolute changes in these outcomes were small, they could represent meaningful changes for a portion of the population. For example, the 0.14 increase in PCP visits per-beneficiary per-year represented about 14,000 additional visits for the approximately 100,000 VBID-eligible beneficiaries in 2017. The 0.19 increase in specialist visits per-beneficiary per-year translated to 19,000 additional specialist visits. VBID was also associated with a decline in laboratory tests, an outcome with an ambiguous relationship to VBID. On the one hand, several POs targeted laboratory tests with their interventions, which would be expected to increase use. On the other hand, better CM/DM could lead to reductions in duplicative or unnecessary laboratory test services.

Counter to expectations, visits for ACS inpatient and ED visits also increased, though the magnitudes were small—less than 1-percentage point for either ACS outcome measure. The 0.54-percentage-point increase in ACS inpatient visits translated to an additional 540 beneficiaries having any ACS inpatient visit, and the 0.67-percentage-point increase in ACS ED visits translated to an additional 670 beneficiaries having an ACS ED encounter. It may be that increased interactions with physicians led to more diagnoses and referrals, which could increase utilization of other services, including inpatient and ED services. In addition, as we noted in Chapter 5, many beneficiaries were confused about which benefits were eligible for reduced cost-sharing and could have incorrectly assumed that VBID benefits extended to inpatient and ED care.

In sensitivity analyses, we found that increases in prescription drug fills were driven primarily by POs with Part D interventions, whereas increases in primary and specialty care visits were driven by POs with participation requirements (usually requirements to participate in CM/DM). Increases in ED and inpatient use also appeared to be concentrated in POs with participation requirements.

7. Effects of Value-Based Insurance Design on Health Care Quality and Health Status

Through its incentives to encourage the use of high-value services, the VBID model has the potential to improve health care quality, which could be reflected by increased delivery of recommended care by providers, as well as increased adherence to treatment by beneficiaries. Several mechanisms could lead to improvements in the quality of care. As shown in the previous chapter, VBID interventions modestly increased beneficiaries' contact with PCPs, which could increase the number of opportunities to deliver recommended care for individuals with VBID-targeted chronic conditions (e.g., periodic hemoglobin A1c testing for diabetics). Contact with care managers, which was included in many VBID interventions, could also spur higher-quality care. Likewise, increased care delivery could increase rates of preventive health care screenings that are recommended but not targeted specifically by VBID. Ultimately, more interactions with high-value providers and enhanced quality of care for chronic medical conditions could lead to better health outcomes.

Key Takeaway Points:

- VBID had no effect on any of the contract- or beneficiary-level measures of health care quality or medication adherence in the first two years of the model test.
- VBID had no effect on any of the self-reported health status measures examined, and it had no effect on beneficiary risk scores in the first two years of the model test.
- We were unable to draw firm conclusions as to VBID's effect on the risk of death.
- Interviews with POs suggested they were not necessarily expecting to see substantial changes in quality or health outcomes at this point in the model test.
- Some POs mentioned that desirable results of participating in VBID, at the plan level, may include a lack of negative effects of the VBID intervention on health outcomes, or helping beneficiaries reach their personal health goals.

Health Care Quality/Adherence

We used MA and PDP Star Ratings data at the contract level to assess the effect of VBID on health care quality. Contract-level measures that we analyzed included the overall Star Rating and composite indexes focused on VBID-targeted conditions, diabetes care, and an index focusing on the quality of care for general preventive services that were not explicitly targeted by VBID. As detailed in Appendix H, the diabetes care index was based on three HEDIS measures related to diabetes monitoring and blood sugar control, together with one measure of diabetes medication adherence; the VBID-targeted conditions index was based on one HEDIS measure related to blood pressure control, together with two measures of adherence to medications for

hypertension and cholesterol; and the general medical care index was based on six HEDIS measures related to recommended preventive care (i.e., cancer screening, weight management) or conditions that were not targeted by VBID (i.e., osteoporosis, rheumatoid arthritis). Contract-level measures were pooled across all beneficiaries and PBPs in a given contract.

We also used beneficiary-level PDE data to assess the effect of VBID on beneficiary adherence to medications for three categories of targeted medical conditions—diabetes, hypertension, and high cholesterol (cholesterol-lowering medications are typically recommended in the management of diabetes and CHF, both of which are VBID-targeted conditions)—and a beneficiary-specific composite measure of the percentage of these medication adherence measures satisfied. We also assessed receipt of breast cancer screening at the beneficiary level using HEDIS data, as a measure of general preventive services for female beneficiaries. Table 7.1 summarizes the quality/adherence outcome measures that we used in the analyses.

Table 7.1. Summary of Quality/Adherence and Health Status Outcome Measures

Outcome Category	Outcome Measures	Study Population
Quality/adherence (contract level)	Overall Star Rating	Contracts containing VBID-participating PBPs or matched comparison PBPs
	VBID-targeted condition index	
	Diabetes index	
	General medical care index	
Quality/adherence (beneficiary level)	Medication adherence for diabetes medications	VBID and matched comparison beneficiaries with diabetes
	Medication adherence for hypertension (RAS antagonists)	VBID and matched comparison beneficiaries with hypertension or CHF
	Medication adherence for cholesterol (statins)	VBID and matched comparison beneficiaries with diabetes or CHF
	Percent of eligible measures satisfied	VBID and matched comparison beneficiaries with diabetes, CHF, or hypertension
	Breast cancer screening	Female VBID and matched comparison beneficiaries
Health status (beneficiary level)	VR-12 PCS	VBID and matched comparison beneficiaries
	VR-12 MCS	
	Activities of Daily Living (ADL) Limitations	
	Independent Activities of Daily Living (IADL) Limitations	
	Number of Days of Poor Physical Health	
	Number of Days Where Health Limited Usual Activities	
	Smoking Status	
	HCC/RxHCC	
	Mortality	

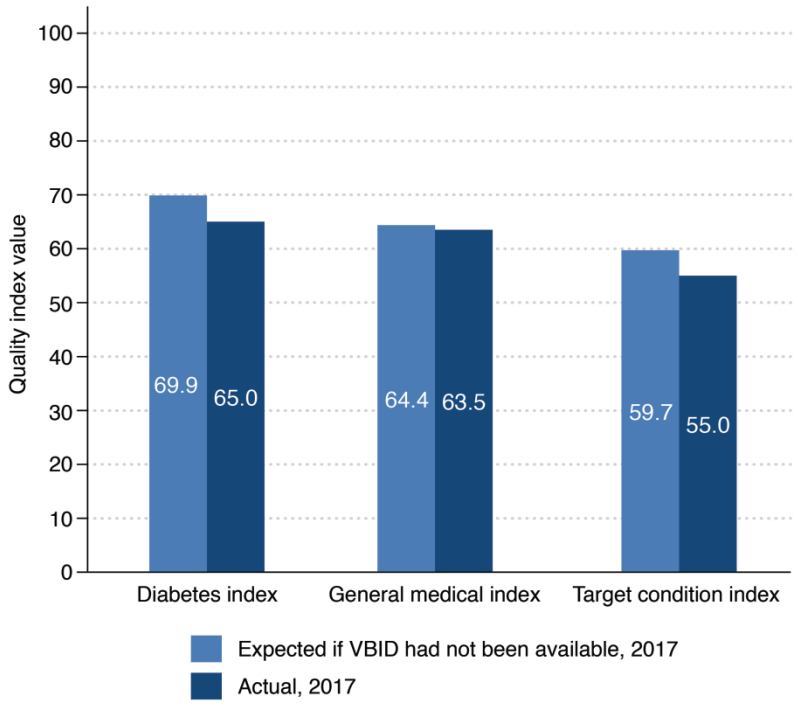
Abbreviations: RAS = renin angiotensin system; VR-12 PCS = Veterans RAND 12-Item Health Survey Physical Component Score; VR-12 MCS = Veterans RAND 12-Item Health Survey Mental Component Score.

Regression Results

Our difference-in-differences regression analyses examined whether VBID caused improvements in the quality of care and medication adherence. Figures 7.1 and 7.2 display the difference-in-differences findings for all outcome measures except for the overall Star Rating, which we describe in the text. For continuous outcome measures (i.e., all measures in Figure 7.1 and the percentage of medication adherence measures satisfied in Figure 7.2), the figures compare the observed values of the measures among VBID contracts and beneficiaries in 2017 or 2018 (dark blue bar) with the expected values in the absence of VBID in the same years (light blue bar). For discrete outcome measures (i.e., all other measures in Figure 7.2), the figures compare the predicted probability of the outcomes of interest among VBID contracts and beneficiaries in 2017 or 2018, both in the presence (dark blue bar) and absence (light blue bar) of VBID.

No statistically significant effects of VBID were found on any of the quality/adherence measures examined. Regression coefficients from these models are presented in Appendix H and show the size and direction of effects.

Figure 7.1. Comparison of Quality Index Values in VBID Contracts, Actual Versus Expected If VBID Had Not Been Available, 2017



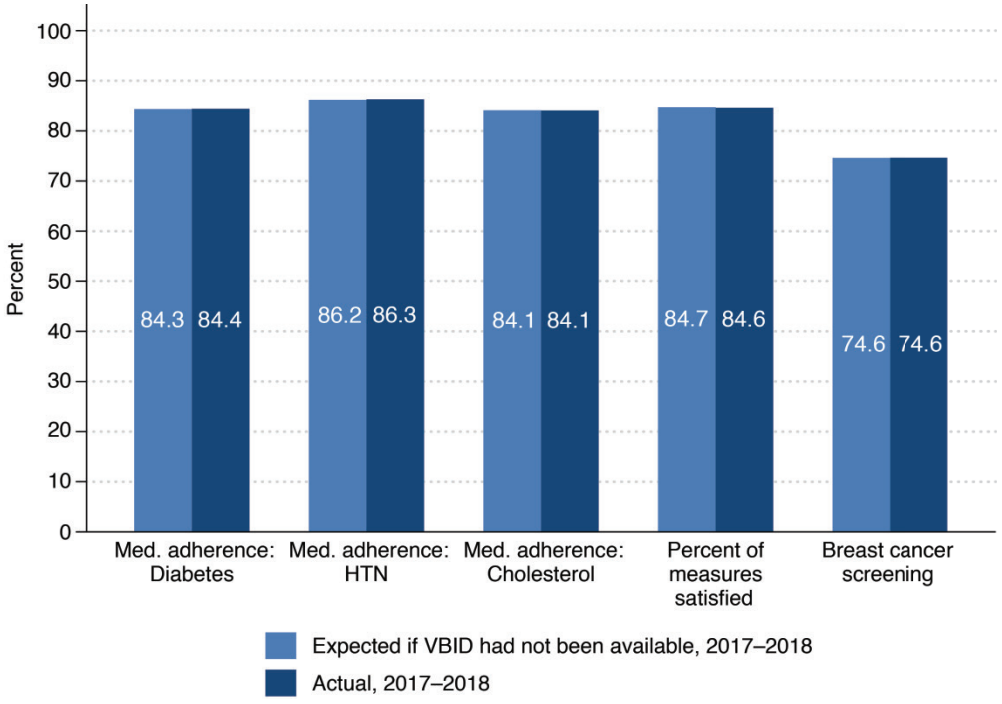
NOTE: Results are derived from difference-in-differences models comparing VBID contracts with comparison contracts. Differences are not statistically significant at conventional levels. See Appendix H for details.

We first examined the effect of VBID on overall Star Ratings of VBID contracts. We found no statistically significant differences in overall Star Ratings as a result of VBID (see Appendix H for regression coefficients). We note that, to be eligible for VBID, participating POs were required to have a minimum Star Rating of 3.5. Given that the required baseline performance was high and relatively close to the maximum possible overall Star Rating (i.e., 5), it is possible that ceiling effects might have constrained our ability to detect improvements in the overall Star Rating, especially given our limited sample size for contract-level measures.

Figure 7.1 shows that there were no statistically significant differences between the actual values of our quality and adherence indexes among VBID contracts in 2017 and the values that would be expected in the absence of VBID.

Figure 7.2 shows that, among VBID beneficiaries, there were no statistically significant differences between the actual values of the medication adherence measures in 2017 and 2018 and the values that would be expected in the absence of VBID. Similarly, there were no statistically significant differences in actual rates of breast cancer screening in 2017 among VBID beneficiaries compared to the values that would be expected in the absence of VBID.

Figure 7.2. Comparison of Beneficiary-Level Quality/Adherence Measures Among VBID Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017 and 2018



NOTE: Results are derived from difference-in-differences models comparing VBID beneficiaries with comparison beneficiaries. For all measures except breast cancer screening, results represent average effects for both 2017 and 2018. For breast cancer screening, results represent effects in 2017 only. Differences are not statistically significant at conventional levels. See Appendix H for details.

Health Status

Finally, we explored the effect of VBID on health outcomes, including self-reported health status, risk scores (HCC and RxHCC scores), and mortality. Self-reported health measures are drawn from the Medicare HOS, an annual longitudinal survey of a sample of MA beneficiaries (see Appendix H for more details). HCC and RxHCC scores are risk adjustment scores that are used to adjust payments made by Medicare to MA and Medicare Advantage and Part D (MA-PD) PBPs, such that payments more closely reflect the health status and expected medical spending (HCC) and prescription drug spending (RxHCC) of beneficiaries. VBID could, in principal, contribute to improvements in health status through increased contact with high-value providers, enhanced quality of care for chronic medical conditions, and increased medication adherence. Improvements in health status could, in turn, be reflected by improvements in self-reported health measure, decreased severity and fewer complications of chronic conditions leading to a decrease in HCC and RxHCC risk scores, and decreased mortality. See Table 7.1 for a summary of the outcome measures used in the health status analyses.

Regression Results

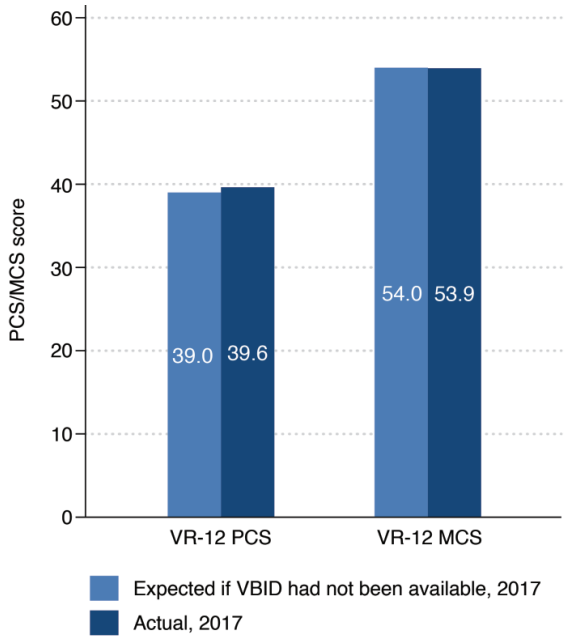
Our difference-in-differences analyses examined whether VBID caused improvements in health status or slowed declines in health status. The difference-in-differences findings for self-reported health measures are presented in Figures 7.3–7.5. For continuous outcome measures (i.e., all measures except smoking status, which is shown in Figure 7.4), the figures compare the observed values of the measures among VBID beneficiaries in 2017 (dark blue bar) with the expected values in the absence of VBID (light blue bar) in the same year. ***No statistically significant effects of VBID were found on any of the self-reported health status measures.*** Regression coefficients from these models are presented in Appendix H.

Figure 7.3 shows that among VBID beneficiaries, there was no statistically significant difference between the observed values of the Veterans RAND 12-Item Health Survey (VR-12) Physical Component Score (PCS) and Mental Component Score (MCS) in 2017 and the values expected in the absence of VBID.

For smoking status, Figure 7.4 shows a comparison of the predicted probability of being a current smoker among VBID beneficiaries in 2017, both in the presence (dark blue bar) and absence (light blue bar) of VBID.

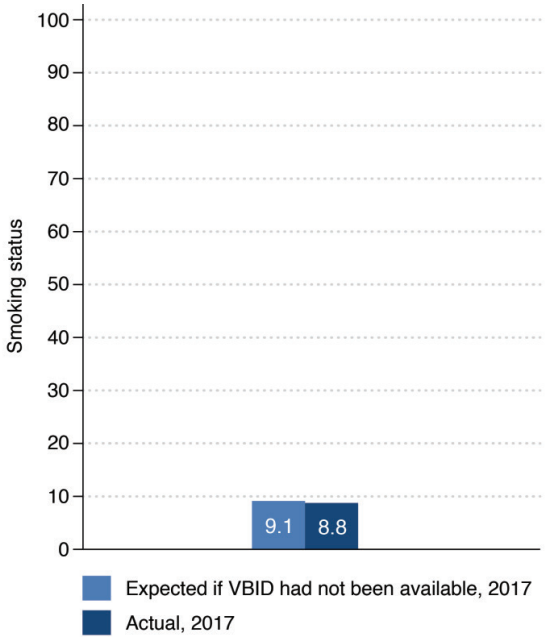
Figure 7.5 shows that, for the remaining self-reported health measures, there was no significant difference between the values observed among VBID beneficiaries in 2017 and the values expected in the absence of VBID. We note several limitations of this analysis, notably the

Figure 7.3. Comparison of Self-Reported Physical and Mental Health Scores (VR-12 PCS, MCS) Among VBID Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017



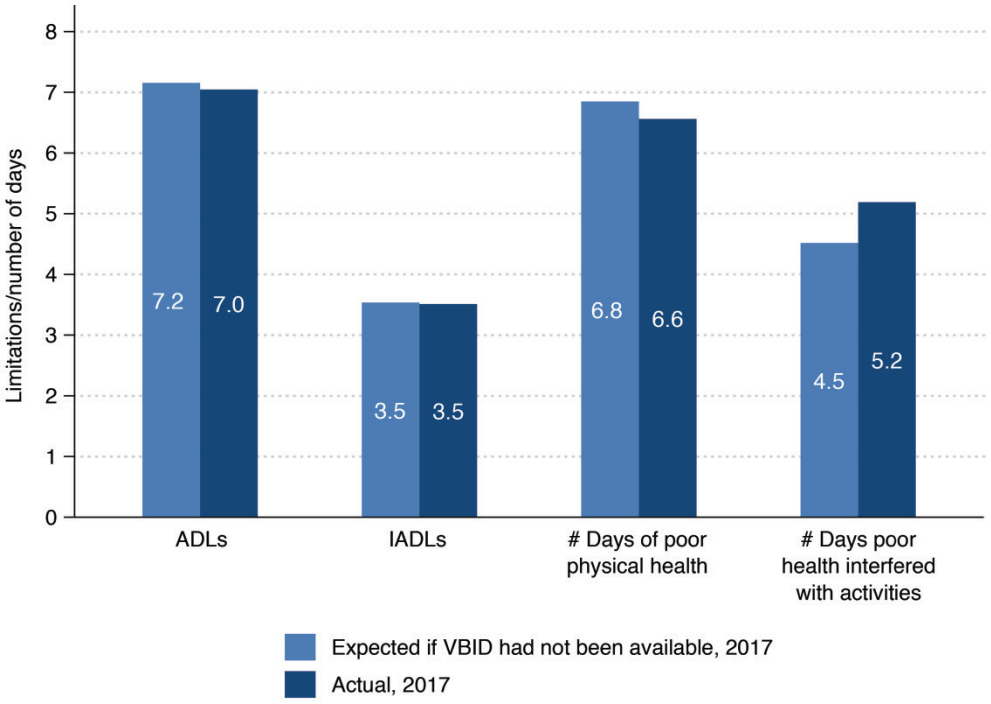
NOTE: Differences are not statistically significant at conventional levels. See Appendix H for details. Sample size for VR-12 PCS is 2,218. Sample size for VR-12 MCS is 2,206.

Figure 7.4. Comparison of Probability of Smoking Among VBID Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017



NOTE: Differences are not statistically significant at conventional levels. Results are derived from ordered logistic difference-in-differences models comparing VBID beneficiaries with matched comparison beneficiaries. Sample size is 2,032.

Figure 7.5. Comparison of Other Self-Reported Health Measures Among VBID Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017



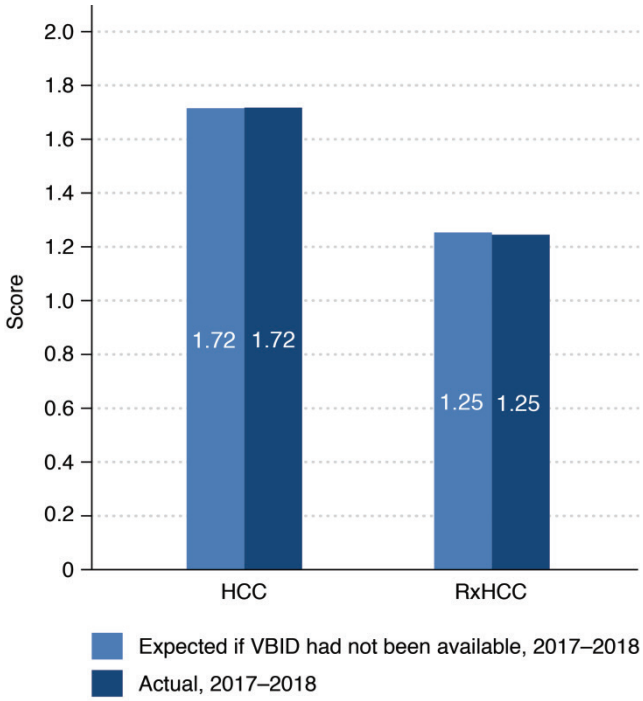
NOTE: Differences are not statistically significant at conventional levels. Results are drawn from the difference-in-differences models comparing VBID beneficiaries with matched comparison beneficiaries. Sample sizes range from 1,858 to 2,156.

relatively small sample sizes for each measure, and the limited follow-up period of only two years between the baseline and follow-up HOS surveys that might not be sufficient to detect changes in health status as a result of VBID. These limitations are reviewed at the end of this chapter and discussed in further detail in Appendix H.

The difference-in-differences findings for HCC and RxHCC scores are presented in Figure 7.6. Figure 7.6 shows that there was no significant difference between the values of HCC scores or RxHCC scores observed among VBID beneficiaries in 2017 and 2018 and the values expected in the absence of VBID.

Finally, ***we were unable to conclude whether VBID had any effect on the risk of death.*** As discussed in greater detail in Appendix H, the conditions needed for a valid survival analysis were not satisfied in our data, which limits our ability to draw firm conclusions about the effect of VBID on mortality.

Figure 7.6. Comparison of HCC and RxHCC Scores Among VBID Beneficiaries, Actual Versus Expected If VBID Had Not Been Available, 2017 and 2018



NOTE: Differences are not statistically significant at conventional levels. Results are drawn from the difference-in-differences models comparing VBID beneficiaries with matched comparison beneficiaries and represent the combined effects of VBID in 2017 and 2018. Sample sizes are 588,252 for both HCC and RxHCC scores.

Parent Organization Perspectives

The absence of statistically significant findings is not surprising given the results of our interviews with PO representatives. Some PO representatives told us that only in 2018 did they start using VBID as an opportunity to address weak spots in their Star Rating performance (referred to as "Star gaps") that arise when beneficiaries do not get recommended care: "We're focusing on these targeted classes of medications that are measured from the Stars perspective. So really having the pairing of the VBID team with the Stars team, doing this very focused reporting to watch that members don't fall out of compliance" (PO B).

Other POs, especially those staying in the VBID model test in 2020, noted that it might be too early to see major positive effects on quality and health outcomes. Instead of taking a short-term view, these POs are taking a long-term view of VBID and its impacts. For these POs, being able to engage more beneficiaries in their own care without seeing a *negative* short-term effect of the model test on care quality and health outcomes is actually a positive outcome.

> *"It's equally as important—almost more important—to say: 'Hey, [VBID] is not hurting anything.' If you can do the right thing, you can encourage the care, and you can equip care managers with what they feel is important to help [beneficiaries manage their conditions], and not negatively impact your Star Rating, I think that's also a valid outcome."*
>
> *PO G representative on the effect of VBID on care quality*

Moreover, from the PO perspective, it may be unrealistic to expect improvements in health outcomes of VBID participants because "of the fact that this is an older population, and as they age, their health deteriorates naturally," a PO E representative stated, Consequently, VBID's goal may be to maintain, rather than improve, beneficiaries' current health status. This sentiment was supported by a PO C representative who said that slowing down the disease progression may be the best health outcome VBID could achieve.

Finally, helping beneficiaries reach their personal health goals and helping them feel better might be another desirable VBID outcome.

> *"When we get member testimonials, even if it's one member, that say . . . I lost 11 pounds or 12 pounds . . . when you think about that, 11 pounds may have changed their life. May have kept them alive for a couple years if they're lowering their blood pressure, if they're just living a healthier life. And that alone to me is worth it. . . . At the end of the day it's about people actually feeling better or trying to get them to feel better."*
>
> *PO C representative on how VBID helped beneficiaries reach their personal goals*

Summary

Overall, our analyses identified no significant effects of VBID on health care quality and adherence for VBID-targeted medical conditions in the short term. Our analyses also identified no significant effects of VBID on general preventive care, indicating no evidence of positive spillovers to nontargeted conditions. There were also no significant negative effects of VBID on health care quality and adherence.

Our analyses also identified no significant effects of VBID on health status as measured by self-reported health and beneficiary risk scores. Again, there were no significant negative effects of VBID on health status. We could not conclude whether VBID had any significant effect on the risk of death.

Several limitations of our analyses, discussed in detail in Appendix H, may have contributed to our findings that VBID was without significant effects on health care quality/adherence or health status. Briefly, these include the limited follow-up period, the relatively small sample size for our contract-level quality and adherence measures, and the limited availability of quality and adherence measures with high relevance to the VBID interventions included in the model test.

8. Effects of Value-Based Insurance Design on Health Care Costs

The VBID model test was intended, among other things, to lower Medicare spending by reducing costly complications that can occur when chronic conditions are poorly managed. Multiple entities shoulder the costs of health care services and prescription drugs covered by Medicare: the federal government (Medicare); Medicare beneficiaries, through spending on premiums and cost-sharing; and POs, through spending on health care services and prescription drugs by PBPs. The effect of VBID on health care costs is uncertain. On the one hand, reducing costly complications can reduce spending, but on the other, lower cost-sharing and increased interaction with the health care system may increase utilization and costs.

In this chapter, we examine the effect of the MA VBID model on health care costs borne by PBPs, beneficiaries, and Medicare. We examine spending by PBPs on health care services and prescription drugs, and by beneficiaries on PBP premiums and prescription drugs. We also evaluate whether VBID affected the costs to Medicare of MA and Part D coverage. Finally, we examine PBP bids, which are the basis for Medicare costs for a given year and are a forward-looking indicator of PBPs' anticipated costs in the coming year. To assess these effects, we considered multiple quantitative measures combined with insights from interviews with PO representatives and beneficiaries to capture their perceptions of changes in spending due to VBID.

Key Takeaway Points:

- We did not find evidence of lower PBP spending on health care services or prescription drugs among VBID-participating PBPs in 2017 or 2018. Data for 2019 were not available.
- While overall MA-PD premiums did not significantly change in VBID-participating PBPs, monthly MA premiums in VBID-participating PBPs grew from baseline by an additional $11.36 in 2017, $21.43 in 2018, and $18.46 in 2019, relative to growth from baseline in the comparison PBPs.[1]
- Part D premiums were statistically significantly lower for VBID-participating PBPs in 2018 (−$9, or 8 percent) only. Part D premiums were lower in 2017 and 2019, but not by a statistically significant amount.

[1] The increases are not cumulative. Rather, they are difference-in-differences estimates that compare changes in premiums between VBID and comparison PBPs over time. These estimates capture the growth in premiums observed for VBID plans over and above changes observed over the same time period for comparison plans. For instance, the estimate for 2018 indicates that MA premiums for VBID plans were $21.43 higher in 2018 than would have been anticipated if premium growth trends for VBID PBPs mirrored trends for comparison PBPs. The approach controls for fixed differences in the levels of premiums between VBID and comparison PBPs, and captures only differential changes in trends over time.

- Part D OOP costs decreased by a statistically significant but small amount in 2017 (−$21, or 4 percent) and in 2018 (−$15, or 3 percent). Data for 2019 were not available.
- Part D costs to Medicare declined in 2017, but this finding appears to be driven by unrelated changes in the risk scores of enrolled beneficiaries rather than savings due to VBID. We found no statistically significant change in these costs for 2018.
- There was no change in MA costs to Medicare in the first two model years (2017 and 2018).
- No significant changes in average MA or MA-PD combined bids were detected for any of the first three years of the model test.
- VBID reduced average bids for Part D coverage in 2018 and 2019, by $10 (14 percent) and $12 (18 percent), respectively.

Health Care Services and Prescription Drug Spending

To examine the effect of VBID on health care costs, we first estimated the effect of VBID on realized spending on health care services and prescription drugs by PBPs and spending on premiums and prescription drugs by beneficiaries. These measures focus on spending on actual health care services, prescription drugs, and beneficiaries' costs of coverage, but they do not address administrative costs. All analyses reported in this chapter focus on PBPs offering both MA and Part D benefits (MA-PD PBPs); these PBPs represent the majority of PBPs participating in VBID. Additional estimates for MA-only PBPs are presented in Appendix I.

Plan Benefit Package Spending

PBP Spending

Spending by the PBP on medical services and prescription drugs on behalf of PBP enrollees.

We measure health care services spending by PBPs using data on per-beneficiary per-month net medical costs reported by POs as part of their bids. Retrospective net costs for health care services are reported as part of supporting data for the bids submitted for MA coverage to be offered in the following year. Bids are submitted on or before the first Monday of June in the year before the coverage will begin, with spending from the preceding year reported as experience data. For example, realized health care spending in 2017 was submitted with bids for coverage for 2019. The 2019 bids were submitted in June 2018. Data on realized medical spending were available for the years 2014–2017.

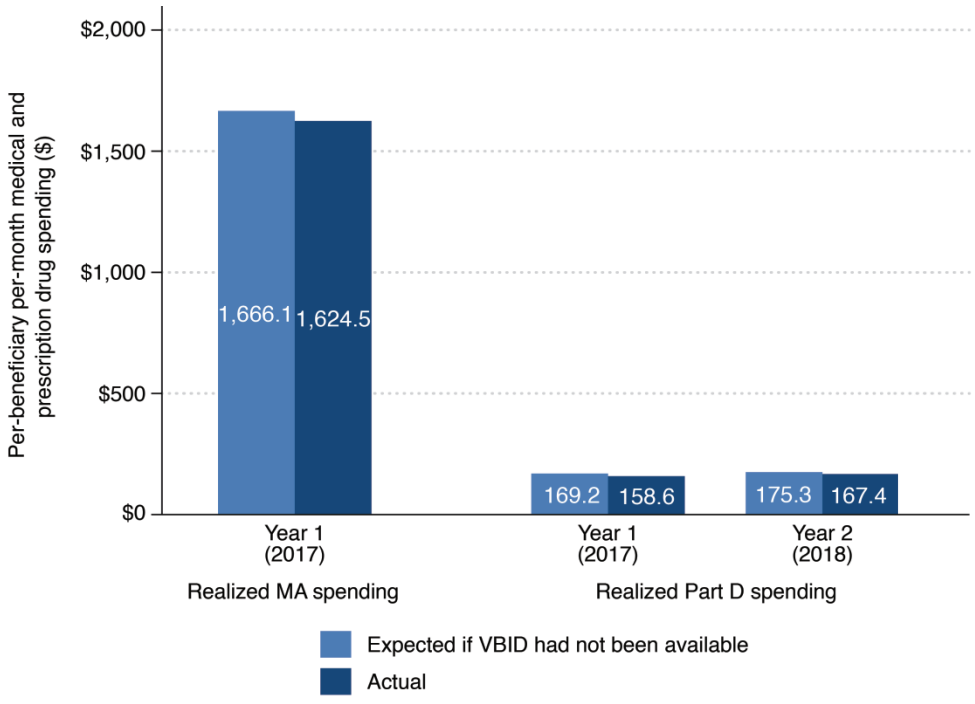
Prescription drug spending by PBPs was measured as total PBP spending for all final prescription drug events for that PBP on the Integrated Data Repository (IDR). Per-beneficiary per-month prescription drug spending in each year was calculated by dividing total PBP spending by the number of beneficiary months of coverage provided by the PBP. We note that our measure of prescription drug spending does not account for direct and indirect remuneration received by PBPs, and so our spending measure reflects gross (point-of-sale) drug costs rather

than net drug costs. Data on realized PBP prescription drug spending were available for the years 2014–2018.

To estimate the effect of VBID on medical and prescription drug spending by PBPs, we followed the general approach to study design described in Chapter 1. We estimated a difference-in-differences regression model comparing changes in outcomes over time between the VBID PBPs and the matched comparison PBPs. For each model year with data available (2017 for medical spending; 2017 and 2018 for drug spending), we calculated the expected level of per-beneficiary per-month spending predicted by our regression model for the VBID-participating PBPs. We compare actual spending by VBID PBPs to the regression-based prediction of what spending would have been in the absence of the VBID model test. Technical details and regression tables are presented in Appendix I.

Figure 8.1 compares per-beneficiary per-month health care services (MA) and prescription drug (Part D) spending by VBID-participating PBPs with the expected level of spending if VBID had not been available, for the model years noted above. ***We did not detect any statistically significant changes in realized MA spending in 2017, nor did we detect any statistically significant changes in realized Part D spending in 2017 or 2018.*** Point estimates for the change in spending associated with VBID are negative, but quite small.

Figure 8.1. Comparison of Per-Beneficiary Per-Month Spending for VBID-Participating PBPs, Actual Versus Expected If VBID Had Not Been Available, 2017 and 2018



NOTE: None of the differences is statistically significant. Counterfactuals estimated using the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. Sample size is 257 PBP-year observations for MA spending and 322 PBP-year observations for Part D spending. Each model contains data from 33 VBID PBPs and 32 comparison PBPs. See Appendix I for details.

Beneficiary Costs

We analyzed four different beneficiary OOP cost variables: MA premiums, Part D premiums, MA-PD premiums (the sum of MA and Part D premiums), and OOP costs for prescription drugs (Part D). We obtained monthly PBP-level premium data for MA and Part D coverage for the years 2014–2019 and constructed annual beneficiary Part D OOP costs for the years 2014–2018. Part D OOP costs were calculated as the sum of patient payment amounts from the PDE data submitted by PBPs. Although MA OOP costs, such as copayments and coinsurance for office visits, are an important component of beneficiaries' total costs, we were unable to examine this outcome due to data limitations.

Beneficiary Costs

Premiums: Monthly amount paid by beneficiaries enrolled in PBPs for MA and Part D coverage.

Part D OOP costs: Annual beneficiary cost-sharing, such as copayments and coinsurance, for covered Part D drugs.

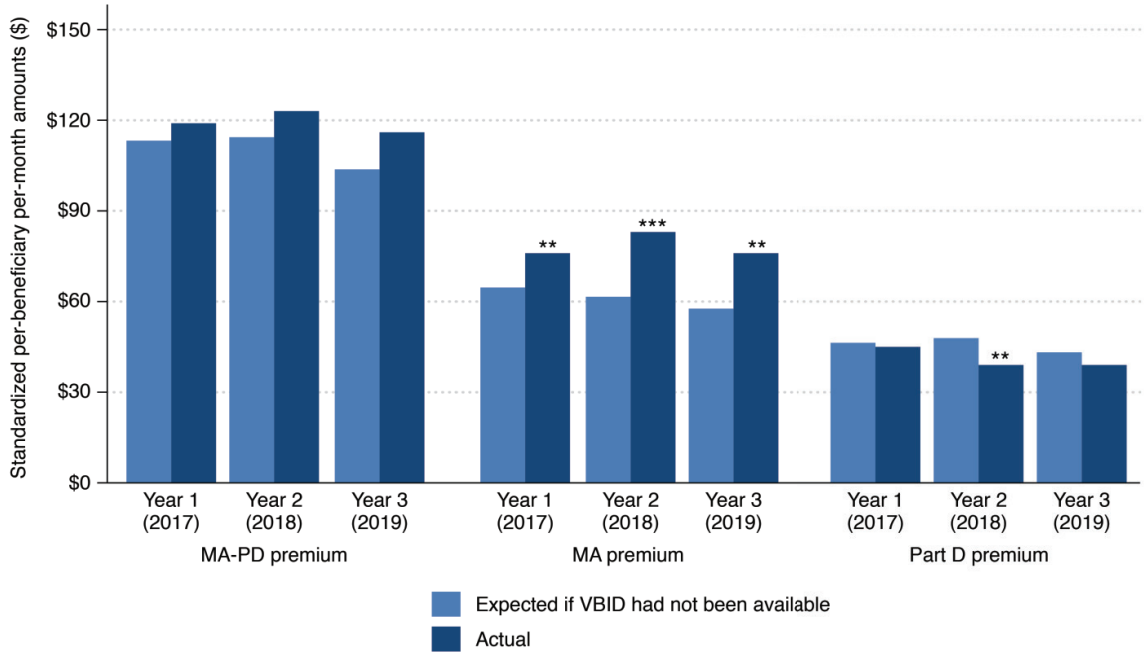
To estimate the effect of VBID on premiums, we used a difference-in-differences approach similar to that used for the PBP spending analyses described above. We estimated a difference-in-differences regression model comparing changes in outcomes over time between the VBID PBPs and the matched comparison PBPs. For each model year with data available (2017–2019), we calculated the expected level of monthly MA and Part D premiums predicted by our regression model for the VBID-participating PBPs. We compared VBID PBP premiums to the regression-based prediction of what premiums would have been in the absence of the VBID model test. Technical details and regression tables are presented in Appendix I.

For Part D OOP costs, we used an approach similar to the utilization analyses described in Chapter 6. That is, we compared annual Part D OOP costs for beneficiaries eligible for VBID and enrolled in VBID-participating PBPs to costs for matched comparison beneficiaries in nonparticipating PBPs. We were able to focus specifically on VBID-eligible beneficiaries because our measure of Part D OOP costs was observed at the individual beneficiary level. Other outcomes analyzed in this chapter relied on data reported at the PBP-year level, meaning that our analyses of PBP spending, premiums, costs to Medicare, and bids reflect changes in PBP-average outcomes associated with VBID rather than changes in costs specifically among VBID-eligible beneficiaries. These distinctions in the level of observation and the populations examined should be noted when interpreting our results.

Premiums

Figure 8.2 shows the estimated association between VBID participation and premiums. We observed that MA premiums for VBID-participating PBPs increased more than MA premiums for a set of comparison PBPs after the model test began. Specifically, in the aggregate, monthly premiums in VBID-participating PBPs grew from baseline by an additional $11.36 in 2017,

Figure 8.2. Comparison of PBP Premiums, Actual Versus Expected If VBID Had Not Been Available, 2017–2019



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Counterfactuals estimated using the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. Sample size is 386 PBP-year observations. Each model contains data from 33 VBID PBPs and 32 comparison PBPs. See Appendix I for details.

$21.43 in 2018, and $18.46 in 2019, relative to growth from baseline in the comparison PBPs.[2] Findings were similar in sensitivity tests in which we excluded VBID PBPs that were outliers and reweighted the data to reflect total PBP enrollment.

While our mixed-methods evaluation was intended to isolate the effect of VBID, this study is ultimately observational in nature. In the absence of a randomized controlled trial, it is difficult to establish whether these findings can be attributed to the VBID model alone, or in part. Importantly, our matching strategy did not include all possible key differences between VBID participants and matched comparators, including differences in benefit design beyond the out-of-pocket maximum and competitive and pricing dynamics in individual PBP markets (see Appendix D). While the difference-in-differences approach is designed to reduce biases introduced by imperfect matching, that approach relies on the assumption that treatment and comparison PBPs would have experienced similar premium changes in the absence of VBID.

[2] Prior to the model test, average MA premiums for VBID-participating PBPs were higher than average MA premiums nationwide, an issue that was only partly addressed via matching. For example, in 2016, MA premiums for MA-PD PBPs that participated in the model test were $67 PMPM on average, compared to $49 PMPM among comparison PBPs and $26 PMPM among all MA-PD PBPs that would have met VBID eligibility criteria if the model were offered in their state. VBID-participating PBPs and their matched comparators were also less likely to offer zero premium PBPs than the general population of PBPs. In 2016, 24 percent of VBID PBPs and 18 percent of comparators had zero premiums, compared to 49 percent of all PBPs that would have met VBID criteria if the model were available to them.

To better understand the premium increases, we analyzed selected factors that may have contributed to these findings. Our interviews with POs did not directly address possible pricing changes due to VBID, but respondents did note several VBID-related factors that could have influenced PBPs' administrative costs and may have subsequently led to premium increases. Examples include IT improvements to support VBID, additional staff to administer CM/DM programs, and communications with beneficiaries to explain benefit changes. Our analysis of PBPs' bid data also found that projected administrative costs for VBID-participating PBPs grew more, relative to administrative costs for comparison PBPs, after the model took effect. Relative increases in administrative costs allocated to Medicare-covered services ranged from $9.52 to $18.36 per beneficiary per month, depending on the year. Administrative costs allocated to supplemental benefits also increased in 2017 and 2018 for VBID-participating PBPs relative to comparators, although these relative increases were small ($0.77 to $2.06 per-beneficiary per-month). See Appendix I for details on these analyses.

These results suggest that growth in administrative costs for VBID-participating PBPs may have been a factor contributing to premium increases. However, we do not have a direct measure of the cost of implementing VBID. To the extent that VBID increased administrative costs, it is possible that some of these costs may have been frontloaded, so that costs borne early in the model test would yield a return on the investment over time. Some of these investments could also serve the PBPs' enrolled population beyond those eligible for VBID.

Increases in the cost of supplemental benefits for VBID-participating PBPs may also have influenced MA premiums because PBPs that provided supplemental benefits to VBID-eligible beneficiaries as part of their interventions were required to fund these benefits through premiums and/or rebate dollars from all enrollees (CMS, 2015). We found a statistically significant change in the direct cost of supplemental benefits in 2018: a $7.35 increase among VBID-participating PBPs, relative to matched comparators. However, the changes in supplemental benefits costs were small, relative to the changes in administrative costs, and the change was statistically significant only in 2018 (see Appendix I).

Other factors that were not related to the VBID model may have also contributed to premium increases in VBID-participating PBPs. Our interviews with POs revealed that at least one PBP expanded its general (non-VBID) care management program around the time that VBID took effect, and that two PBPs eliminated cost-sharing for primary care visits for all beneficiaries. Other VBID-participating PBPs may have made similar changes that they did not report during the course of the evaluation. If such changes in benefit design increased costs and did not occur in comparison PBPs, the estimated increase in MA premiums among VBID-participating PBPs may reflect these general, non-VBID related effects.

Given these factors, it is not possible to conclude with certainty the degree to which observed premium increases are associated with participation in the VBID model; it is likely that a number of factors may have contributed to this increase in premiums.

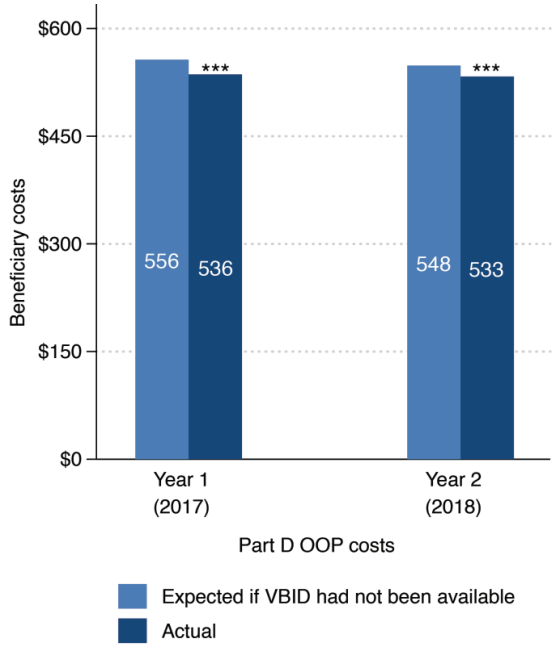
We found a significant decrease in Part D premiums (–$9, or 8 percent), but only in 2018. It is worth noting that beneficiaries enrolling in MA-PD PBPs pay the combined MA and Part D

premiums on a monthly basis. Though the premium analyses found statistically significant changes in MA and Part D premiums due to VBID, there was no statistically significant effect of VBID on the combined MA-PD premiums, as shown in Figure 8.2.

Part D OOP Costs

Figure 8.3 shows the results of the difference-in-differences model for Part D OOP costs, comparing actual versus expected beneficiary Part D OOP costs if VBID had not been available in 2017 and 2018. ***We found a small but statistically significant decrease in Part D OOP costs during 2017 and 2018 (−$21 and −$15, respectively).***

Figure 8.3. Comparison of Beneficiary Part D OOP Costs, Actual Versus Expected If VBID Had Not Been Available, 2017–2018



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Counterfactuals estimated using the difference-in-differences models comparing VBID-participating beneficiaries to matched comparison beneficiaries. Sample size is 58,463 in 2017 and 42,220 in 2018. Beneficiary costs are annual. See Appendix I for details.

Beneficiary Perspectives on Costs

Although we were not able to use CMS data to analyze beneficiaries' costs associated with MA OOP, we asked beneficiaries about this topic as part of our interviews. Specifically, we asked them to identify what they liked and did not like about their current health insurance plan. Beneficiary OOP costs (which encompass both MA and Part D costs) ranked highly on both lists. Of 100 beneficiaries, 21 indicated that they felt their PBP's costs were reasonable, while 33 indicated they did not like the high OOP costs in their PBP. One beneficiary from PO G noted that his or her PO was going to raise specialist visit copays from $20 to $25, which could make health care unaffordable for those with chronic conditions: "I don't think that's right because if

you have a heart condition, you know what I mean? You have to see them every so often." Another PO G beneficiary agreed that cardiology copayments were too high, which made it difficult for him or her to seek needed care. Similarly, some beneficiaries were concerned about their medication costs: "My wife's got diabetes, and she's on insulin. It's very expensive, and I wish they'd pay more on that. Once you hit that donut hole, all of her medication goes way up."

Costs of Coverage

To complement our estimates of the effects of VBID on health care services and prescription drug spending by PBPs and beneficiaries, we examined the effect of VBID on per-beneficiary per-month costs to Medicare as well as the effect on prospective bids submitted by PBPs. In the context of MA and Part D, costs to CMS reflect payments from Medicare to POs to cover the federal government's share of the costs of MA and Part D coverage, with adjustments to payments made to account for any risk sharing between the government and the PBPs (e.g., for risk corridors in Part D). Further details on our cost measures are presented below and in Appendix I. PBP bids influence but do not completely determine costs to Medicare, so the effects of VBID on PBP bids likely differ from the effects of VBID on costs to Medicare. PBP bids reflect the PO's best estimate of the cost of coverage for the coming year, including both expected costs for health care services (or prescription drugs) and administrative costs. VBID may lower bids over time if POs project lower costs associated with the model in comparison to what costs would have been in the absence of VBID. The bid data are standardized across PBPs, enabling us to assess the effect of VBID across participating PBPs and matched comparison PBPs.

Costs to Medicare

Costs to Medicare

Costs paid by Medicare for health and drug coverage. These include the portion of the MA and Part D PBP bids paid by Medicare, adjusted by the risk score for each enrollee in the PBP; MA PBP rebates paid; and additional Medicare costs associated with Part D reinsurance and low-income subsidies.

We measured costs to Medicare by calculating the per-beneficiary per-month payments made by Medicare to POs for PBPs offering benefits under MA and Part D benefits. We used PBP bids, final risk score data, and per-beneficiary MA rebates (also from the bid) to calculate MA costs to Medicare. To estimate Part D costs to Medicare, we used publicly available information on the components of Part D payments for each year, Part D risk score data, premium data to estimate the low-income premium subsidy, and actual low-income cost-sharing payment data from the PDE data. Additional details on these sources and the formulas used in our calculations are presented in Appendix I. We note that our measure of Part D costs to Medicare did not include risk corridor payments because PBP-level data on risk corridor payments were not readily available for this study. We think it is unlikely that the risk

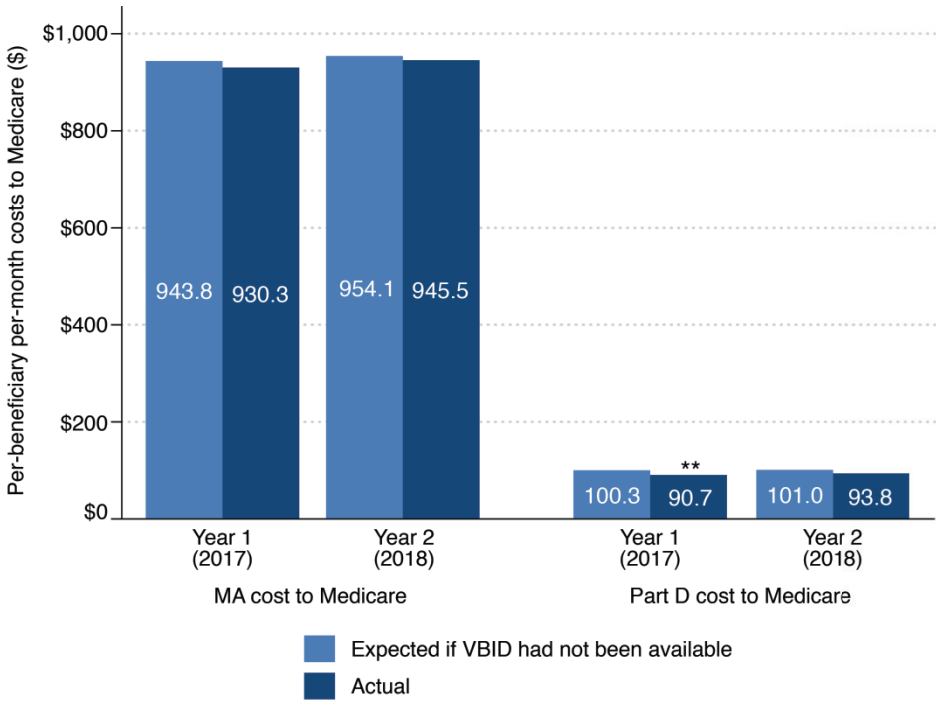
corridor payment data would change the findings of our difference-in-difference models, but this potential limitation of our measure of costs to Medicare should be considered when interpreting our findings. Additional background on the risk corridors is presented in Appendix I.

Data on MA and Part D costs to Medicare were available for the 2014–2018 contract years, allowing us to examine changes in costs associated with VBID over the first two years of the VBID model test (2017 and 2018).

For each of the two model years with data available (2017 and 2018), we calculated the expected costs to Medicare predicted by our regression model for the VBID PBPs. Figure 8.4 compares actual costs to Medicare for VBID PBPs with expected costs had VBID not been available, in each model year. ***There were no statistically significant changes in per-beneficiary per-month MA costs to Medicare in the first two model years.***

Per-beneficiary per-month Part D costs to Medicare in 2017 were statistically significantly lower for VBID PBPs compared to what would have been expected in the absence of VBID. We estimated that VBID was associated with a $9.63 reduction in Part D costs (9.6 percent of what we would have expected in the absence of VBID) in 2017. ***However, we did not find any statistically significant change in Part D costs to Medicare in 2018.*** The estimated change in

Figure 8.4. Comparison of Per-Beneficiary Per-Month Costs to Medicare for VBID-Participating PBPs, Actual Versus Expected If VBID Had Not Been Available, 2017 and 2018



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. Sample size is 322 PBP-year observations from 33 VBID PBPs and 32 comparison PBPs. See Appendix I for details.

Part D costs for 2018, a statistically insignificant reduction of $7.21 in Part D costs, was slightly smaller than the estimated change in 2017.

This pattern of results may seem surprising: If VBID were, in fact, responsible for lowering Part D costs, we would expect reductions in Part D costs to grow over time as beneficiary health improves, rather than fading away after the first year of implementation. The fact that the apparent reduction in Part D costs estimated in 2017 did not persist into 2018 raises the question of whether the estimated change in Part D costs truly reflects the effect of VBID. Because PBP bids play a central role in determining Part D costs to Medicare, we revisit the interpretation of these findings below, after presenting our results for PBP bids.

Plan Benefit Package Bids

PBP Bids

POs' projected costs of health and prescription drug coverage for coming year, submitted about seven months before start of plan year.

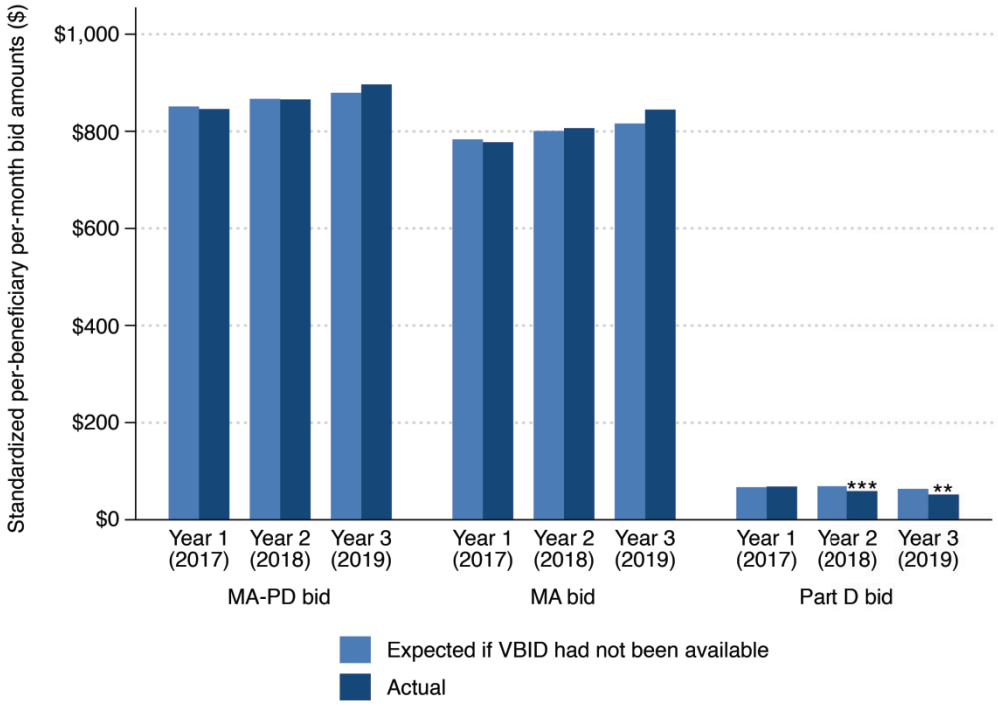
We analyzed PBP bids for 2014–2019, which included three years of data after the start of the model test. Data on MA bids for VBID and comparison PBPs were extracted by the CMS Office of the Actuary (OACT) from bid pricing tool (BPT) spreadsheets submitted for MA coverage in 2014–2019. Bids represent the projected costs of providing Medicare coverage to beneficiaries for the CY. The bids are submitted as monthly per-beneficiary cost estimates and are standardized to reflect a 1.0 beneficiary risk score. This facilitates comparison of the bids across different PBPs. POs submit bids for both MA (Parts A and B) benefits and Part D benefits separately.

Because bids are submitted prospectively (on the first Monday in June) for PBPs to be offered in the following year, changes in PBP bids associated with VBID for a given contract year may be reasonably interpreted as a reflection of POs' expectations of PBP costs based on the first half of the year preceding the contract year. The prospective nature of bids enabled us to examine outcomes for the third model year (2019), but we note that these bids reflected PO experiences only through mid-2018. Similarly, we note that bids for the first model year (2017) would have been submitted well before the PBPs had any experience with VBID implementation and therefore were submitted under greater uncertainty regarding anticipated savings.

Figure 8.5 compares actual bids for VBID-participating PBPs to the bids expected had VBID not been available in each model year. For each of the three years after the implementation of VBID (2017, 2018, and 2019), we present the expected bid predicted by our regression model for the VBID-participating PBPs.

Part D bids for coverage in 2018 and 2019 were lower than would have been anticipated in the absence of VBID, by $9.93 per-beneficiary per-month in 2018 and by $11.69 in 2019. These reductions are sizable compared to the average bids actually submitted by VBID plans, amounting to 17 percent of the average Part D bid of $59.03 in 2018 and 23 percent of the average Part D bid of $51.89 in 2019. However, ***there were no statistically significant changes***

Figure 8.5. Comparison of Bids for VBID-Participating PBPs, Actual Versus Expected If VBID Had Not Been Available, 2017–2019



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs with matched comparison PBPs. Sample size is 387 PBP-year observations from 33 VBID PBPs and 32 comparison PBPs. See Appendix I for details.

in MA bids or in MA-PD bids (which combine MA and Part D bids) in the first three model years (2017–2019). Because MA bids are an order of magnitude larger than Part D bids, a relatively small and statistically insignificant increase in MA bids ($29 compared to an average bid of $845, or a 3-percent increase) offsets the statistically significant and relatively large decline in Part D bids.

The absence of statistically significant changes in MA-PD bids over the first three years of the model test is somewhat surprising: Participating POs were required not only to provide financial projections indicating net savings to Medicare over the first five years of the model, but also to incorporate the assumptions underlying these projected savings into their PBP bids. Costs to Medicare are affected by many factors besides bids, so it is possible that costs to Medicare could have changed without systematic changes in PBP bids. However, the data presented in Figure 8.5 do not suggest that substantial changes in costs to Medicare would be expected.

The absence of statistically significant changes in Part D bids also raised questions about the interpretation of our finding that Part D costs were lower in 2017 than we would have expected in the absence of VBID. To understand this pattern of results, we estimated the difference-in-differences regression model separately for each factor that enters the formula for Part D costs to

Medicare. Results and further discussion are presented in Appendix I. VBID-associated changes in 2017 were small and insignificant for most factors affecting payment to PBPs—as they were for PBP bids. Instead, we found that ***reduced Part D costs to Medicare in 2017 reflect, at least in part, a compositional shift in PBP enrollment toward beneficiaries with lower Part D risk scores and a marginally significant (p = 0.071) reduction in Medicare reinsurance payments.*** Given the modest size of the estimated 2017 effect on Part D costs and the smaller and insignificant effects estimated for 2018, we do not view these results as providing strong or robust evidence of Part D savings associated with VBID.

Parent Organization Perspectives on Plan Benefit Package Bids and Costs to Medicare

POs did not generally comment, during interviews, on whether they had experienced savings for the VBID population, and several noted that savings were not a primary outcome they were trying to achieve: "I've done projects and signed off on them when there isn't an ROI [return on investment] because I understand it's for the betterment of the member. It's really more our mission," said a PO C representative. Some POs commented on changes in costs associated with VBID but were not yet ready to confirm the trends. Of these, POs A and C indicated that costs were lower for the VBID group, but POs B and I indicated that they thought costs were increasing, at least initially, owing to increased utilization of high-value services. PO I noted that the cost of the monitoring equipment and low beneficiary enrollment in the intervention had a negative effect on ROI: "It's a very high cost per member because we have not been able to reach the break-even point to pay for . . . the cost of renting the equipment, having 24/7 staffing to do the monitoring on such a small patient population, it is a little bit cost prohibitive." Of the four POs that discussed costs, two reported seeing positive results, whereas the other two did not. It is worth noting that the analytic approach many POs used to evaluate cost trends within their own organizations differed from the approach used in this report. POs cannot access data on cost trends for competitors, so they would have been unable to use comparison groups to disentangle the effects of VBID from other influences on costs. This fundamental difference in methodologies may explain any differences between our results and the POs' own estimates.

Summary

Changes in realized medical and drug spending by PBPs after the start of the model test were small and statistically insignificant. ***MA premiums grew more from baseline in 2017, 2018, and 2019 among VBID-participating PBPs than among non-VBID participating PBPs, while Part D premiums decreased by a statistically significant amount in 2018.*** When combined, there was no statistically significant increase in total (MA and Part D) premiums for POs that offered both benefits. ***Beneficiaries' Part D OOP costs fell by a statistically significant magnitude in 2017 and 2018.*** For Part D, our findings that 30-day fills per beneficiary increased on average,

combined with lower estimated Part D OOP costs, suggests that beneficiaries eligible for VBID may have shifted utilization to lower-cost drugs.

Changes in MA costs to Medicare associated with VBID were statistically insignificant. VBID was associated with a statistically significant reduction in Part D costs to Medicare in 2017. However, in comparison to actual costs in 2018 that would have been expected had VBID not been available, changes in Part D costs were not statistically significant. The 2017 results appear to reflect lower payments by Medicare for drug costs above the catastrophic threshold and, perhaps, a shift in enrollment toward beneficiaries with slightly lower Part D risk scores.

Neither MA-PD bids nor MA bids exhibited statistically significant changes in any of the first three years of the model test. In contrast, ***Part D bids were lower by a statistically significant amount in 2018 and 2019;*** however, any reduction in Part D bids among these PBPs appears to have been offset by (insignificantly) higher MA bids. The lack of observed significant reductions in MA bids is somewhat unexpected, given that POs proposing to participate in VBID were required to show actuarial certifications of anticipated savings during the five years of the model test.

9. Conclusions

Between 2017 and 2019, 11 POs participated in CMS' MA VBID model test, the first of its kind within the Medicare population. Many of the POs that participated in the model test's inaugural year in 2017 continued their participation during 2018 and 2019, and two new POs began offering VBID interventions in 2018 and 2019. Nonetheless, only four 2019 model participants continued their participation in 2020.

For this report, we evaluated the initial effects of the MA VBID model test using one to three years of primary and secondary data analyzed quantitatively or qualitatively. This mixed-methods approach allowed us to examine the model test's effects from the vantage point of POs, beneficiaries, and, in some cases, providers, while also analyzing the early effects of MA VBID model on health care utilization, enrollment, patient experiences, quality, health outcomes, and costs. These findings led us to several conclusions about the effects of the VBID model test.

Value-based insurance design appears to be working mostly as intended

VBID aims to motivate beneficiaries to use high-value care, which may ultimately improve health care quality, avoid costly complications, reduce spending, and improve beneficiary health. As a necessary first step toward achieving VBID's long-term goals, beneficiaries must increase their use of high-value services. In the context of the MA VBID model test, these increases appear to be occurring. We found statistically significant increases in use of more than half of the specific services targeted by POs for reduced cost-sharing (10 of 18 VBID-targeted services), such as visits with high-value providers, completion of diabetes-monitoring "scorecards," and use of targeted COPD treatments. VBID-eligible beneficiaries reported that they appreciated lower copays (although they also cited other barriers to receiving care, such as transportation), and our spending analyses showed that eligible beneficiaries experienced statistically significant declines in OOP costs for Part D prescription drugs. These lower Part D costs suggest that reduced cost-sharing for VBID targeted drugs had a financial effect for beneficiaries, or that medication reconciliation was taking place during PCP visits or as part of required CM/DM activities.

Across all eligible beneficiaries in VBID-participating POs, we found increases in primary care visits, specialist visits, prescription drug fills, and use of DME. These across-the-board increases may have occurred because participating POs were effective in encouraging eligible beneficiaries to interact with care managers and primary care providers, who would have enabled beneficiaries to use a range of high-value services relevant to their needs. If expectations about VBID are correct, these increases in utilization could eventually lead to improved health and lower Medicare spending, among other benefits.

There were a few unintended results

Although most of our results suggest that VBID is having the intended effect, a few findings were not consistent with the goals of VBID. Specifically, we did not expect to see increases in ACS inpatient and ED utilization, and in fact VBID aims to reduce unnecessary or avoidable inpatient or ED utilization through better management of chronic conditions and prioritization of primary care. It is possible that VBID prompted people to use more care, which may have led to increased referrals to emergency and inpatient settings in the short run. To the extent that they are avoidable, these types of encounters may decline over time, particularly if VBID enables beneficiaries to develop stronger relationships with care managers and primary care providers, or if VBID reduces avoidable health complications.

In addition, we found increases in beneficiaries' MA premiums in all three years of the model test compared to relative growth from baseline in premiums in comparison PBPs, although beneficiaries' Part D premiums and OOP costs for Part D decreased significantly in some years. One of VBID's goals is to decrease overall beneficiary spending by reducing the need for costly care. The increases in MA premiums raise questions as to whether this reduction in beneficiary cost is occurring. However, we were unable to assess all components of beneficiary costs; notably, we did not have reliable data on MA OOP spending (e.g., copayments and coinsurance).

Finally, not all eligible beneficiaries participated in the model test. Most participating POs required beneficiaries to complete participation requirements to receive VBID benefits, and less than one-third of beneficiaries in POs with such requirements participated in the model test. Poor communication between POs and beneficiaries, as well as beneficiaries' confusion over the VBID model, could also have dampened effects on utilization and other outcomes.

More time is needed to gauge effects on downstream outcomes

For the most part, available data allowed us to analyze outcomes only through the end of 2017, which was the first year of VBID implementation. At the end of 2017, POs were still fine-tuning their interventions and overcoming initial implementation challenges (detailed in our first annual report, Eibner et al., 2018). For example, although eligible beneficiaries were not uniformly aware of VBID in the initial years, POs have continued to refine strategies to engage beneficiaries and encourage participation.

Most of our statistically significant findings (18 of 24) are for utilization outcomes that might be relatively responsive to changes in cost-sharing and would be expected to precede other changes such as improvements in health care quality, adherence, and beneficiary health. For example, although prescription drug use may increase quickly in response to reduced copays, becoming adherent to a drug regimen requires complying with treatment recommendations over time. Even after becoming adherent, it may take months or years for beneficiaries to experience measurable health benefits of high-value treatment.

Further, most of the VBID approaches implemented by POs diverged from those previously tested in the literature, which could influence the time horizon needed to achieve desired effects. Seven POs adopted CM/DM approaches, which could require a relatively long runway to yield an effect, particularly if it takes time for beneficiaries to enroll into VBID and for care managers to establish a rapport with beneficiaries. Two POs provided benefits in the form of rebates, which break the link between receiving high-value care and getting an immediate discount at the point of service. One PO (PO B) reduced cost-sharing for high-value providers, which, for some beneficiaries, meant they had to switch providers to receive VBID benefits. Compared to simpler interventions that reduce cost-sharing for specific services, these VBID interventions may require more time to achieve impact.

Changes to the model test and other factors affected POs' participation

Although VBID appears to be moving outcomes in the intended direction, only 4 of 11 POs continued to participate in the model test in 2020. POs described a variety of reasons for leaving the model test, including perceived lack of ROI and the ability to implement VBID outside the model test via the increased flexibility offered under CMS' reinterpretation of the uniformity requirement. Other factors affecting POs' participation decisions included the perceived burden of participating in the model test (e.g., due to data submission and monitoring requirements) and the requirement to implement WHP as part of VBID starting in 2020.

Flexibility in benefit design limits ability to compare across interventions

CMS offered POs substantial flexibility to design VBID benefits, possibly more flexibility than could have been fully used by the relatively small number of POs that chose to participate in the model test. Although VBID was available in 25 states by the end of the third year of the model test, VBID was only implemented in six states. Similarly, by 2019, CMS had approved VBID for 11 health conditions and offered POs the flexibility to propose VBID benefits for any condition, yet POs targeted only five conditions. None of the participating POs offered VBID for mental health conditions, rheumatoid arthritis, or dementia.

Despite this apparent concentration of VBID in certain geographic areas and targeted conditions, VBID as implemented across the 11 participating POs was far from homogeneous. POs differed in which conditions they targeted, whether they focused on beneficiaries with or without comorbidities, how they communicated with beneficiaries, what beneficiaries had to do to receive benefits, and other factors. Many of these decisions could influence the outcomes that we analyzed, such as utilization and spending.

Because each PO's intervention was unique, it is difficult to draw conclusions about which interventions were most effective and what specific design features led to the changes we observed. For example, because PO D was the only participant to target beneficiaries with hypertension and one of only three POs to focus on drugs, any comparison of PO D's results to

those of other POs will be hampered by fundamental differences in the population targeted and the outcomes of interest. Our analysis controlled for this implementation variability as best as possible, but our concluding observations are necessarily broad.

Thus, although our analysis speaks to the overall effects of the MA VBID model test given the range of interventions adopted by participants, the results may not generalize to other VBID interventions, and it is difficult to conclude which components of POs' VBID designs had the biggest impact. These limitations are important considerations as CMS embarks on changes to the VBID model in 2020, which allows new flexibilities and will involve a different set of participating POs. VBID as implemented by the 2017–2019 participants appears to have moved utilization in the intended direction, but there is much unexplored territory going forward, and POs will continue to refine their benefit designs.

Looking Ahead

Significant changes to the VBID model test began in 2020 and will bring significant changes for the four POs that will continue their participation, as well as any new POs that join this model test. Most POs participating in the model test during the first three years expressed that they learned valuable lessons on implementing VBID and appreciated the opportunity to experiment with benefit designs. These lessons will serve them well as the MA environment in which they operate continues to evolve.

Appendix A. Methods for Year 2 and Year 3 Interviews with Parent Organizations

This appendix describes the methods used to conduct interviews with new and returning POs in Years 2 and 3 and to analyze the PO interview data. Analyses of the data are presented in Chapters 2–4.

Interviews with New Parent Organizations

One new participating PO joined the VBID model test in 2018, and one new PO joined in 2019. We conducted semistructured in-person interviews with representatives from these POs in July 2018 and August 2019, respectively. Interviews explored how POs chose their intervention designs, why they joined the VBID model test, and what their early implementation experiences were. For the new model participants, we followed the interview guide used for 2017 participants.

Interviews with Returning Parent Organizations

We conducted a series of semistructured interviews with returning VBID participants to better understand their VBID intervention designs, beneficiary participation conditions, approaches used to communicate with beneficiaries and providers, implementation experiences, intervention uptake, and expected intervention outcomes.

For Year 2 (2018) participating POs, we conducted in-person or telephone interviews with 90 PO representatives across all ten model participants. We allowed each PO to select representatives who were the most knowledgeable about VBID implementation. Interviewed representatives held a variety of positions, including Medicare product specialists, Medicare compliance officers, actuarial directors, directors of regulatory affairs, care management directors and staff, informatics specialists, and medical directors of government programs.

Between June and September 2018, two researchers conducted individual or small-group interviews with VBID-participating POs that lasted approximately 60–90 minutes. Small-group interviews contained two to six participants. We interviewed representatives of four POs during in-person site visits; representatives of the other six POs were interviewed by telephone.

All interviews with returning model participants followed a semistructured format covering topics such as:

- VBID intervention design and changes made from previous years
- CM/DM participation requirements
- Communication with beneficiaries and providers
- Intervention uptake

- Implementation experiences, including implementation barriers and facilitators
- Anticipated mechanisms through which the VBID model may affect health care quality and costs
- Expected VBID outcomes
- Thoughts about uniformity rule changes
- Thoughts on ways to improve the VBID model.

For Year 3 (2019) participating POs, we conducted in-person or telephone interviews with 47 PO representatives across nine model participants (one PO refused to participate in the evaluation). Several POs that left the model test in 2020 only agreed to participate in an abridged interview, which lasted about 30 minutes. We typically interviewed one or two people from such POs and focused these interviews on their decision to leave the model test.

Year 3 interviews were conducted between June and September 2019 by the same team of researchers following similar data collection procedures. We interviewed representatives of three POs during in-person site visits; representatives of the other six POs were interviewed by telephone. As before, most interviews were small group interviews with about five representatives, lasting for approximately 60–90 minutes.

Year 3 interviews with returning model participants followed a semistructured format covering topics such as:

- Changes made to the VBID intervention design, including CM/DM participation requirements
- Communication with beneficiaries and providers
- Intervention uptake
- Implementation experiences, including implementation barriers and facilitators
- VBID outcomes to date
- Thoughts about changes to the VBID model test for 2020 and 2021.

We supplemented these semistructured interviews with a review of POs' VBID application materials. Results of these interviews are presented in Chapters 2–4.

The RAND Institutional Review Board, RAND's internal human subjects protection committee, determined this study to be exempt from review.

Analysis of the Interview Data

All interviews were audio recorded and transcribed. We developed a codebook based on the main topics addressed in the protocol and the codebook used last year. Two authors (DK and CB), who are experienced qualitative researchers, used MAXQDA (a qualitative data analysis software program) to apply the codebook to each transcript and identify key themes independently (MAXQDA, 1989–2019). Each transcript was coded by one coder. Then DK and CB divided the codes and reviewed all text coded into these codes to ensure consistency of codebook application. All disagreements were discussed until consensus was achieved. Some examples of

earlier points of disagreement that were later reconciled include the nuances of CM/DM requirements and enrollment procedures, as well as how to categorize the outcomes, such as whether engagement in care management should be considered an outcome.

Once all data were coded, we developed key themes through identifying commonly addressed topics in consultation with the entire team. We also highlighted issues identified as significant concerns, even if only by one PO.

Appendix B. Methods for Beneficiary Interviews

To learn about beneficiary awareness of the VBID model test and their perception of the VBID benefits, we interviewed 100 beneficiaries who have either received VBID benefits or who were eligible for but did not enroll into VBID. These semistructured interviews were conducted by a team of eight researchers by telephone between August and October 2018 (the second half of Year 2 of the VBID model test).

Sampling

To identify potential interviewees, we used the VBID participation data that participating POs reported to CMS via MARx. POs are required to report whether each VBID-eligible beneficiary is participating ("full status") or is not participating ("unearned status") in VBID. If POs required beneficiaries to participate in CM/DM activities to receive VBID benefits, they had to use an "unearned status" to identify beneficiaries who did not participate in CM/DM activities, but did not actively opt out of VBID. We note that some POs misunderstood this requirement and did not report beneficiary VBID participation status correctly.

In June 2018, we randomly sampled 500 beneficiaries from each Year 2 participating PO using VBID participation data reported through MARx for a total sample of 4,737 (one PO had less than 500 beneficiaries in MARx). If a PO reported some beneficiaries as having "unearned status," we sampled 250 beneficiaries with "full" and 250 beneficiaries with "unearned" status. The information on the sampled beneficiaries also included the beneficiary's name, age, gender, mailing address, and number of months participating in VBID. We were able to identify telephone numbers for 3,185 beneficiaries by using their name and mailing address.

Our *a priori* goal was to interview a total of 100 beneficiaries. We used a quota sampling strategy to recruit ten beneficiaries from each participating PO (five in full status and five in unearned status, if applicable), balance the sample in terms of beneficiary gender, and ensure variation on age (younger than 75 and 75 and older). For each PO, we grouped the beneficiaries by status, gender, and age; put them in random order within each of these sub-groups; and used the resulting ordered list as the basis of recruitment activities within each subgroup.

Recruitment Activities

To meet our sampling goals and to facilitate recruitment, we divided POs into three recruitment waves and staggered recruitment activities. Our recruitment approach included mailing out a one-page letter to beneficiaries and then calling them three days after initial mailing. For each PO, we mailed invitation letters to an initial 25–30 respondents. We allowed an average of three telephone calls to recruit each beneficiary to reduce respondent burden.

Once we reached the maximum number of calls to a given beneficiary, we mailed invitation letters to an additional batch of beneficiaries from the same PO.

After reaching out to beneficiaries in the first recruitment wave, we made a decision to exclude beneficiaries aged 80 and older due to concerns about hearing problems and cognitive abilities to answer interview questions. By adding an age restriction, we reduced the number of potential interviewees from 4,737 to 1,692. We sent invitation letters to 117 and subsequently interviewed three beneficiaries 80 and older from the first recruitment wave.

When calling potential participants, recruiters used a script that explained the purpose of the interview and its length, confidentiality and the voluntary nature of research participation, and the fact that the interview would be audio recorded. If the beneficiary was interested and cognitively able to participate, the recruiter scheduled an interview based on interviewer availability. All potential participants were provided with a number that they could call to schedule an interview. All interviews were conducted in English and audio recorded. As compensation for participating, all participants were sent a $75 check payment after the interview was completed.

Of the total 1,692 potential participants, we mailed letters to 982 beneficiaries. Of these, 234 got only invitation letters and 748 also received at least one call. Among those who received a call, 97 had phone numbers that had been disconnected, 8 were identified as deceased, 68 were ineligible (i.e., older than 80, did not speak English, or were cognitively impaired), 96 refused, 372 did not respond to our phone call(s), 6 were scheduled for but failed to complete an interview, and 101 completed an interview.

Interview Protocol

We created and pilot tested a semistructured interview protocol with open-ended questions and tailored it to the nuances of the VBID interventions designed by each participating PO. Because many POs have participation requirements and do not automatically provide VBID benefits to all eligible beneficiaries, we created two versions of the protocol: one for beneficiaries with full status and another one for those with unearned status (described above). Both versions included general questions about Medicare and MA, clinical conditions that pertain to the interview participant, barriers to care, and VBID awareness. A version of the protocol for VBID-participating beneficiaries also included questions about various VBID benefits, participation requirements, attitudes toward VBID, and expected outcomes. A version of the protocol for beneficiaries who were eligible for VBID but chose not to participate (those who have not completed participation requirements) included questions about reasons for not joining the model test and things that could motivate them to participate in VBID (see Table B.1).

Table B.1. VBID Beneficiary Interview Topics

Interview Topics	Included in Interviews with Beneficiaries in "Full Status"	Included in Interviews with Beneficiaries in "Unearned Status"
Health coverage	X	X
Health conditions	X	X
Barriers to care	X	X
VBID program awareness	X	X
Reasons for not participating in VBID		X
VBID benefits	X	
CM/DM requirements	X	
Attitudes toward VBID	X	
VBID outcomes	X	

Interview Process

All interviews were conducted by one of the eight experienced interviewers trained by the project codirector. Two interviewers were assigned to conduct interviews with beneficiaries from a given PO. Following protocols for this study, a beneficiary was contacted by the assigned interviewer on the scheduled date and time. Once the beneficiary was on the phone, the interviewer obtained oral informed consent to conduct the interview. Each interview lasted for 30–45 minutes and was conducted using the approved protocol for a given PO and beneficiary type.

All interviews were audio recorded and transcribed. After each interview, the interviewer wrote a brief, half-page interview summary. After transcription of the audio recording was completed, the interviewer reviewed the transcripts to fill in any gaps in their summary and the transcript.

Data Analysis

All interview transcripts were uploaded into MAXQDA, a qualitative data analysis software program (1989–2019).

A codebook was collaboratively developed by the team to organize and streamline the identification of relevant themes. The codebook directly maps onto the interview protocol; i.e., each parent code represents a section of interview questions, and subcodes represent the array of possible reactions to the questions posed. The codebook also makes note of potential new or unanticipated findings.

The coding team followed a two-phase (Butler-Kisber, 2010) approach to qualitative data analysis. The first phase (Butler-Kisber, 2010) consisted of reviewing transcripts, discussing what is being revealed, and writing reflective memos to broadly classify emerging themes. Two members of the coding team (KB and CA) trained in qualitative data analysis double-coded

20 interview transcripts. The researchers met weekly to review findings and confirm agreement on the code applications. To ensure rigor and transparency, interrater reliability of the coding across the two coders was evaluated by calculating a pooled Cohen's kappa coefficient and Cohen's kappa for each of the codes (McHugh, 2012). In a small number of instances where kappas fell below 0.8 for a large sample of coded excerpts, the coders reviewed discrepancies and worked through the differences until they could come to consensus.

The second phase of analysis, the fine-grained phase, consisted of the identification of specific words, quotes, and ideas that represent larger themes (Butler-Kisber, 2010). These more discrete units of data were used to support abstract concepts or ideas identified in the first phase of the analysis. The two coders shared these findings with the full qualitative team during weekly meetings.

After completing the coding process, the team jointly developed both qualitative (thematic descriptions) and quantitative (code frequency counts) descriptions of study results.

Sample Characteristics

Our final beneficiary interview sample included 101 participants. We excluded one beneficiary because of his or her limited cognitive abilities to understand and answer the

Table B.2. Description of the Beneficiary Sample (*N* = 100)

Participant Characteristics	Percent
Gender	
Male	48
Age	
<75	44
Ethnicity	
White	96
Education	
Eighth grade or less	2
Some high school	12
High school graduate	40
Some college or a 2-year degree	30
College degree	11
Advanced college degree	5
Employment status	
Retired	81
Work full time	2
Work part time	9
Disabled	6
Volunteer	2
VBID status	
Full	75

interview questions. Of 100 interviewees, 48 percent were males, 44 percent were younger than 75, and 96 percent were white. For education, 40 percent had a high school diploma, 30 percent had completed some college or obtained a two-year degree, 16 percent had completed a four-year degree or higher, and 14 percent had not obtained a high school diploma. The majority (81 percent) were retired. Most (75 percent) were designated full status. On average, participants were VBID eligible for more than one year (mean = 13.25 months; standard deviation [SD] = 5).

Appendix C. Case Study of Parent Organization I's Value-Based Insurance Design Intervention

To provide a more detailed description of one PO's VBID intervention, we conducted a descriptive case study of PO I's telehealth intervention. We chose PO I because it offered a unique approach to designing VBID benefits, which may become more common given the recent changes to the VBID model test design (see Chapter 9). Moreover, PO I is among a very small number of POs with VBID interventions that directly affect how providers deliver care to their patients.

For this case study, we used the results of our interviews with PO I's representatives (see Appendix A) and beneficiaries (see Appendix B), as well as the review of this PO's VBID application materials. We supplemented these data with additional semistructured telephone interviews we conducted with one telehealth nurse[1] and 12 clinicians whose patients participated in the PO I VBID intervention. We describe how we conducted provider interviews below.

Provider Sampling

Eligible providers included all providers (physicians or nurse practitioners) who had at least one patient participating in the PO I VBID intervention. PO I offered the names of 75 unique clinicians in two versions of a list. We obtained the first list in December 2018, and it contained a total of 51 providers (23 primary care providers and 28 cardiologists). The list was then updated in March 2019. Of 61 providers in the second list, 37 providers had already been included in the first list and 24 were new providers (6 primary care and 18 cardiologists). Further, 14 providers (11 primary care providers and 3 cardiologists) from the first list did not appear on the second list, likely because they no longer had any patients participating in VBID or were no longer in PO I's network of providers. Thus, our sampling frame consisted of 75 providers (51 from the first list and 24 from the second list).

Recruitment

Two experienced research coordinators conducted participant recruitment between February and March 2019. They used a multipronged recruitment approach that included outreach through FedEx, telephone, fax, and email to ensure that all 75 clinicians had an opportunity to share their experiences. Providers were first sent a copy of the recruitment letter and CMS endorsement letter via FedEx, and they were offered a compensation of $200 for participation in the interview.

[1] We also interviewed another telehealth nurse as part of our PO interviews focused on care management.

Additional follow-up was conducted by phone, fax, and email. Because PO I modified the list of providers after the start of recruitment, we discontinued active follow-up with the 14 providers who did not appear on the updated list in mid-March.

Sample Characteristics

Of the 75 providers in the sample frame, 12 completed 45- to 60-minute telephone interviews.

Three-quarters of our interviewees were male (See Table C.1). The majority (58 percent) were cardiologists, one-quarter specialized in family medicine, and 17 percent specialized in internal medicine.

Table C.1. Summary of Characteristics of Providers Who Participated in Interviews

Participant Characteristics	*N*	Percent
Gender		
Male	9	75
Specialty		
Cardiology	7	58
Internal medicine	2	17
Family medicine	3	25

Of the 63 remaining clinicians, 51 did not respond to our invitation, 5 refused participation, 4 were no longer employed at the site, 2 had interviews scheduled that were not completed, and 1 reported having no patients from PO I. We do not have additional information on nonresponsiveness of potential participants.

Provider Data Collection

We created and pilot tested a semistructured interview protocol for providers participating in a VBID intervention through PO I. The semistructured nature of the interview protocol allowed us to cover a range of topics while giving providers an opportunity to elaborate on their own perspectives and experiences. The interview protocol included such topics as an understanding of underlying unmet needs and barriers to accessing care; patient motivation to engage in care management; familiarity with the telehealth intervention and with VBID more broadly; perceived advantages, limitations, and burdens of participating in the telehealth intervention; and perspectives on future telehealth interventions for Medicare beneficiaries.

All interviews were conducted by three experienced interviewers between March and April 2019 (the first half of Year 3 of the VBID model test). The interviews were conducted by telephone at the date and time scheduled by the recruitment team. Interviewers obtained oral consent to conduct and record the interview. Each interview lasted 45–60 minutes, and they were

audio recorded and transcribed. After each interview, the interviewer wrote a brief, half-page interview summary. After transcription was completed, the interviewer reviewed the transcripts to fill in any gaps in the summary description and to ensure that all transcripts were scrubbed of personally identifiable data.

Analysis of Provider Interviews

After interviewers verified the accuracy of the interview transcriptions, the transcriptions were uploaded into MAXQDA, a qualitative data analysis software program (1989–2019).

A codebook was collaboratively developed by two members of the research team to systematize the identification of relevant themes. The codebook was developed both deductively, based on the interview protocol, and inductively, based on responses to interview questions. In addition to codes that directly map onto the interview protocol (i.e., there is a code for each section of interview questions), the codebook included subcodes based on the range of responses to the questions posed, which included some unanticipated findings.

Given that the number of interviews was small ($N = 12$), one member of the research team coded the interviews and another member of project leadership reviewed the coding for accuracy. Memos were used to note any questions on the appropriateness of applied codes, which were resolved by a larger team of four members of the research team during weekly calls.

The coded excerpts were analyzed by two members the research team in a detailed review of specific words, quotes, and ideas that represent larger themes (Butler-Kisber, 2010). Repetition of themes, as well as the range of responses, was reported in the description of the study results.

Results

Description of the Telehealth Intervention

PO I designed a telemonitoring intervention for its VBID beneficiaries with CHF to ensure that their vital signs are monitored daily by a nurse. Such monitoring from the patient's home can help identify fluctuations in weight, detect CHF exacerbations, and intervene early, all of which may ultimately reduce avoidable emergency room visits and hospital readmissions. Through this particular intervention, eligible beneficiaries could receive free weight scales, blood pressure cuffs, and pulse oximeters, which would be monitored remotely by the PO's CM/DM staff seven days a week.

The remote monitoring equipment needed for this intervention was about the size of a small alarm clock. It needed an outlet, but it did not require a Wi-Fi connection because the information was transmitted to CM/DM staff using cell towers. Therefore, the equipment could be installed anywhere in the beneficiary's residence, so long as its placement did not compromise patient safety or pose a fall hazard. The monitoring equipment was installed on behalf of PO I by a vendor, though installation could have been curtailed if the vendor found signs of cockroach or

bedbug infestations. Once the equipment was installed, representatives of PO I involved in the intervention contacted the beneficiary's primary care physician or cardiologist to discuss participation in the telehealth monitoring intervention and to confirm agreement with the diuresis protocol (described below).

To participate in the intervention, beneficiaries were required to complete daily measurements and answer yes/no questions about their health, such as "Are you having chest pains?" or "Do you need to talk to your clinician today?" The monitoring equipment was programmed to remind beneficiaries to take their vital signs about every 15 minutes, for up to an hour, if they failed to do so by a time predetermined by each beneficiary. If no readings were transmitted, the beneficiary would receive a phone call from a telemonitoring nurse. Nurses also called the beneficiary if their readings were abnormal (e.g., a weight change of more than 2 pounds from the previous day, or a blood pressure reading or heart rate outside normal limits), or if the beneficiary answered "Yes" to any of the health questions.

Telehealth nurses followed a standard diuresis protocol to increase the beneficiary's diuretic medication dosage over the phone, for up to three days. If a provider did not sign the diuresis protocol, the nurse would call or send a message to the provider through the EHR to determine the next steps. Any time the nurses adjusted medications, providers would receive a notification so they could discuss this medication adjustment with the beneficiary, if needed. If the diuresis protocol was not successful, the nurse would help coordinate an appointment with the treating physician. If a beneficiary was going to be away, he or she could notify the CM/DM staff, and the equipment would be suspended until the beneficiary returned. Every month, providers whose patients were in VBID received trend reports on their patients' vital signs collected through the telehealth intervention, either via an EHR notification or by fax.

Beneficiaries' Perspective on the Telehealth Intervention

VBID beneficiaries we interviewed were generally happy with the telehealth monitoring intervention. Some said that the home monitoring was both convenient and made them feel cared for and valued: "I'm really impressed with the program. They put it in my home, brought it, put it up, set it up, showed me how to use it," said one beneficiary. Another beneficiary said that machine monitoring makes her feel secure, knowing that nurses will notify her of any problems or issues that may occur: "There's somebody that keeps an eye on my vitals every day and if my blood pressure's too low, I'll know immediately." A PO I representative told us that they serve many rural beneficiaries who "love living where there's not very much hustle and bustle…[and] don't want to go anywhere [to get care where they] have to get up and go so far away and have to be on the road all the time." A small number of beneficiaries we interviewed, however, took issue with the telemonitoring benefit, citing a preference for face-to-face interaction with their physicians. One said: "I don't want virtual anything. I want to see face-to-face my doctor. I want to get to know my doctor. I want him to get to know me. I don't want to be a number."

Providers' Perspective on the Telehealth Intervention

Providers whose patients were in the telemonitoring intervention generally considered it to be useful for increasing patients' awareness of their condition, helping them take better care of their health, and reducing unnecessary emergency room visits and hospitalizations. Although most were enthusiastic about continuous monitoring of their CHF patients, some providers noted that they did not fully understand how the intervention works, including how patients are enrolled, how and who (i.e., nurse or provider) should make medication adjustments, whether nurses could adjust medications without consulting them first, and the frequency with which they receive reports about their patients. This lack of clarity led some providers to feel overwhelmed by the seemingly constant flow of data on their intervention-participating patients and frustrated by the fact that they are not directly compensated for the extra work and time incurred by the intervention. "It's a lot of data in addition to the work I already have," said one provider. "I don't have extra time to review it," said another provider. Others noted that some of their colleagues voiced concerns about the intervention, claiming that providers whose patients are in it could "be liable for the decisionmaking, or lack of decisionmaking, based upon the [monitoring] data" that sometimes include information that is "noncardiac," such as complaints about diarrhea or recent falls.

Cardiologists and primary care physicians we interviewed suggested ways to make the intervention more provider friendly, such as simplifying the way vital sign information is presented in the reports they receive, ensuring that data are automatically transmitted to the patients' EHRs, providing additional information about a patient when clinical decisions need to be made quickly, and further automating the intervention by reaching out to treating physicians only with "a specific summary question asking what they think should [be done to the patient only in urgent clinical situations]." Several providers noted the need to train telehealth monitoring nurses to "look at [longer term] trends [in a patient's data], initiate changes based on a good algorithm, and ask the right screening questions to make sure the person is compliant with the plan of care that has been outlined to them" by the physician. Essentially, providers expressed needing "a little more meat" in the form of more long-term data trends to be able to make "faster, more informed judgment calls."

Summary

Unlike other VBID interventions described in this report, providers with patients participating in the PO I telehealth intervention receive remotely collected information about their patients' conditions. This feedback makes PO I providers active VBID participants. The providers we interviewed generally expressed support for the intention of the telehealth intervention, even if some found issues with its implementation, such as the level of detail provided in reports and lack of understanding of how data are collected or who should take ownership of using the data to manage patients' symptoms. Several providers recommended

ways to improve how information on vital signs, for example, is communicated between the telehealth nurses and providers, or how telehealth nurses might be trained to take the long view of patient data trends into account. Though one beneficiary expressed a dislike of the intervention, preferring instead to develop personal relationships with providers during face-to-face office visits, other beneficiaries commented that they were happy to participate. One reported feeling secure because of the close symptom management. Others were happy with the ease of participation, particularly with the equipment setup.

Appendix D. Methods for Quantitative Analyses

This appendix describes the methods used throughout the quantitative analyses presented in Chapters 6–8, as well as Appendices E–I. Specifically, this appendix covers the approach to matching VBID participants and nonparticipants at the plan and beneficiary level, the difference-in differences models used, and the parallel trends assumption.

Approach to Matching Value-Based Insurance Design-Participating and Value-Based Insurance Design-Nonparticipating Plan Benefit Packages

MA PBPs were not randomly selected to participate in the VBID model; thus, VBID-participating MA PBPs may differ in significant and potentially unobservable ways from those that chose not to participate. We used propensity score matching methods to adjust for any observed differences between participating and nonparticipating PBPs. The general approach was to match VBID-participating PBPs with PBPs that are not participating in VBID. We initially considered three different comparison groups: (1) PBPs outside VBID states, (2) PBPs in VBID states that did not elect to participate in the VBID model, and (3) PBPs in VBID states in the same PO as VBID-participating PBPs. We immediately ruled out the third comparison group because less than half of the participating PBPs had potential comparison PBPs in both the same state and PO. Although we retained the within-state matching approach as an option, our qualitative interviews suggested that there are likely important unobservable differences between participating and eligible but nonparticipating PBPs. For example, eligible nonparticipants seemed more risk averse and concerned about ROI than participating POs, while participating POs expressed enthusiasm about the opportunity to be at the forefront of MA benefit design. These differences suggest that the out-of-state comparison group might be preferable to the within-state comparison group. Because out-of-state PBPs were ineligible to participate in the VBID model, the out-of-state comparison group might be more likely than the within-state comparison group to contain POs that share VBID participants' innovation-focused philosophy.

Appendix D of the first annual report contains a more detailed comparison of the out-of-state and within-state PBP comparison group matching approaches. Here, we focus on the out-of-state comparison group, because this was identified as the primary comparison group for the evaluation. We limited the comparison group to PBPs drawn from within ten states that had metropolitan regions that resembled those in VBID-eligible states, and that were not among states in which VBID was expected to become available during the early years (2017–2019) of the model test.[1]

[1] These ten states were Arkansas, Connecticut, Illinois, Louisiana, Missouri, New York, Oklahoma, Ohio, Washington, and Wisconsin.

From within these ten states, we determined which PBPs would have been eligible for the VBID model if it were available to them. We used the September 1, 2015, CMS criteria to identify PBPs in both comparison states that would be considered eligible to offer MA VBID benefits. However, we modified the minimum enrollment size requirement (2,000 enrollees) to reflect the fact that CMS allowed some smaller PBPs to participate as long as they were in a contract with a larger plan.[2] More details on the specific eligibility criteria and data sources used to for identification are available in the first annual report.

Table D.1 lists the characteristics we used to match VBID-participating and comparison PBPs, along with the data source and year for each characteristic. Generally, POs entered all segments[3] within a given PBP into the model test. However, one PO (PO E) entered selected segments from some PBPs. In these cases, we defined the treatment PBP using data only from participating segments. We performed the matching using a greedy nearest neighbor propensity score match, implemented with the R package *Matching* (Sekhon, 2011). A greedy matching approach cycles through each VBID-participating PBP, finds the best match for that PBP from the yet-to-be matched comparison PBPs, and does not reassess the match. That is, once two PBPs are matched, the algorithm does not change its mind. We used logistic regression predicting PBP participation in VBID to estimate propensity scores, including the main effects of all characteristics in Table D.1,[4] except for the indicator that a PBP offers Part D coverage, which was matched exactly. PBPs were matched one-to-one without replacement based on the propensity score.

We made several refinements to the matching strategy originally described in our first annual evaluation report. First, we forced VBID-participating PBPs that offer Part D to be matched to comparison PBPs that offer Part D (and likewise for participating PBPs that did not offer Part D). We made this modification because some analyses are subset to PBPs that either offer Part D coverage, or those that do not. Without including the Part D indicator into the matching process, we have no guarantee that a VBID-participating PBP that offers Part D will be matched to a comparison PBP that also offers Part D. In addition, we refined the Medicare cost variable used in the matching procedure to reflect standardized Medicare costs instead of unstandardized Medicare spending. Standardized Medicare costs remove geographic variation in payment rates and adjust for beneficiary health using the same risk adjustment model used for payment of MA plans (CMS, 2019b).

---

[2] CMS reserved the right to grant exceptions to the criteria and some were relaxed in approving 2017 plans. We do not account for any exceptions other than the minimum enrollment exception in identifying eligible plans.

[3] A segment is a portion of the service area for each PBP that can consist of one or more counties. POs can choose to break up a service area into segments for their Part C benefits only (Part D service areas cannot be broken into segments). The segmentation allows POs to charge different premiums or cost sharing for Part C benefits to a portion of the service area (such as a rural and urban county within the service area).

[4] Median household income, standardized Medicare costs, enrollment size, and OOP maximums were included in the propensity score model on the log scale.

Table D.1. Characteristics Used to Match Comparison PBPs

Variable	Data Source
County-level measures	
Percentage of population older than 65	Area health resource file
Median household income	Area health resource file
Standardized Medicare costs (per capita)	Medicare geographic variation public use file
PBP or PO measures	
OOP maximum	PBP data
PO market penetration	MA enrollment file
Enrollment size	Beneficiary fact table
Offers Part D	Public plan-level enrollment data file
Beneficiary-level measures	
Mean age	Beneficiary fact table
Percentage male	Beneficiary fact table
Percentage non-Hispanic white	Medicare Bayesian improved surname geocoding version 2.0 (Haas et al., 2019)
Percentage non-Hispanic black	Medicare Bayesian improved surname geocoding version 2.0 (Haas et al., 2019)
Percentage Hispanic	Medicare Bayesian improved surname geocoding version 2.0 (Haas et al., 2019)
Percentage dually eligible for Medicare and Medicaid	Beneficiary fact table
Percentage with each of four chronic and comorbid conditions (CHF, diabetes, COPD, cancer)	Beneficiary risk score data
Mean risk score	Beneficiary risk score data

Table D.2 provides the means and standard deviations of the characteristics for VBID-participating PBPs, all eligible and nonparticipating PBPs outside VBID states, and the set of matched nonparticipating PBPs. The 45 participating PBPs that joined the VBID model in 2017 were matched to nonparticipating PBPs using information from 2016, whereas the one participating PBP that joined in 2018 was matched to a nonparticipating PBP using information from 2017. The matching for the one participating PBP that joined the VBID model test in 2018 was not matched using a propensity score, but instead, matched directly on the covariates using the Mahalanobis distance metric (Stuart, 2010). Plans were matched a single time, and we retained the matched comparator for the duration of the study. Because we intentionally selected matched comparators from among ten states ($N = 243$ PBPs) that were unlikely to be offered VBID before 2020, there were no instances in which a matched comparison PBP joined the model test.

We find large differences in the means between the groups prior to matching, with VBID-participating PBPs tending to have older beneficiaries and a higher percentage of white beneficiaries, fewer dual-eligible beneficiaries, and lower OOP maximums, and tending to serve beneficiaries in counties with higher Medicare spending than nonparticipating PBPs.

Table D.2. Mean and Standard Deviation of Characteristics Before and After Matching

Measures	Participating PBPs Mean	Standard Deviation	Nonparticipating PBPs Outside VBID States[a] Mean	Standard Deviation	Matched Nonparticipating PBPs Outside VBID States[a] Mean	Standard Deviation
Total number of PBPs	46		243		46	
Total number of PBPs offering Part D	33		204		33	
County level						
Population older than 65 (percent)	14.6	1.8	13.7	1.7	14.4	1.3
Median household income	57,758	10,568	52,348	10,255	53,038	11,745
Standardized Medicare costs (per capita)	9,646	382	9,416	1,114	9,546	1,210
PBP or PO						
OOP maximum	4,413	1,228	5,351	1,455	4,735	1,571
Enrollment	9,759	11,535	8,505	12,209	9,527	18,608
PO market penetration (percent)	33.5	12.1	36.1	10.7	33.5	9.4
Beneficiary level						
Age (mean)	76.7	4.2	73.4	3.2	75.5	3.1
Male (percent)	45.5	9.4	46.0	8.6	46.9	9.4
Race/white (percent)	91.1	5.8	81.9	17.9	90.6	4.6
Race/black (percent)	3.9	4.9	9.3	11.1	4.3	3.9
Race/Hispanic (percent)	1.7	0.6	4.1	7.0	1.7	0.9
Dually eligible for Medicare and Medicaid (percent)	6.0	3.3	11.6	10.8	6.0	4.2
Risk score (mean)	1.1	0.2	1.0	0.2	1.0	0.2
COPD (percent)	12.1	3.8	12.0	3.6	11.2	3.8
CHF (percent)	11.1	4.7	9.5	3.1	9.8	3.6
Diabetes (percent)	22.5	5.7	23.6	5.3	21.7	6.2
Cancer (percent)	12.4	3.8	8.9	2.7	10.7	2.4

[a] VBID states refers to states with at least one participating PBP.

Table D.3 summarizes the balance of the characteristics before and after matching. The average absolute standardized difference before matching is 0.73.[5] After matching, this is reduced to 0.18. This indicates that matching improved the similarity between the comparison group and the VBID-participating PBPs. Standardized differences of 0.2, 0.5, and 0.8 are often considered small, medium, and large (Austin, 2009). However, there is no consensus as to what

Table D.3. Standardized Differences of Characteristics Before and After Matching

Measures	Standardized Difference Before Matching	Standardized Differences After Matching
Number of plans		
VBID	46	46
Comparison	243	46
County level		
Percentage of the population older than 65	0.50	0.12
Median household income	0.51	0.45
Standardized Medicare costs (per capita)	0.60	0.26
Plan or PO		
Offers Part D	−0.27	0.00
OOP maximum	−0.76	−0.26
PO market penetration	−0.22	−0.01
Enrollment	0.11	0.02
Beneficiary level		
Age (mean)	0.79	0.30
Male (percent)	−0.06	−0.15
Race/white (percent)	1.59	0.1
Race/black (percent)	−1.11	−0.09
Race/Hispanic (percent)	−3.84	−0.03
Dually eligible for Medicare and Medicaid (percent)	−1.73	−0.03
Risk score (mean)	0.40	0.32
COPD (percent)	0.04	0.25
CHF (percent)	0.35	0.27
Diabetes (percent)	−0.21	0.13
Cancer (percent)	0.92	0.45
Average absolute standardized difference	0.73	0.18

[5] The *standardized difference* is the mean of VBID-participating PBPs minus the mean of the comparison PBPs divided by the standard deviation of the VBID-participating PBPs.

constitutes a sufficiently small standardized difference, with some suggesting thresholds as low as 0.03, 0.05, or 0.10 (Austin and Stuart, 2015; Caliendo and Kopeinig, 2008; Normand et al., 2001). Regardless of the exact threshold, it is clear that important differences between our groups remain, even after matching. Notably, VBID-participating PBPs serve beneficiaries in counties that have higher standardized Medicare costs and a higher median income compared to the matched comparison PBPs (standardized difference of 0.26 and 0.45, respectively). Further, while our matching strategy controls for a wide range of PBP characteristics, some factors—such as baseline MA premiums, benefit design features other than the out-of-pocket maximum, and measures of local market competition, are not included. We account for the differences between participating and comparison PBPs that remain after matching by using a difference-in-differences model, described in greater detail below.

Several other approaches were implemented in an attempt to improve the similarity of the VBID-participating and comparison PBPs. These included, but are not limited to, matching on the covariates directly using the Mahalanobis distance metric, using a logistic regression-based propensity score weight (Lunceford and Davidian, 2004), using a gradient boosting-based propensity score weight (Ridgeway et al., 2017), using the covariate-balancing propensity score (Imai and Ratkovic, 2014), using entropy balancing weights (Hainmueller, 2012), and modifying the set of characteristics being matched. No method was superior to all of the others on the basis of the average absolute standardized difference.

Approach to Matching Eligible Beneficiaries in Participating Plan Benefit Packages with Eligible Beneficiaries in Nonparticipating Plan Benefit Packages

To evaluate the effect of the VBID model test on beneficiary-level outcomes, we matched eligible beneficiaries in participating PBPs with beneficiaries who would be eligible for VBID, but who are enrolled in a matched nonparticipating PBP. The first step of this process was to obtain detailed information from each PO regarding how they identify eligible beneficiaries and to replicate each participating PO's approach. More details are available in the first annual report.

We implemented a greedy Mahalanobis distance matching using the R package matching (Sekhon, 2011) separately for each PO, matching 1:1 without replacement within each participating PO but with replacement across participating POs. This was done for two reasons. First, each PO has different eligibility requirements; therefore, the set of eligible comparison beneficiaries is different for each participating PO. Second, the number of eligible comparison beneficiaries is not sufficient to match without replacement across participating POs. We included a similar set of characteristics from the PBP matching. The full list can be found in Table D.4, along with mean characteristics before and after the beneficiary matching.

Table D.4. Mean of Characteristics Before and After Beneficiary Matching

Measure	Eligible Beneficiary in Participating PBP	Eligible Beneficiary in Nonparticipating PBP[a]	Matched Beneficiary in Nonparticipating PBP
PBP characteristics			
Offers Part D (percent)	95.6	93.4	95.6
PO market penetration (percent)	38.7	40.0	38.5
OOP maximum ($)	4,606	4,843	4,633
County characteristics			
Percentage of the population older than 65	14.9	13.7	14.3
Median household income ($)	57,514	54,254	54,638
Standardized Medicare costs ($, per capita)	9,670	9,711	9,685
Beneficiary characteristics			
Age	75.0	74.4	75.0
Percent female	53.1	54.5	53.0
Native American	0.2	0.5	0.2
Asian Pacific Islander	1.4	1.0	1.4
Black	3.7	8.0	3.7
Hispanic	1.6	1.8	1.6
Multiple races	1.6	2.0	1.6
Dually eligible for Medicare and Medicaid (percent)	11.0	12.9	11.0
Low income subsidy (percent)	15.6	17.9	15.2
Disabled (percent)	16.4	20.9	16.2
Months enrolled in PBP	11.8	11.8	11.8
Risk score	1.6	1.6	1.6
COPD	30.3	27.7	30.3
CHF	31.9	27.4	31.9
Diabetes	41.5	46.9	41.5
Cancer	14.3	12.3	14.1
Hypertension	52.1	58.1	52.2

[a] Eligible beneficiaries in nonparticipating PBPs includes all beneficiaries in nonparticipating PBPs that meet at least one of the POs' eligibility requirements.

Table D.5 provides the standardized differences before and after the beneficiary matching. We note that PBP and county characteristics were included in the match with the goal of improving the similarity of the PBP characteristics between the beneficiary-level matched set. There is a modest gain in the balance of the PBP and county characteristics in the beneficiary-level analysis compared to the PBP-level analyses reported in Table D.3, with the average absolute standardized difference being 0.17 before matching and 0.08 after matching. The beneficiary characteristics are very similar prior to matching, with only a few notable differences. These include eligible beneficiaries in participating PBPs being less likely to be black (3.7 versus 8.0 percent); more likely to have COPD (30.3 versus 27.7 percent), CHF (31.9 versus 27.4 percent), and cancer (14.3 versus 12.3 percent); less likely to have diabetes

Table D.5. Standardized Differences of Characteristics Before and After Beneficiary Matching

Measure	Eligible Beneficiary in Nonparticipating PBP[a]	Matched Beneficiary in Nonparticipating PBP
PBP characteristics		
Offers Part D (percent)	0.11	0.00
OOP maximum ($)	−0.18	−0.02
PO market penetration	−0.07	0.01
County characteristics		
Percentage of the population older than 65	0.45	0.24
Median household income ($)	0.28	0.24
Standardized Medicare costs ($, per capita)	−0.11	−0.04
Average absolute standardized difference, PBP, and county level characteristics	0.17	0.08
Beneficiary characteristics		
Age	0.06	−0.01
Percent female	−0.03	0.00
Race/Native American	−0.45	−0.02
Race/Asian Pacific Islander	0.04	0.01
Race/black	−0.24	0.01
Race/Hispanic	−0.03	0.01
Multiple races	−0.32	−0.04
Dually eligible for Medicare and Medicaid (percent)	−0.06	0.00
Low income subsidy	−0.06	0.01
Disabled	−0.12	0.00
Months enrolled in PBP	0.01	−0.01
Risk score	0.04	0.02
COPD	0.06	0.00
CHF	0.10	0.00
Diabetes	−0.11	0.00
Cancer	0.05	0.00
Hypertension	−0.12	0.00
Average absolute standardized difference, beneficiary characteristics	0.10	0.01
Average absolute standardized difference, all characteristics (PBP, county, and beneficiary level)	0.13	0.03

[a] Eligible beneficiaries in nonparticipating PBPs include all beneficiaries in nonparticipating PBPs that meet at least one of the POs' eligibility requirements.

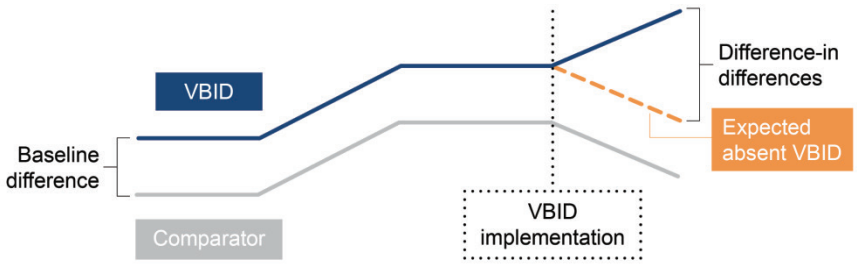
(41.5 versus 46.9 percent) and hypertension (52.1 versus 58.1 percent); and more likely to have a higher risk score (1.64 versus 1.59). The balance of the beneficiary characteristics improves after matching, with the average absolute standardized difference going from 0.10 before matching to 0.01 after matching. When we consider all characteristics (PBP level, county level, and beneficiary level), the average absolute standardized difference falls from 0.13 to 0.03 after matching.

The beneficiary matching was successful at improving the similarity between the beneficiaries in participating PBPs and those in nonparticipating PBPs.

Difference-in-Differences Model and Parallel Trends Assumption

We accounted for any differences between VBID-participating and matched comparison PBPs using difference-in-differences analyses. The difference-in-differences approach analyzes whether trends for treated and nontreated observations diverged after the intervention (in this case, the VBID model test). Figure D.1 provides a heuristic overview of the difference-in-difference approach. The trend for VBID-participating PBPs (the blue line) and the trend for comparison PBPs (the grey line) are parallel prior to the implementation of VBID but diverge after VBID implementation. The approach assumes that the trend for VBID-participating PBPs would have followed that of comparison PBPs absent the VBID model (the orange dashed line), and the difference-in-differences estimate of the effect of VBID is the difference between the actual trend in VBID-participating PBPs and this expected trend absent VBID.

Figure D.1. Conceptual Illustration of the Difference-in-Differences Methodology



The difference-in-differences approach should account for both observed and unobserved differences between participating and comparison PBPs. However, a key assumption of the difference-in-differences analysis is that—without the intervention—trends in outcomes between participating and comparison PBPs would have been similar. This assumption must hold for our final comparison group, which matches PBPs and beneficiaries within PBPs.

To test the parallel trends assumption, we assessed whether trends for key outcome variables were similar before the model test was implemented. Let y_{cpt} be the outcome for PBP p in PO c at year t, let $VBID_{cp}$ be an indicator that the pth PBP in PO c is a VBID-participating PBP, and let X_{cpt} be a set of additional characteristics to be controlled for in the analysis, including the PBP OOP maximum and the PBP premium. We consider three years of data prior to implementation of the test model, so that $t \in \{2014, 2015, 2016\}$. Our model for assessing the parallel trends assumption is given by

$$y_{cpt} = \alpha + \alpha_t + \theta_c + \delta * VBID_{cp} + \gamma_t * VBID_{cp} + \beta^T X_{cpt} + \varepsilon_{cpt}, \qquad \text{(D.1)}$$

where

- α = overall intercept
- α_t = year fixed effect (with $\alpha_{2014} = 0$) that captures the trend over time in the comparison PBPs
- θ_c = PO fixed effect capturing time-invariant differences between POs
- δ = baseline difference between VBID and comparison PBPs
- γ_t = interaction effect between time and VBID-participating PBPs (with $\gamma_{2014} = 0$) that captures the difference in trends over time between participating and comparison PBPs
- β = effect of the additional characteristics included in the model.

Under this model, the error terms ε_{cpt} are not independent within PBP, and different approaches were used to account for the nonindependence. The primary approach was to fit the model using a generalized estimating equation (GEE) assuming an order 2 autoregressive (AR-2) working correlation structure (Zeger, Liang, and Albert, 1988). The parameters from the GEE are consistently estimated even if the working correlation structure is misspecified, and a robust standard error (SE) was used to ensure valid inference. An unstructured working correlation structure was also considered, but the number of parameters needed for an unstructured correlation matrix increases as the number of time points increases, whereas the AR-2 structure always has two parameters. Thus, the AR-2 structure was chosen based on a tradeoff between parsimony and flexibility.

The model formulation given in Equation (D.1) was modified to estimate the final difference-in-differences effect of VBID by setting the interaction term γ_t to be zero during the period prior to the implementation of VBID. That is, $\gamma_t = 0$ for $t \in \{2014, 2015, 2016\}$. In addition, we also consider a model that pooled the effect of VBID by constraining the γ_t to have the same value γ; i.e., $\gamma_t = \gamma$ for $t \geq 2017$. Note that this is a standard difference-in-differences model that includes a time-fixed effect (α_t), a VBID main effect (δ), and the difference-in-differences coefficients (γ_t).

The regression model described in Equation (D.1) is linear, which we used for outcomes that were continuous or approximately continuous. For other outcome types, the GEE model was adjusted to a specification that is appropriate for the outcome, for example, logistic regression for dichotomous outcomes and negative-binomial regression for count outcomes. The model in Equation (D.1) is described at the PBP level. Some outcomes were measured at the beneficiary level or contract level. For contract-level outcomes, the model formulation remains the same but without indexing by PBP. For beneficiary-level outcomes, the model formulation is expanded to index by beneficiary (i.e., y_{cpit} is the outcome for beneficiary *I* in PBP *p* in PO *c* at year *t*) and to include additional beneficiary characteristics (e.g., age, gender, and disability status).

The test of parallel trends tests the hypothesis that H_0: $\gamma_t = 0$ for all $t \in \{2014, 2015, 2016\}$. The assumption of parallel trends is critical to the validity of the difference-in-differences

models that we apply in our analysis. However, because of the breadth of outcomes considered in our analysis, parallel trends did not hold in all cases.

When the parallel trends assumption did not hold, we implemented a reweighting approach that combines aspects of the methods described by Stuart et al. (2014) and Abadie (2005). We start with the set of matched comparisons described earlier, which draws only from 2016 data. Then, for each outcome, the comparison PBPs are weighted to ensure that the trends in that outcome for comparison PBPs are as similar as possible to VBID-participating PBPs.

We implemented two different approaches for reweighting the comparison PBP trends, one based on a propensity score similar to Stuart et al. (2014) and one based on entropy balancing Hainmueller (2012). In both cases, we sought to balance two quantities: The mean difference in the outcome between 2014 and 2015 and the mean difference in the outcome between 2015 and 2016. This was achieved by including the corresponding differences at the PBP level as covariates in the two approaches.

The propensity score–based weights were the preferred approach and were derived as the usual average treatment effect on the treated weights from a logistic regression with only the two differences as predictors. If the propensity score–based weights achieved parallel trends, then they were used. Otherwise, the entropy balancing weights were used. This strategy avoids the inference issues of the synthetic control method, and our approach matches only *trends* in outcomes over time, rather than the *levels* of outcomes.

One nuance of the reweighting approaches is how we handled missing data. For the propensity score weights, missing values were imputed for the derivation of the weights only, but the difference-in-differences models were restricted to observed data only. For the entropy balancing weights, we modified the algorithm to allow for missing data. This was achieved by limiting the moment conditions to the observed data relevant for that moment (e.g., the mean of a covariate would be the mean among the observed data for that covariate). Otherwise, the algorithm is identical to the original entropy balancing algorithm.

Appendix E. Medicare Advantage & Prescription Drug Plan CAHPS Methods and Results

Participation in DM programs, lower cost-sharing, and additional supplemental benefits may change patients' experiences with their PBP over time. The MA & PDP CAHPS Survey collects patient responses regarding their experiences with MA and Part D plans. In addition to measuring experiences of care, MA & PDP CAHPS also measures beneficiaries' awareness of the VBID model benefits. It is possible that beneficiaries are not aware of the model benefits, yet still report improved experiences with their PBP due to receipt of the VBID benefits.

Study Population and Data Source

The population of interest was VBID-eligible beneficiaries in VBID-participating PBPs compared to VBID-eligible beneficiaries in comparison to PBPs. The VBID eligibility status of beneficiaries was determined by RAND's implementation of the beneficiary eligibility algorithms (see first Annual Report, Eibner et al., 2018, for more details). MA & PDP CAHPS data from VBID-participating PBPs and matched comparison PBPs for the years 2014–2018 were used to explore patterns of patient experience.

Measures

MA & PDP CAHPS measures many different experiences of care and summarizes these experiences with a set of six composite measures and five overall rating measures. These measures are as follows:

- Composites
  - Getting Needed Prescription Drugs
  - Doctors Who Communicate Well
  - Getting Needed Care
  - Getting Appointments and Care Quickly
  - Customer Service
  - Care Coordination
- Ratings of care
  - Health Care Quality
  - Doctor
  - Specialist
  - Health Plan
  - Prescription Drug Plan.

All MA & PDP CAHPS composites and ratings are coded on a 0–100 scale regardless of the original scale of the items. For example, beneficiaries are asked to rate their health care on a

0–10 scale, but this item is converted to a 0–100 scale for consistency across measures and ease of presentation.

Methods

The sampling design of the MA & PDP CAHPS survey is focused on collecting data sufficient for summarizing the experiences of beneficiaries at the MA & PDP contract level. As such, it is not designed to collect a sample of beneficiaries within any particular PBP, and it does not guarantee that VBID-eligible beneficiaries will be sampled. For these reasons, we were unable to use the beneficiary match described in Appendix D, because there were too few survey respondents among the matched comparison beneficiaries. Instead, we used data from all VBID-eligible beneficiaries in the VBID-participating and matched comparison PBPs who completed the MA & PDP CAHPS survey for each year of 2014–2018, including a supplemental sample that was fielded in 2016.

For each patient experience measure, let Y_{pit} be the 0–100 score of the ith beneficiary in the pth PBP in year t. Let X_{pit} be the set of MA & PDP CAHPS case-mix adjusters, and let $VBID_p$ be an indicator of VBID participation for the pth PBP. Consider the following case-mix adjusted difference-in-differences linear regression model:

$$y_{pit} = \alpha + \alpha_t + \lambda_p + \delta * VBID_p + \gamma_t * VBID_p + \beta^T X_{pit} + \varepsilon_{pit}, \quad \text{(E.1)}$$

where

- α_t = year fixed effect (with $\alpha_{2014} = 0$)
- λ_p = PBP-level random effect
- δ = baseline difference between VBID and comparison PBPs
- γ_t = interaction between time and VBID-participating PBPs (with $\gamma_t = 0$ for $t \leq 2016$ and $\gamma_{2017} = \gamma_{2018}$)
- β = effect of the case-mix adjusters included in the model.

The effect of interest is γ_t, which represents the difference-in-differences coefficient. The parallel trends assumption was tested using the same strategy as described in Appendix D, but using the mixed model specification described here. This model included PBP-level random effect in place of the PBP-level (or PO-level) fixed effect that was used elsewhere in the report due to the limited sample size of the MA & PDP CAHPS survey.

Results

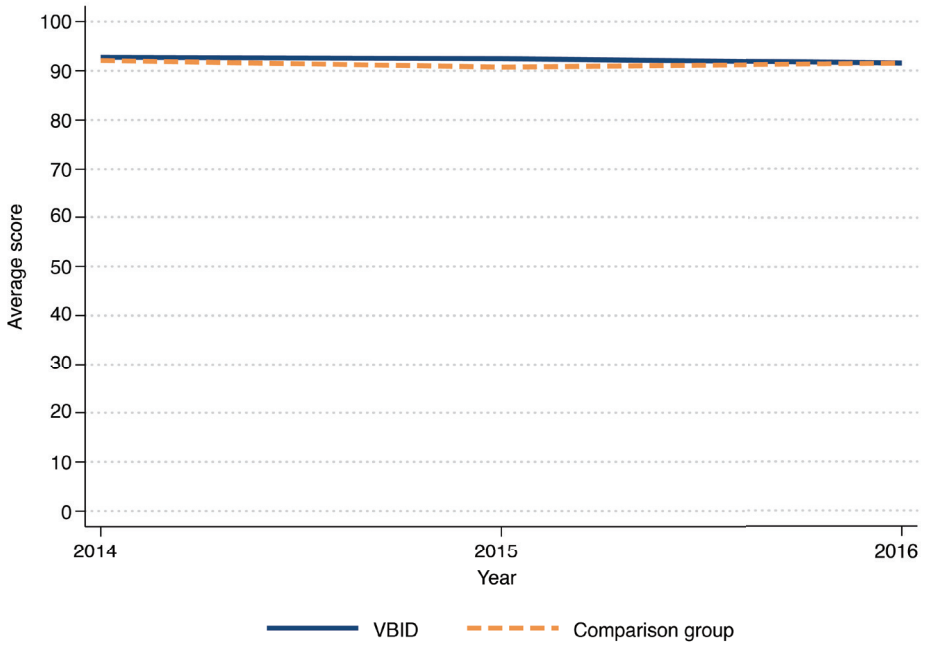
Table E.1 provides tests of parallel trends comparing VBID-participating PBPs to matched comparison PBPs for the MA & PDP CAHPS measures from 2014 to 2016. Five of the 11 measures reject the test of parallel trends between VBID-participating and comparison PBPs,

Table E.1. MA & PDP CAHPS Tests of Parallel Trends, 2014–2016

Measure	Parallel Trend Test p-Value	Sample Size
Getting Needed Prescription Drugs	0.12	10,136
Doctors Who Communicate Well	0.02	9,450
Getting Needed Care	0.06	7,975
Getting Appointments and Care Quickly	0.00	10,204
Customer Service	0.38	3,917
Care Coordination	0.01	9,791
Rating of Health Care Quality	0.03	11,112
Rating of Doctor	0.01	9,373
Rating of Specialist	0.39	6,622
Rating of Health Plan	0.08	10,928
Rating of Prescription Drug Plan	0.78	10,198

indicating a departure of trends during the pre-VBID period. Figures E.1–E.11 show the trends in the MA & PDP CAHPS measures from 2014 to 2016 by VBID participation status. Focusing on the five measures with p-values less than 0.05 from Table E.1, we note that there is an increase in scores from 2014 to 2015 among VBID-eligible beneficiaries in VBID-participating PBPs followed by a drop from 2015 to 2016, as exemplified by Getting Appointments and Care

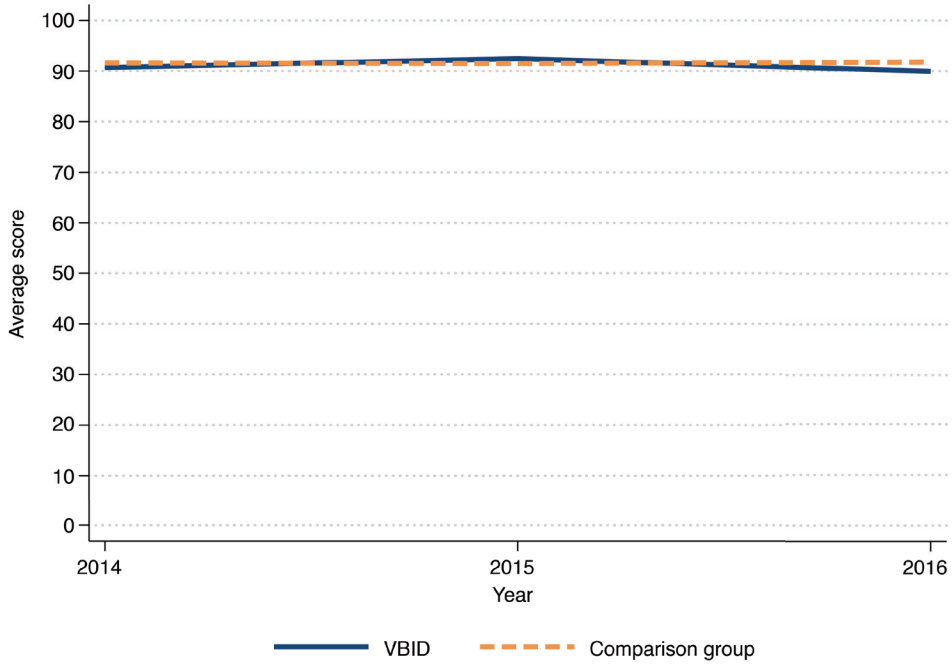
Figure E.1. Average Score of Getting Needed Prescription Drugs by VBID-Participation Status and Year, 2014–2016 (*N* = 10,136)



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.12.

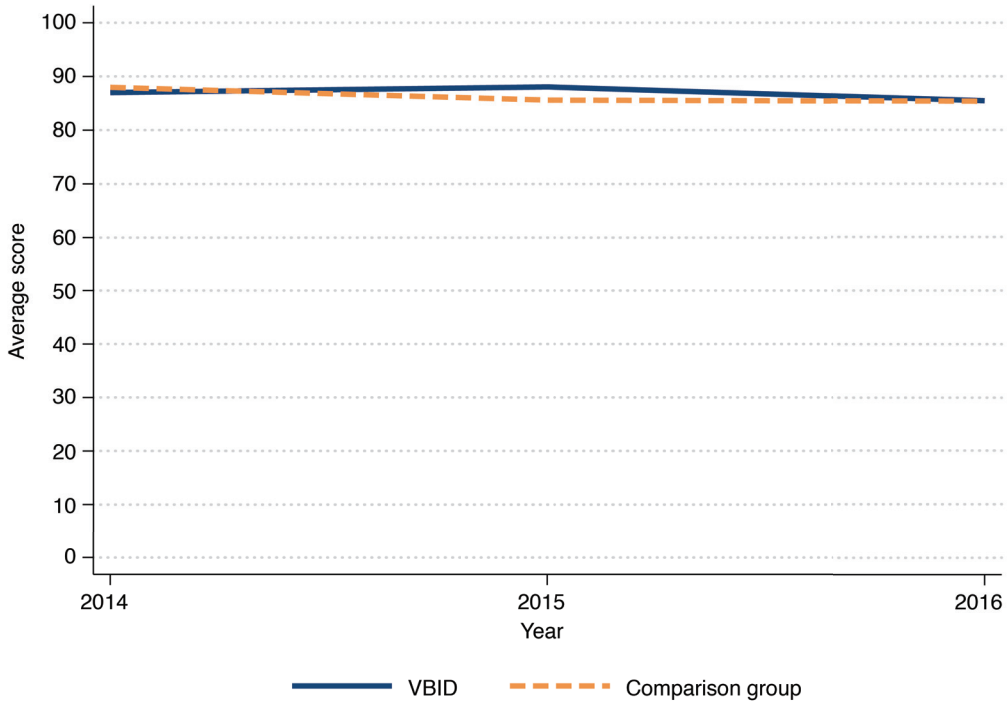
Quickly in Figure E.4. This pattern also occurs in beneficiaries who are not eligible for VBID in VBID-participating PBPs (not shown), although not to the same extent as VBID-eligible beneficiaries. This indicates that the one-year spike in scores is systematic to VBID-participating PBPs and not specific to VBID-eligible beneficiaries in VBID-participating PBPs. This pattern of scores could be explained by actual improvements in experiences of care between 2014 and 2015, random variation due to MA & PDP CAHPS surveying only a sample of beneficiaries, or other survey-related sources of error. Because we are unable to explain the pattern observed in VBID-participating PBPs, we chose to avoid the reweighting strategies for ensuring parallel trends that are described in Appendix D. Instead, we interpret the differences as random variation, proceed with the difference-in-differences analyses, and interpret the results with caution.

Figure E.2. Average Score of Doctors Who Communicate Well by VBID-Participation Status and Year, 2014–2016 (*N* = 9,450)



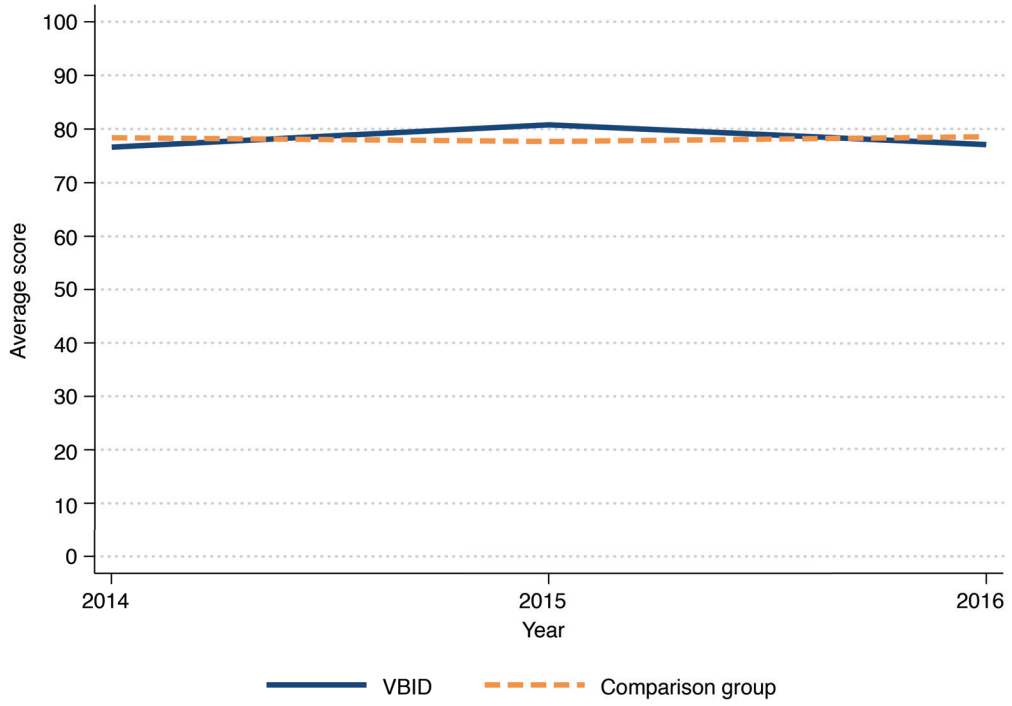
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.02.

Figure E.3. Average Score of Getting Needed Care by VBID-Participation Status and Year, 2014–2016 (*N* = 7,975)



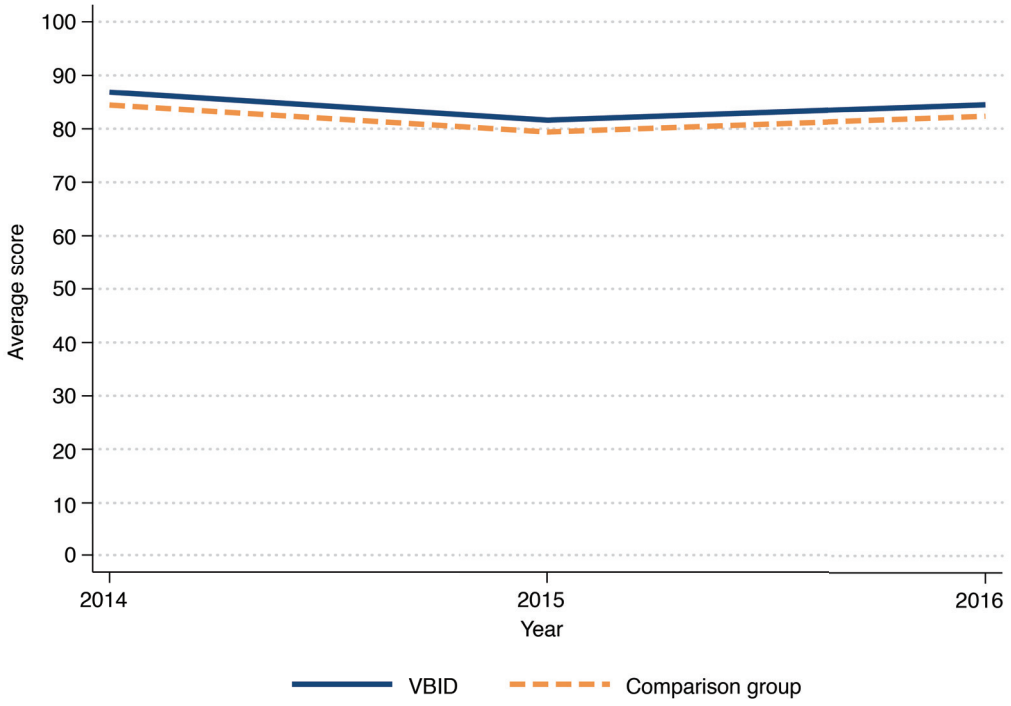
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.06.

Figure E.4. Average Score of Getting Appointments and Care Quickly by VBID-Participation Status and Year, 2014–2016 (*N* = 10,204)



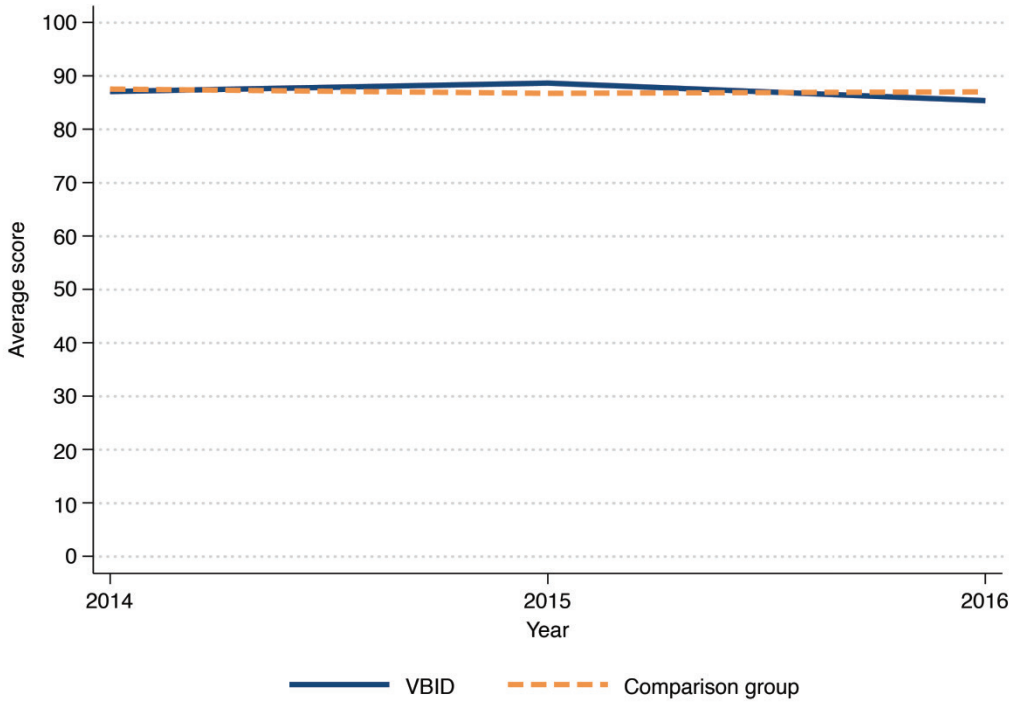
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.00.

Figure E.5. Average Score of Customer Service by VBID-Participation Status and Year, 2014–2016 (*N* = 3,917)



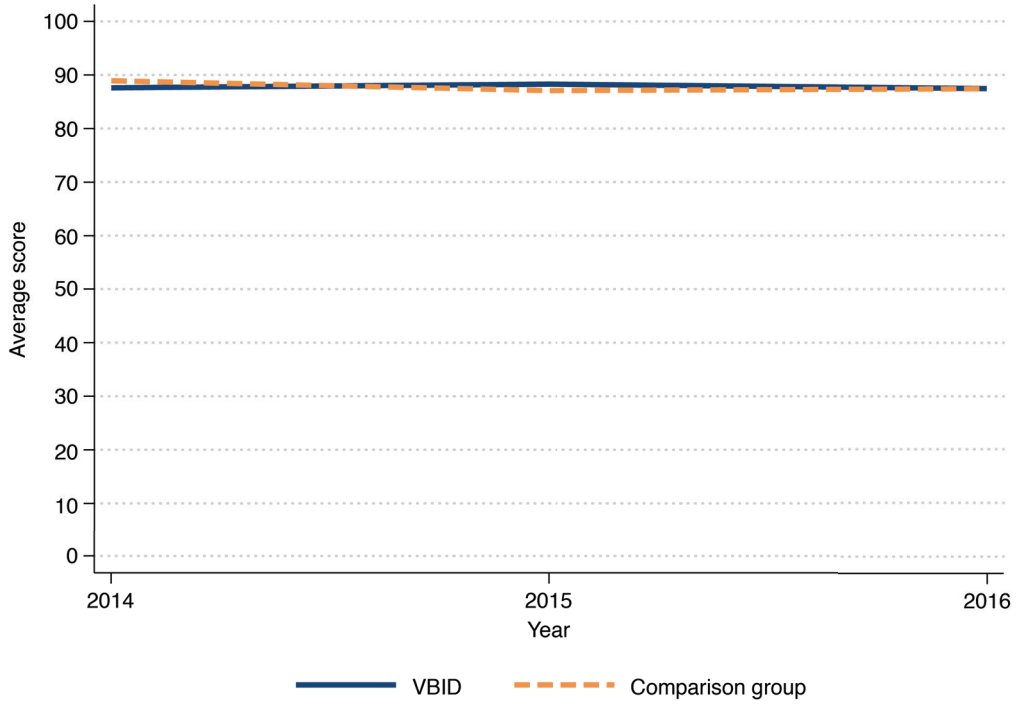
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.38.

Figure E.6. Average Score of Care Coordination by VBID-Participation Status and Year, 2014–2016 (*N* = 9,791)



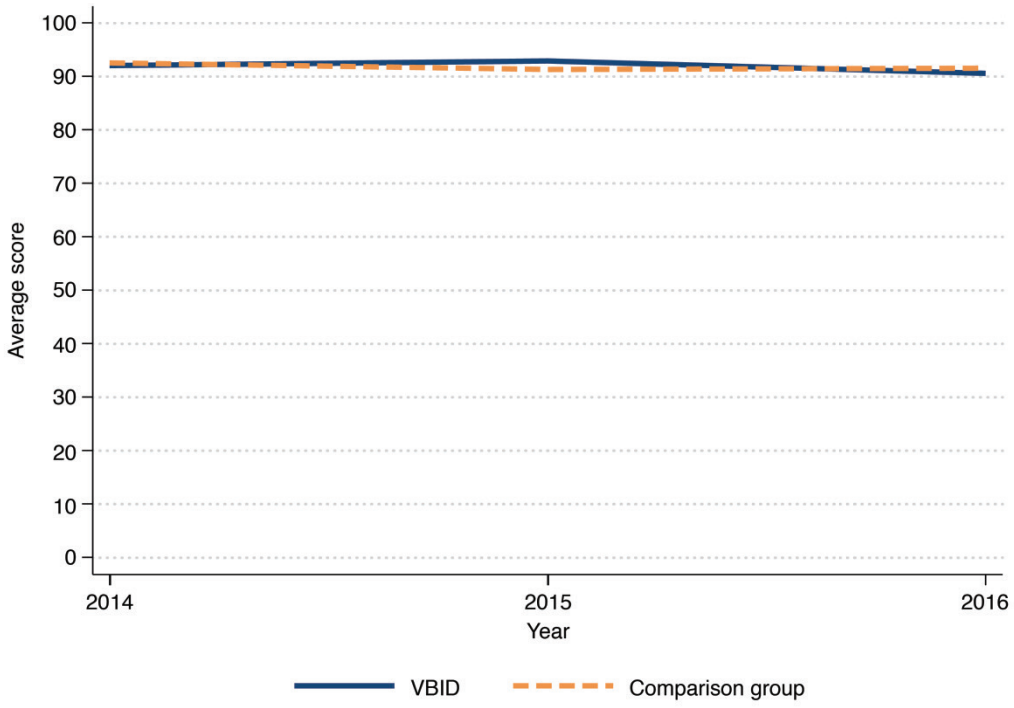
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.01.

Figure E.7. Average Rating of Health Care Quality by VBID-Participation Status and Year, 2014–2016 (*N* = 11,112)



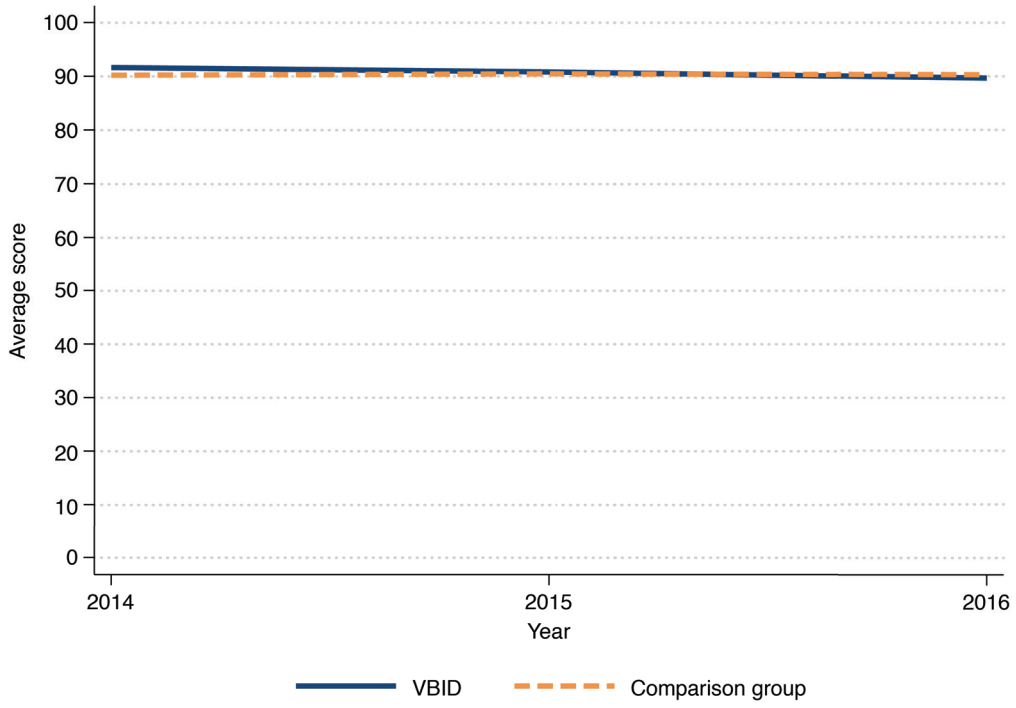
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.03.

Figure E.8. Average Rating of Doctor by VBID-Participation Status and Year, 2014–2016 (*N* = 9,373)



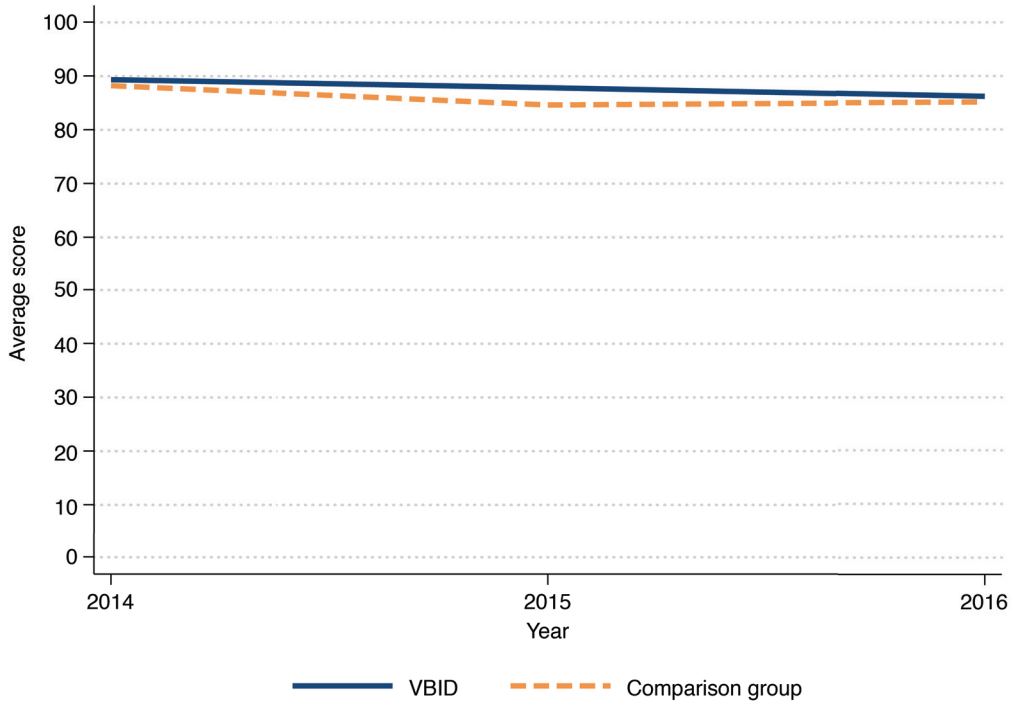
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.01.

Figure E.9. Average Rating of Specialist by VBID-Participation Status and Year, 2014–2016 (*N* = 6,622)



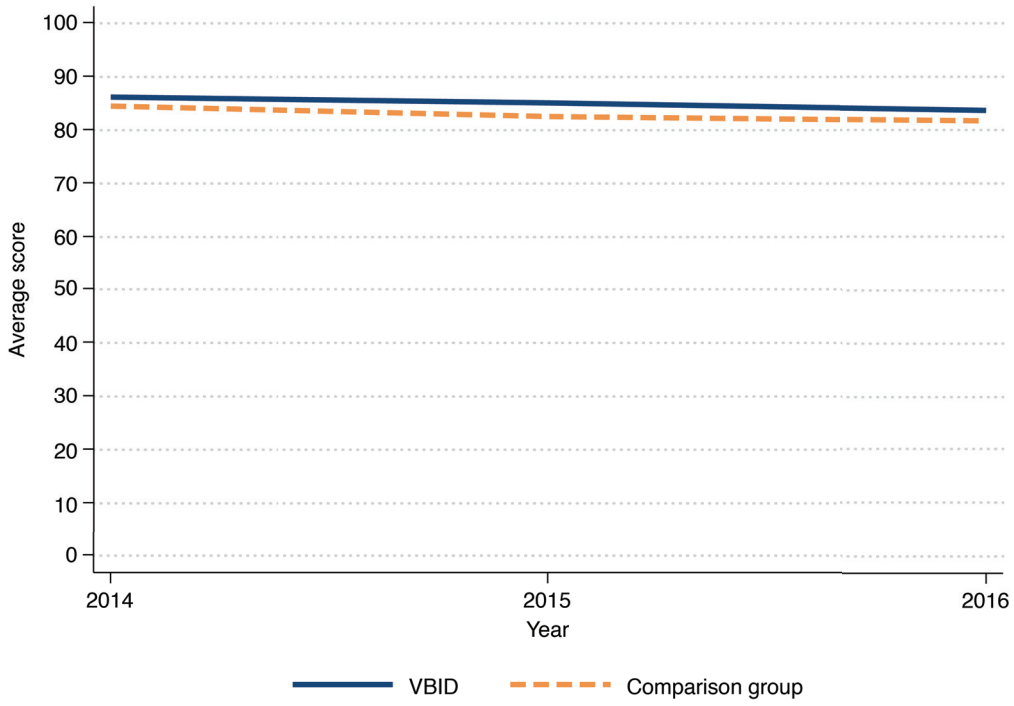
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.39.

Figure E.10. Average Rating of Health Plan by VBID-Participation Status and Year, 2014–2016 (*N* = 10,928)



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.08.

Figure E.11. Average Rating of Prescription Drug Plan by VBID-Participation Status and Year, 2014–2016 (*N* = 10,198)



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.78.

Tables E.2–E.4 provide the case-mix adjusted difference-in-differences model results for each outcome. The sample sizes range from 7,757 to 19,289, and the difference-in-differences estimates are shown in the final row of the tables.

Table E.2. Case-Mix Adjusted Difference-in-Differences Model Results for MA & PDP CAHPS Composite Measures (Part 1)

	Getting Needed Prescription Drugs	SD	Doctors Who Communicate Well	SD	Getting Needed Care	SD
Number of observations	17,696		16,557		16186	
Random effect variance	1.19		3.65		1.78	
Intercept	79.85***	0.80	80.12***	0.79	73.85***	1.05
Year						
2014	*Reference*		*Reference*		*Reference*	
2015	-1.07**	0.45	0.24	0.43	−1.25**	0.62
2016	−0.72	0.45	0.03	0.44	−1.81***	0.63
2017	−0.67	0.45	1.01**	0.45	−0.97	0.59
2018	−1.19**	0.47	0.62	0.46	−1.79***	0.61
Age	0.47***	0.09	−0.47***	0.08	−0.12	0.11
Education	−0.40***	0.11	−0.02	0.10	−0.31**	0.14
General health status	1.17***	0.16	1.53***	0.15	1.99***	0.21
Mental health status	2.15***	0.15	2.32***	0.14	2.46***	0.20
Any proxy help	0.34	0.58	−0.62	0.54	0.09	0.76
Proxy answer	1.47*	0.89	2.20***	0.83	3.99***	1.15
Medicaid dual eligible or LIS	1.47***	0.42	−0.03	0.41	−2.23***	0.57
Fielded late in 2016	−1.80	1.66	2.11	1.53	−0.42	2.28
VBID PBP	0.69	0.53	−0.79	0.62	0.20	0.70
Difference-in-Differences	**−0.29**	**0.63**	**0.96**	**0.60**	**−0.72**	**0.82**

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.
Abbreviation: LIS = low-income subsidy.

Table E.3. Case-Mix Adjusted Difference-in-Differences Model Results for MA & PDP CAHPS Composite Measures (Part 2)

	Getting Appointments and Care Quickly	SD	Customer Service	SD	Care Coordination	SD
Number of observations	17,978		7,757		17196	
Random effect variance	21.69		10.51		3.25	
Intercept	61.62***	1.26	71.73***	1.56	77.35***	0.89
Year						
2014	*Reference*		*Reference*		*Reference*	
2015	-0.07	0.64	−4.34***	0.91	−0.16	0.4
2016	−0.14	0.65	−1.12	0.93	−0.65	0.50
2017	2.32***	0.67	0.15	0.90	0.10	0.51
2018	2.55***	0.69	1.25	0.92	−0.96*	0.52
Age	0.09	0.12	−0.10	0.16	−0.55***	0.09
Education	0.75***	0.15	0.12	0.21	−0.17	0.12
General health status	1.56***	0.22	1.73***	0.30	1.63***	0.18
Mental health status	2.37***	0.21	1.94***	0.28	2.10***	0.17
Any proxy help	0.27	0.80	−1.89*	1.11	1.78***	0.63
Proxy answer	1.96	1.24	0.66	1.75	0.33	0.97
Medicaid dual eligible or LIS	−0.78	0.59	−1.10	0.82	−0.49	0.47
Fielded late in 2016	2.31	2.20	6.32*	3.77	1.73	1.81
VBID PBP	−0.73	1.19	1.60	1.20	−0.42	0.66
Difference-in-Differences		**0.89**	**−0.26**	**1.20**	**1.44****	**0.69**

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.
Abbreviation: LIS = low-income subsidy.

Table E.4. Case-Mix Adjusted Difference-in-Differences Model Results for MA & PDP CAHPS Overall Ratings

	Rating of Health Care Quality	SD	Rating of Doctor	SD	Rating of Specialist	SD	Rating of Health Plan	SD	Rating of PDP	SD
Number of observations	19,289		16,428		12,245		18988		17654	
Random effect variance	3.75		1.77		0.85		11.14		13.59	
Intercept	71.74 ***	0.77	82.91 ***	0.70	82.68 ***	0.84	75.92 ***	0.86	68.82 ***	1.04
Year										
2014	*Reference*		*Reference*		*Reference*		*Reference*		*Reference*	
2015	−0.78*	0.41	−0.62	0.39	0.01	0.48	−2.67***	0.42	−1.09**	0.52
2016	−0.52	0.42	−0.90**	0.39	−0.52	0.49	−2.94***	0.43	−2.34***	0.53
2017	0.02	0.43	−0.24	0.40	0.34	0.48	−1.68***	0.45	−1.22**	0.55
2018	0.05	0.44	−0.02	0.41	0.24	0.49	−1.12**	0.46	−0.63	0.57
Age	0.01	0.08	−0.19***	0.07	−0.22**	0.09	0.62***	0.08	1.07***	0.10
Education	−0.19*	0.10	−0.32***	0.09	−0.81***	0.12	−1.05***	0.10	−1.30***	0.12
General health status	2.58***	0.15	1.26***	0.14	1.33***	0.17	2.03***	0.15	3.23***	0.18
Mental health status	2.36***	0.14	1.97***	0.13	2.00***	0.16	1.84***	0.14	1.35***	0.17
Any proxy help	−0.22	0.53	−0.01	0.49	1.45**	0.60	0.12	0.53	−0.32	0.64
Proxy answer	1.28	0.82	0.77	0.75	−1.08	0.93	0.07	0.82	0.59	0.99
Medicaid dual eligible or LIS	−0.36	0.40	0.06	0.37	0.90*	0.48	0.87**	0.40	7.11***	0.47
Fielded late in 2016	−0.62	1.47	1.15	1.39	−0.28	1.79	4.72***	1.47	4.70**	1.85
VBID PBP	0.45	0.61	0.17	0.50	0.19	0.53	0.97	0.84	1.61	1.01
Difference-in-Differences	**−0.98***	**0.59**	**0.51**	**0.54**	**0.32**	**0.66**	**0.03**	**0.60**	**−0.36**	**0.72**

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.
Abbreviation: LIS = low-income subsidy.

Appendix F. Enrollment Analytic Results

This appendix describes the methods used for the enrollment analyses. The general approaches for the study design were described in Chapter 1, and the plan matching processes were described in Appendix D.

Data Sources

We used the enrollment table in the IDR to identify beneficiaries enrolled in each VBID-participating and matched comparison PBP in July of each year, 2014–2019. We also used data on beneficiary hierarchical condition categories (HCC) to identify beneficiaries with the chronic conditions of interest; these data were also accessed via the IDR.

Study Population

We analyze enrollment at the PBP level for all VBID-participating PBPs compared to their matched comparison PBPs.

Outcome Measures

This section describes how we constructed each of the outcome measures used for the enrollment analyses.

Plan Benefit Package-Level Enrollment

We define *PBP-level enrollment* based on enrollment reported as of July 1 during each CY. We selected July because it is a month when enrollment has generally stabilized. In a prior work (Eibner et al., 2018), we found that the correlation between July enrollment and an enrollment measure based on the number of beneficiaries enrolled at any month in each CY is 0.95. We focus on July enrollment because it allows us to include data through 2019.

PBP-level enrollment was constructed using the beneficiary-level enrollment data in the IDR. Analyses focused on VBID-participating PBPs and the matched comparison PBPs. All PBP enrollees were included, regardless of VBID eligibility.

Plan Benefit Package-Level New Enrollment

In addition, we derive *PBP-new enrollment* as the number of beneficiaries to newly enter a specific PBP and construct it using the IDR enrollment data. We define *new enrollment* as enrollment during year *t* in PBP *j* when the beneficiary was either previously enrolled in a different PBP, in FFS Medicare, or not yet eligible for Medicare. The beneficiary's status in

each year was defined based on enrollment as of July 1 in each CY. For each beneficiary, new enrollment in a PBP was defined as equal to zero if the beneficiary was enrolled in the same PBP in both years *t* and *t-1*, and one if the beneficiary was enrolled in another MA PBP, FFS Medicare, or not eligible for Medicare during year *t-1*.

Enrollment for Patients with Chronic Conditions

We identified beneficiaries with four targeted VBID chronic conditions: COPD, CHF, diabetes, and hypertension. Beneficiaries with these conditions were identified according to the following HCC codes: 108 (COPD), 80 (CHF), 15–19 (diabetes), and 88 (hypertension). Among patients with these conditions, we calculated PBP-level enrollment and new enrollment using the same metrics described above.

Analysis

We assessed unadjusted trends in July total enrollment and July new enrollment by visually and statistically testing preimplementation period differences in enrollment trends between VBID-participating and matched comparison PBPs.

Next, we used linear difference-in-differences regressions to examine the difference in enrollment trends between the VBID-participating and comparison PBPs. The regression results measure the change in enrollment for the VBID-participating PBPs before and after the 2017 implementation of VBID, relative to the nonparticipating PBPs. We included PBP fixed effects so the regression results allow for a within-PBP estimate of the effect of VBID participation on enrollment trends.

In particular, let y_{pt} be one of the enrollment outcomes for PBP p in year t, and let $VBID_p$ be an indicator that the pth PBP is a VBID-participating PBP. We estimated the following difference-in-differences regression:

$$y_{pt} = \alpha + \alpha_t + \theta_p + \gamma_t * VBID_p + \beta^T X_{pt} + \varepsilon_{pt,} \quad \text{(F.1)}$$

where

- α is the overall intercept,
- α_t is a year fixed effect (with $\alpha_{2014} = 0$) that captures the trend over time,
- θ_p is a PBP fixed effect capturing time-invariant differences between PBPs,
- γ_t is the interaction effect between time and VBID-participating PBPs (with $\gamma_t = 0$ for $t \leq 2016$) that captures the difference-in-differences estimates between participating and comparison PBPs, and
- β is the effect of the additional characteristics included in the model.

The effects of interest are the difference-in-differences coefficients (γ_t for $t \geq 2017$).

Results

Descriptive Statistics

Figures F.1 and F.2 present unadjusted July enrollment by year. These results show an approximately 8-percent increase in total and new enrollment between 2014 and 2015. However, over the 2015–2017 period, total enrollment only changes by a maximum of 1-percentage point per year. Both participating PBPs and matched comparison PBPs experienced decline in new enrollment between 2015 and 2017 (a maximum of −29 percent for participating PBPs and −23 percent for matched comparison PBPs).

Parallel Trends

Table F.1 presents the regressions that we used to examine differential trends by year. We find no statistically significant differences in divergences in trends in the preperiod between VBID-participating and nonparticipating PBPs (total July enrollment p = 0.86; new July enrollment p = 0.40). Trends within the four chronic conditions were also not statistically different, with p-values ranging from 0.33 to 0.91. The magnitude of the coefficients is also small. Thus, we do not find evidence of nonparallel trends in the preimplementation period.

Figure F.1. Unadjusted Trends in Average PBP-Level Total Enrollment

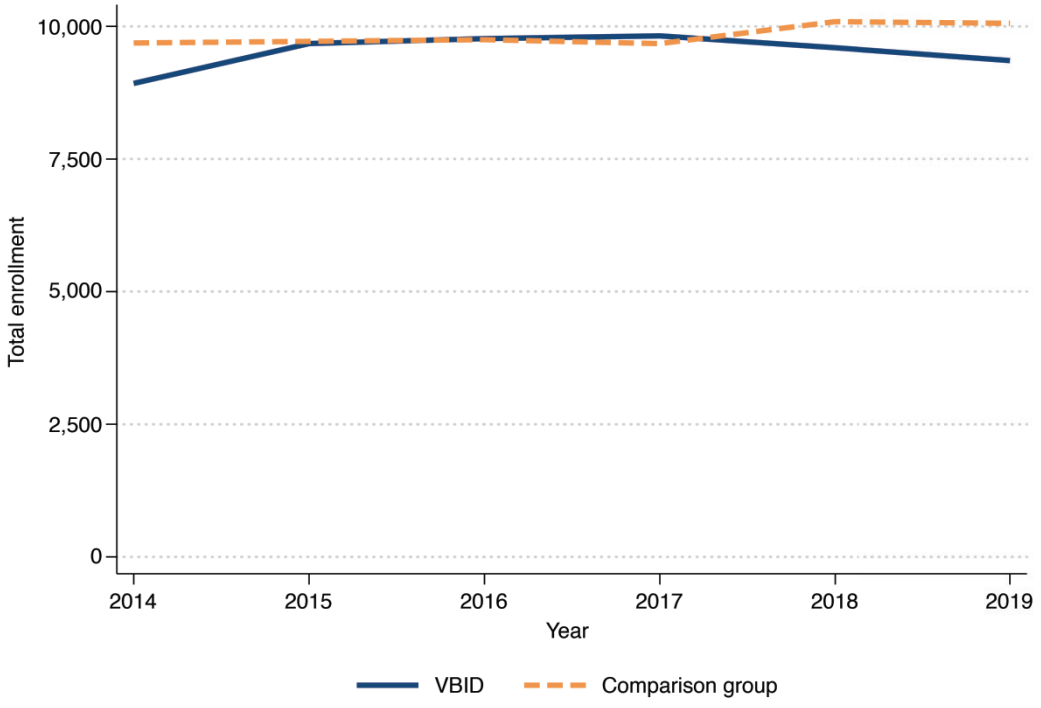


Figure F.2. Unadjusted Trends in Average PBP-Level New Enrollment

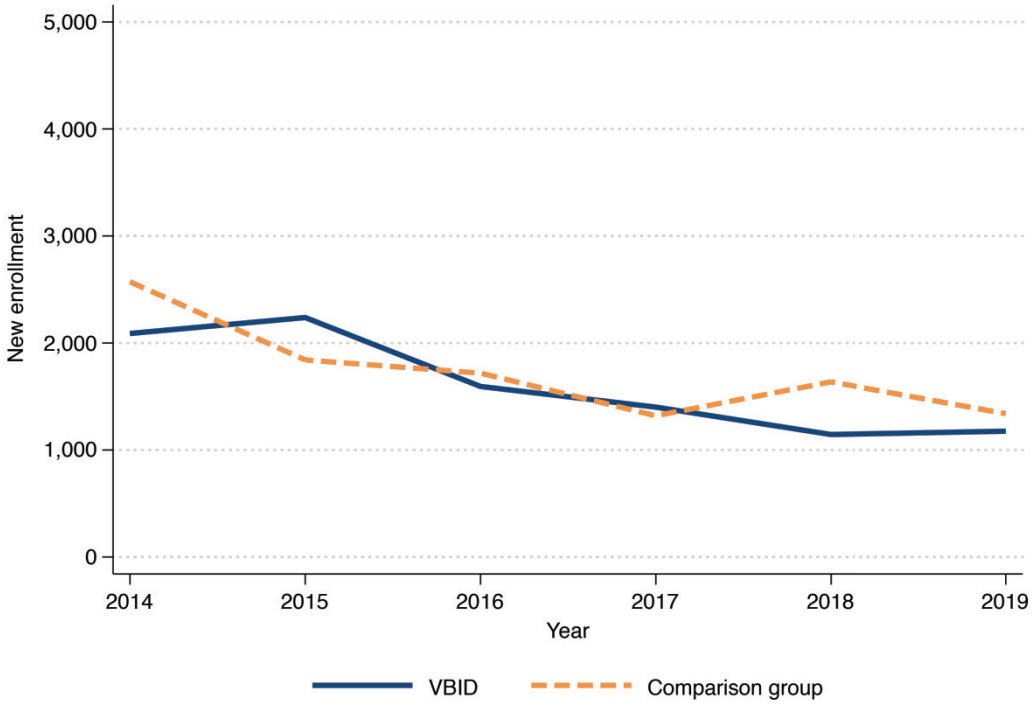


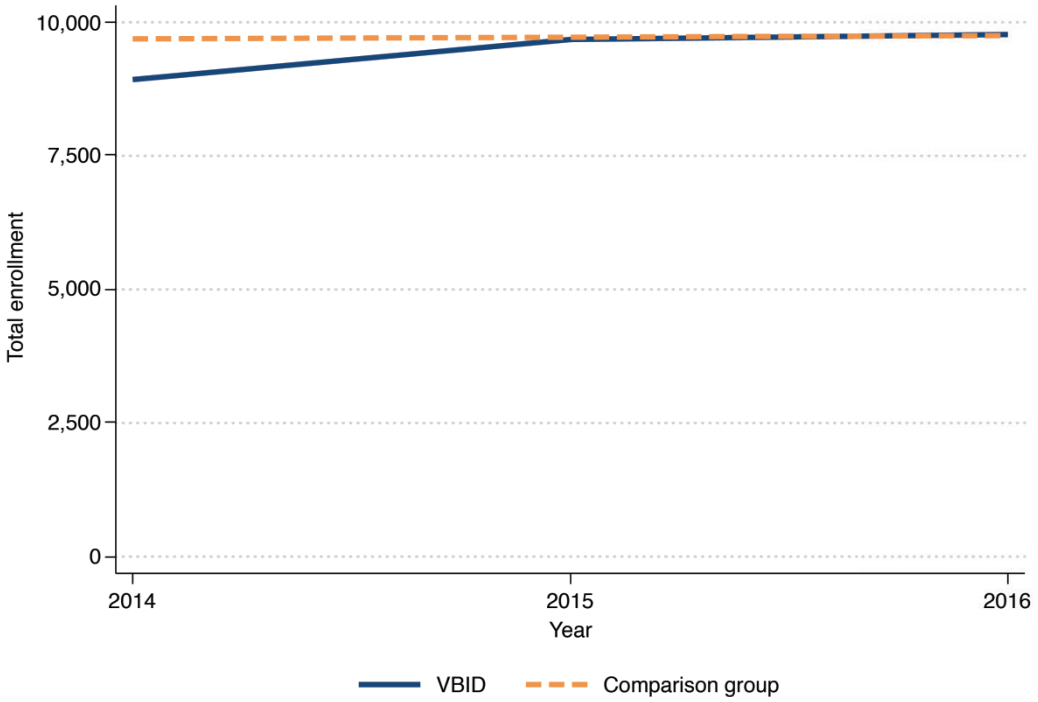
Table F.1. Regression-Adjusted Trends in Enrollment

Enrollment Measure	Overall Enrollment		COPD Enrollment		CHF Enrollment		Diabetes Enrollment		Hypertension Enrollment	
	Total	New	Total	New	Total	New	Total	New	Total	New
VBID × 2015	454	733	54	89	54	61	95	165	238	388
	(802)	(457)	(108)	(55)	(93)	(42)	(190)	(114)	(419)	(263)
VBID × 2016	680	223	267	10	69	5	139	11	354	46
	(1,215)	(665)	(451)	(81)	(131)	(58)	(291)	(163)	(647)	(366)
Observations	543	543	543	543	543	543	543	543	543	543
R-squared	0.942	0.775	0.903	0.750	0.936	0.718	0.940	0.724	0.944	0.712

NOTE: SEs are in parentheses. ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.

Visual trends support this conclusion for total enrollment (Figures F.3 and F.4). Visual inspection of new enrollment trends suggests different patterns over time (Figures F.5 and F.6); however, there is considerable variability in the means, with SEs ranging from 41 to 243. Thus, when combined with the nonsignificant statistical test results (Figures F.5 and F.6), we interpret these trends as representing random variability.

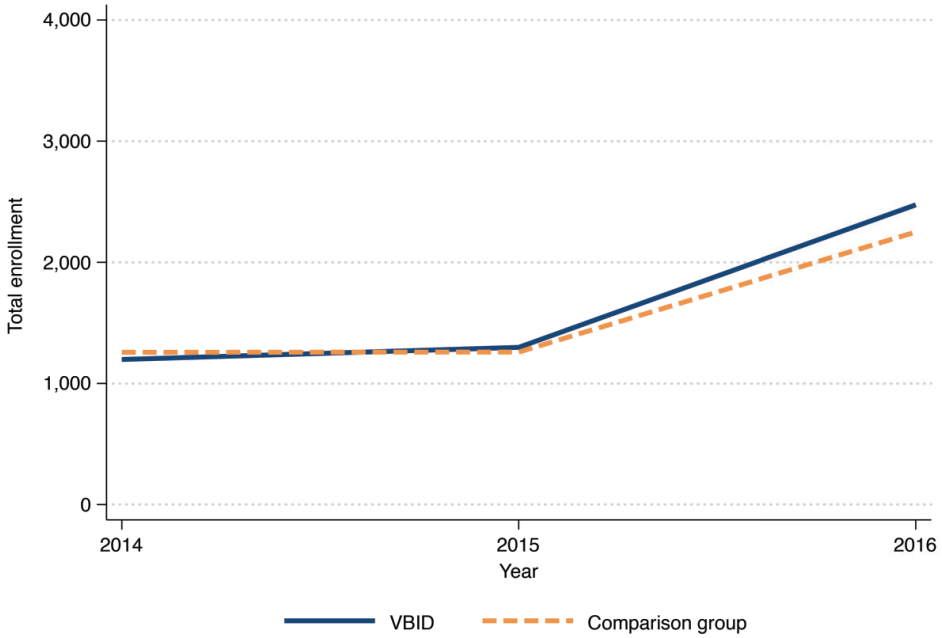
Figure F.3. PBP-Level Average of Total July Enrollment, 2014–2016



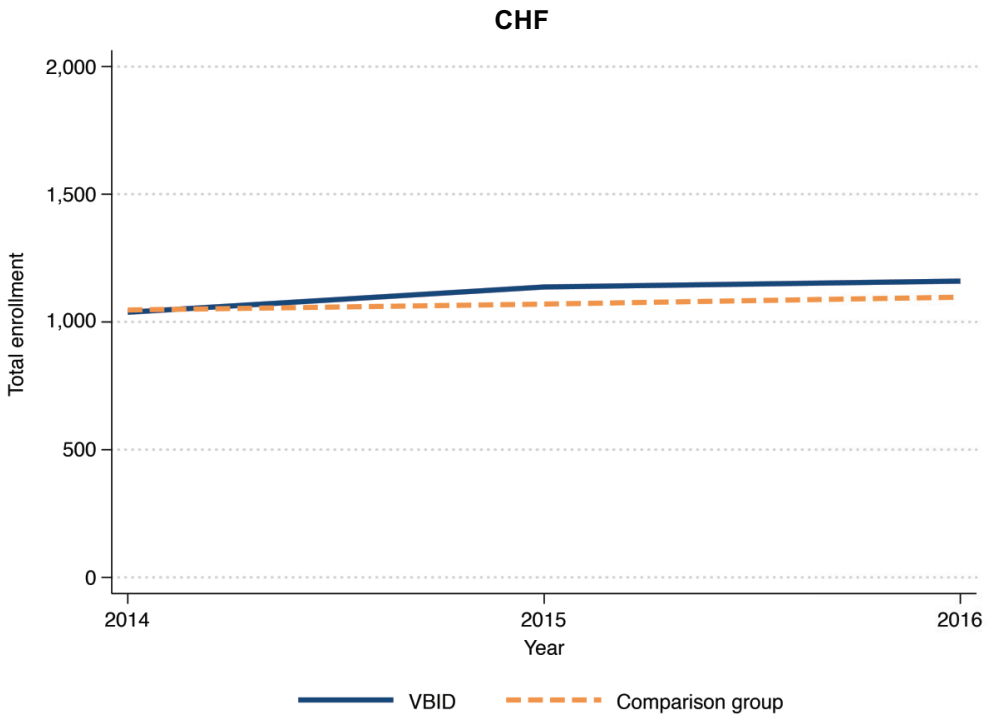
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.86.

Figure F.4. PBP-Level Average of Total July Enrollment by Chronic Condition, 2014–2016

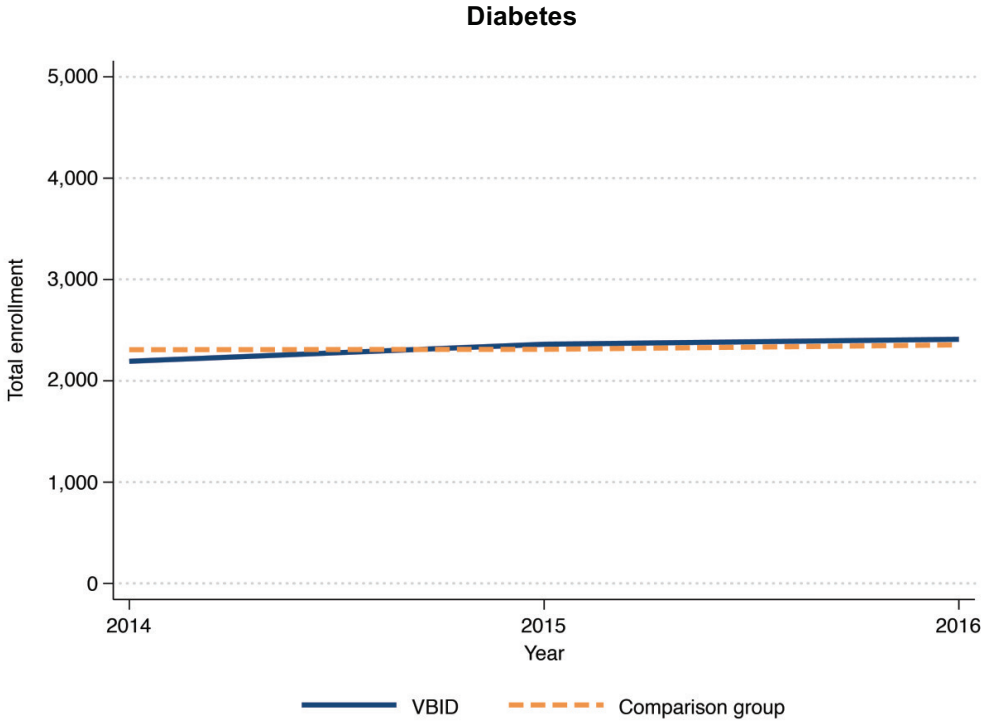
COPD



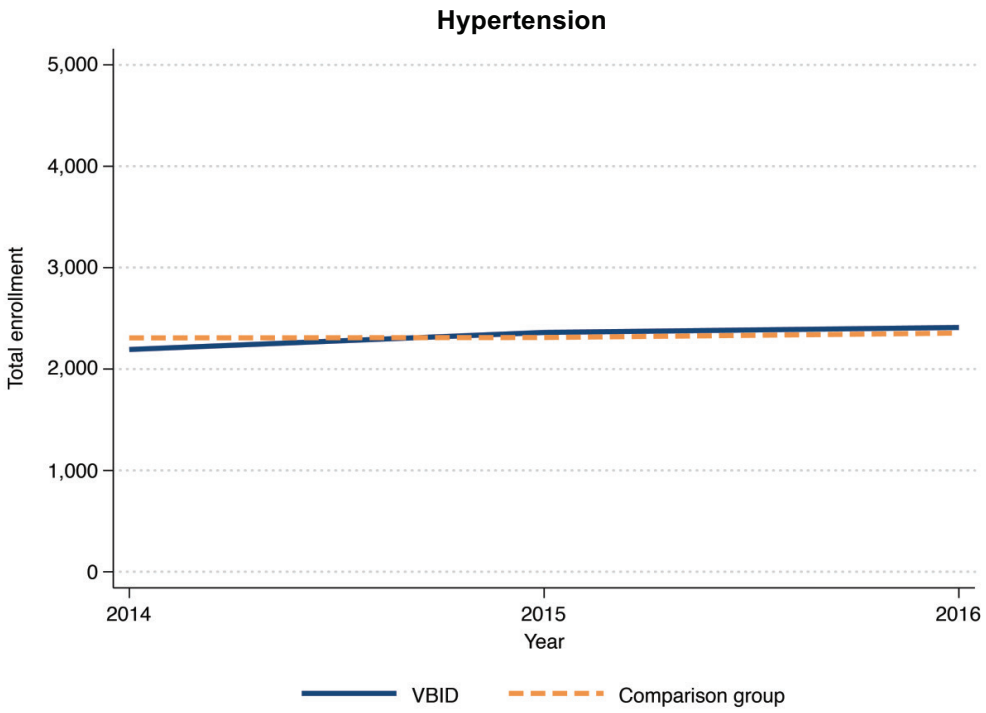
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.55.



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.88.

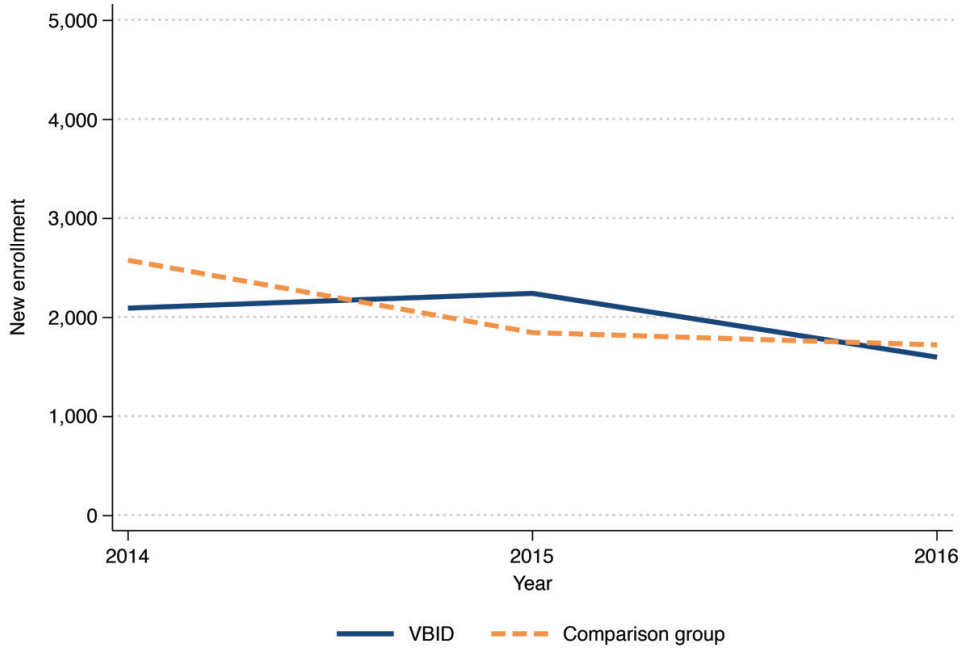


NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.91.



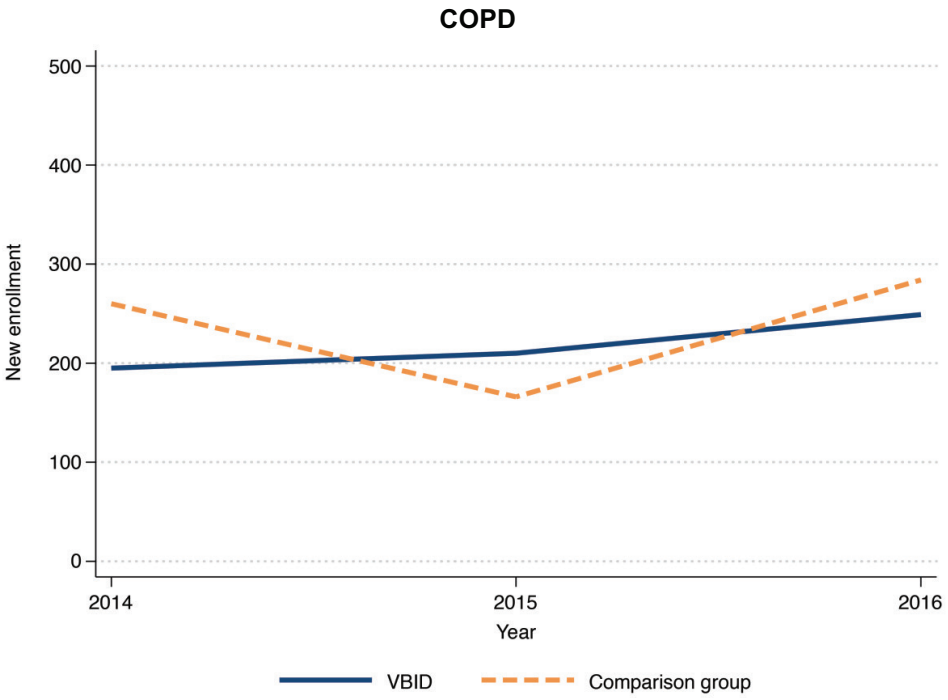
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.87.

Figure F.5. PBP-Level Average of New July Enrollment, 2014–2016

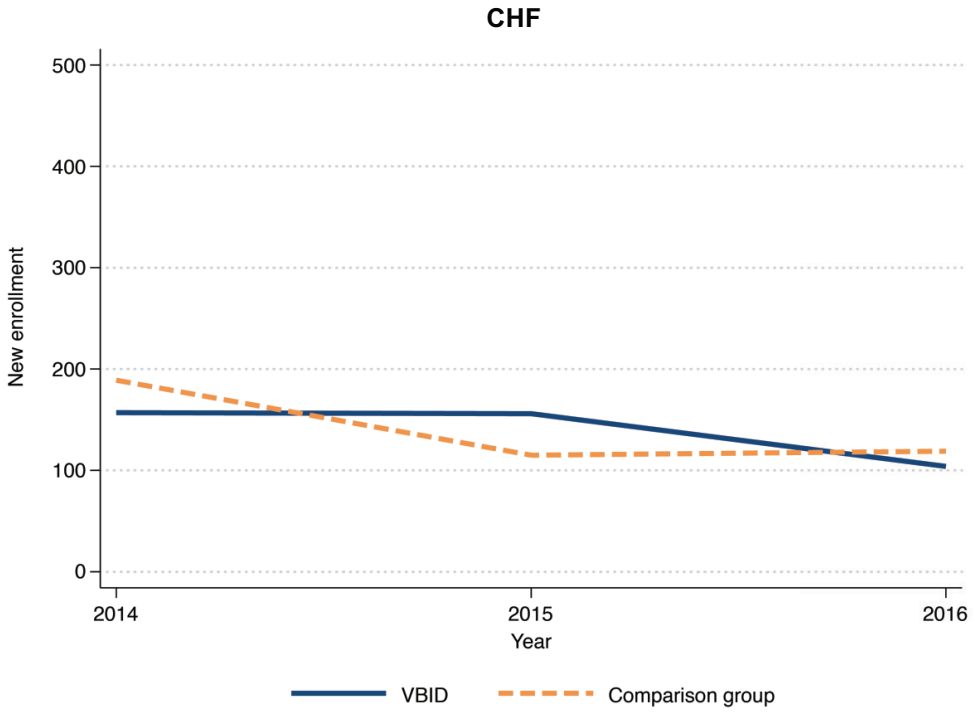


NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.40.

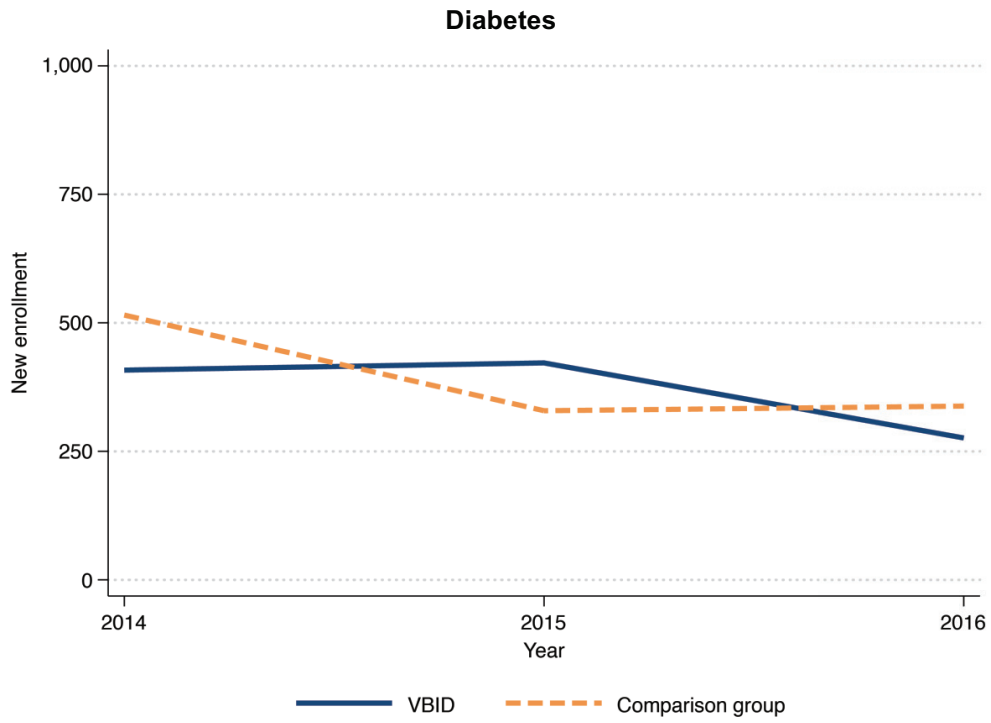
Figure F.6. PBP-Level Average of New July Enrollment by Chronic Condition, 2014–2016



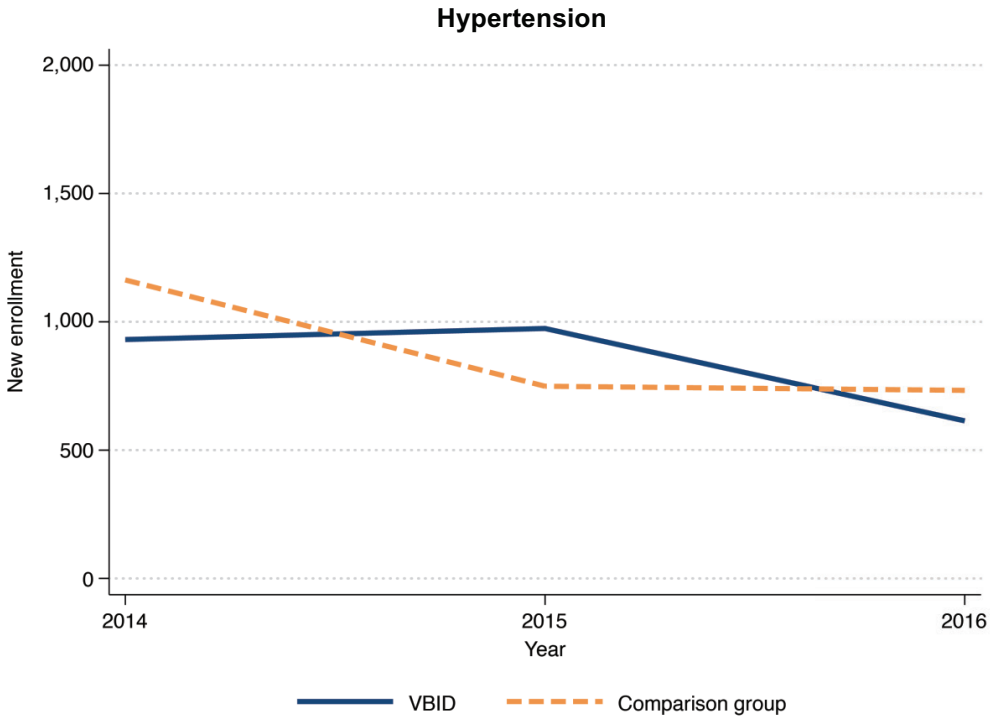
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.39.



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.36.



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.37.



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.33.

Regression-Adjusted Trends

Table F.2 presents the regression-adjusted trends in VBID enrollment. The enrollment results suggest that the implementation of VBID was not associated with changes in PBP-level enrollment, including new enrollment. Although we see a decrease in total and new enrollment in 2018 and 2019, the results are not statistically significant.

Table F.2. Regression-Adjusted Trends in July Enrollment

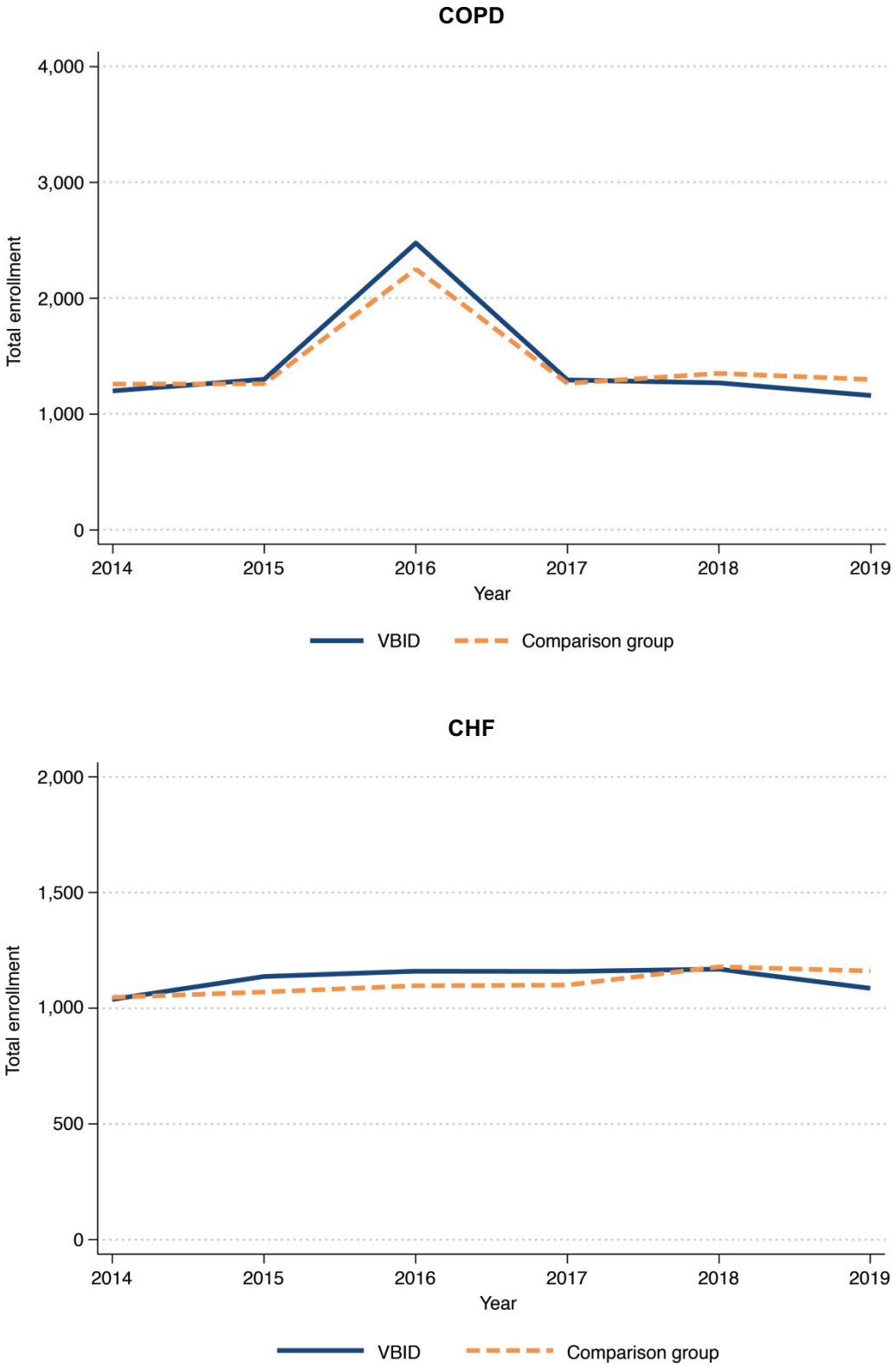
Enrollment Measure	Total Enrollment	New Enrollment
VBID × 2017	387	88
	(918)	(363)
VBID × 2018	−209	−446
	(1,438)	(363)
VBID × 2019	−400	−167
	(1,760)	(782)
Observations	543	543
R-squared	0.942	0.774

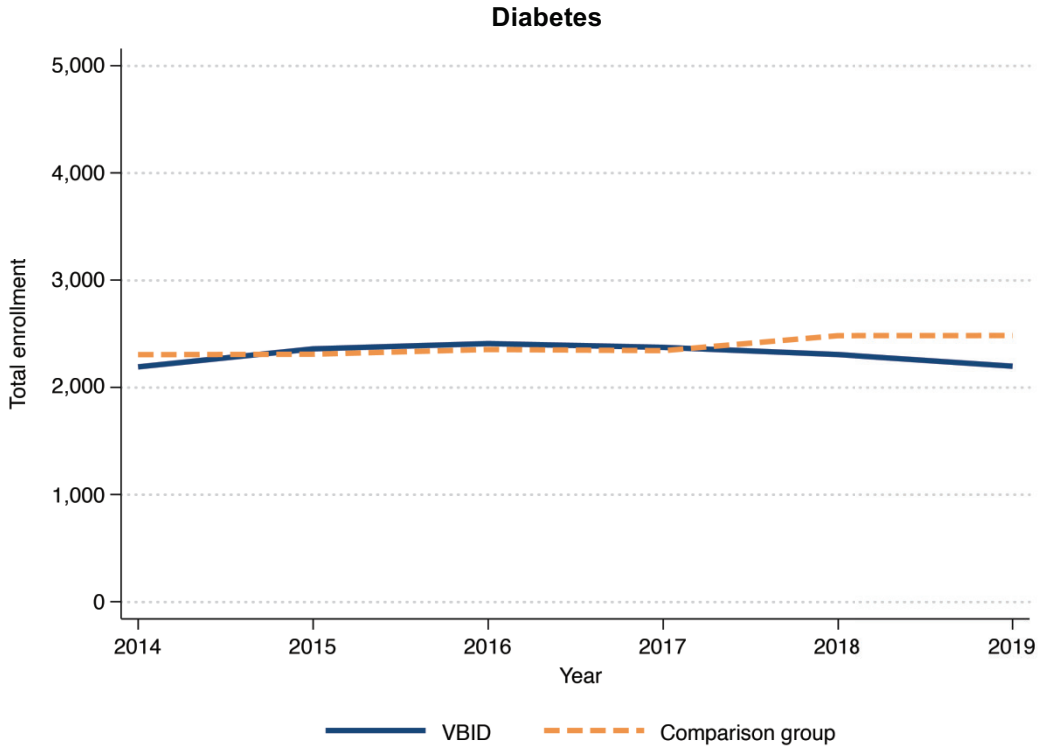
NOTE: SEs are in parentheses. ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.

Unadjusted results for chronic condition enrollment (Figures F.7 and F.8) show a large increase in enrollment for beneficiaries with COPD for both VBID (91 percent) and matched comparisons (79 percent) between 2015 and 2016. These numbers drop back down in 2017 for both beneficiary types and remains steady in 2018 and 2019. Total enrollment is fairly consistent for VBID and matched comparison beneficiaries with CHF, hypertension, or diabetes (across all six years). New enrollment for VBID beneficiaries with CHF, hypertension, or diabetes decreases between 2016 and 2018 (maximum decrease of −35 percent), with a slight increase in 2019 (15–24 percent). Matched comparison beneficiaries decrease in new enrollment between 2016 and 2017, increase again in 2018, and decrease in 2019.

Table F.3 presents the regression-adjusted trends for patients with each of the four chronic conditions. For all four chronic conditions, we do not see any differential trends in total enrollment or new enrollment between the VBID-participating and comparison PBPs.

Figure F.7. Unadjusted Trends in PBP July Total Enrollment, by Chronic Condition




Diabetes
5,000
4,000
3,000
2,000
1,000
0
Total enrollment
2014
2015
2016
2017
2018
2019
Year
VBID
Comparison group

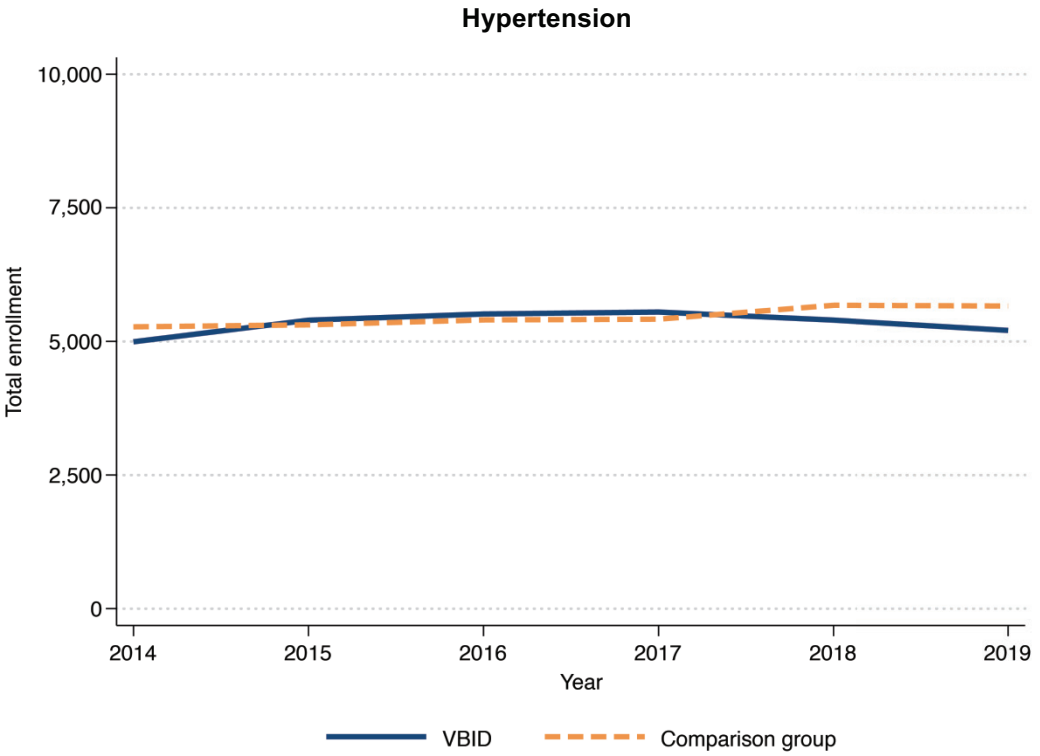
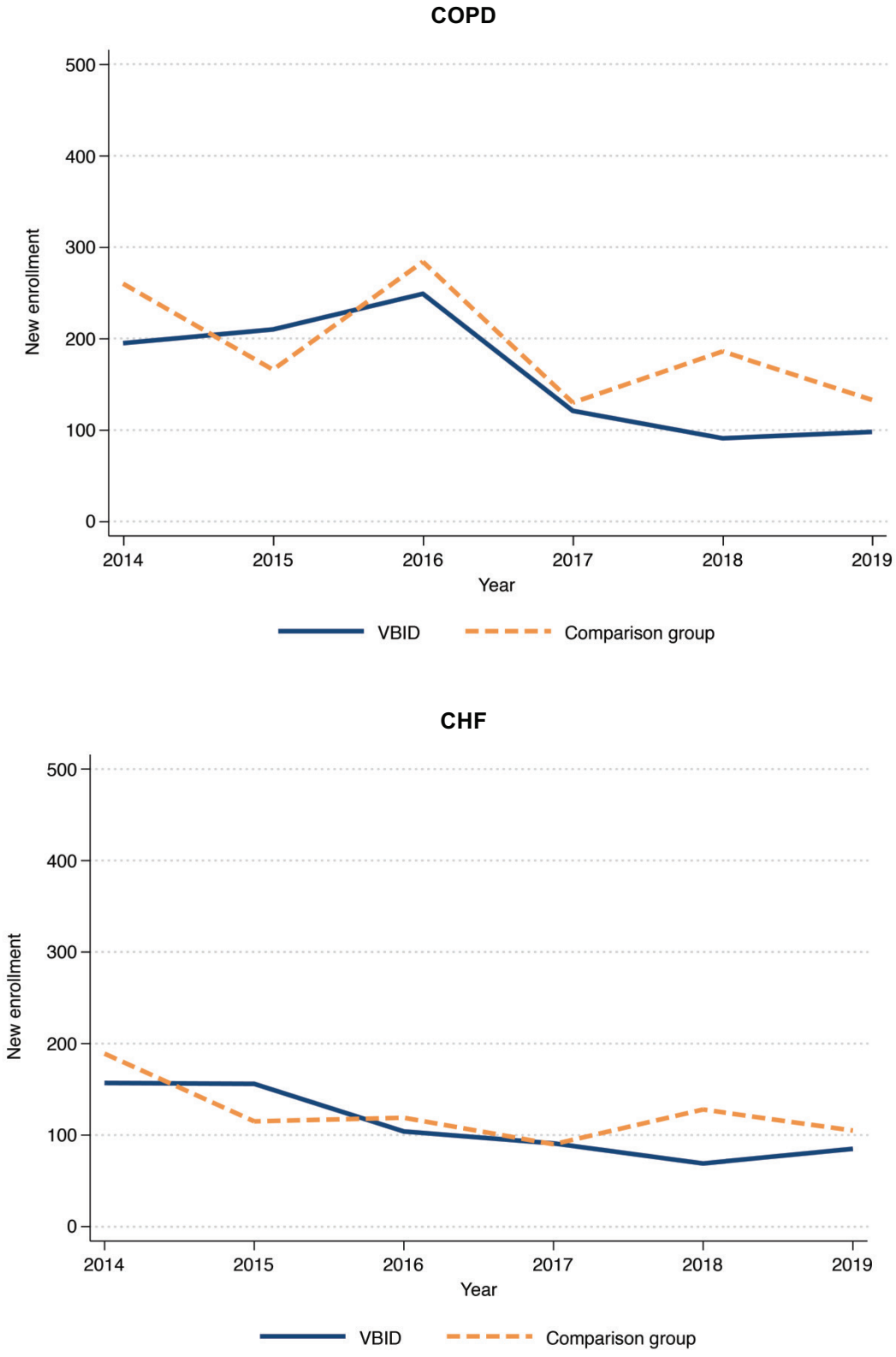
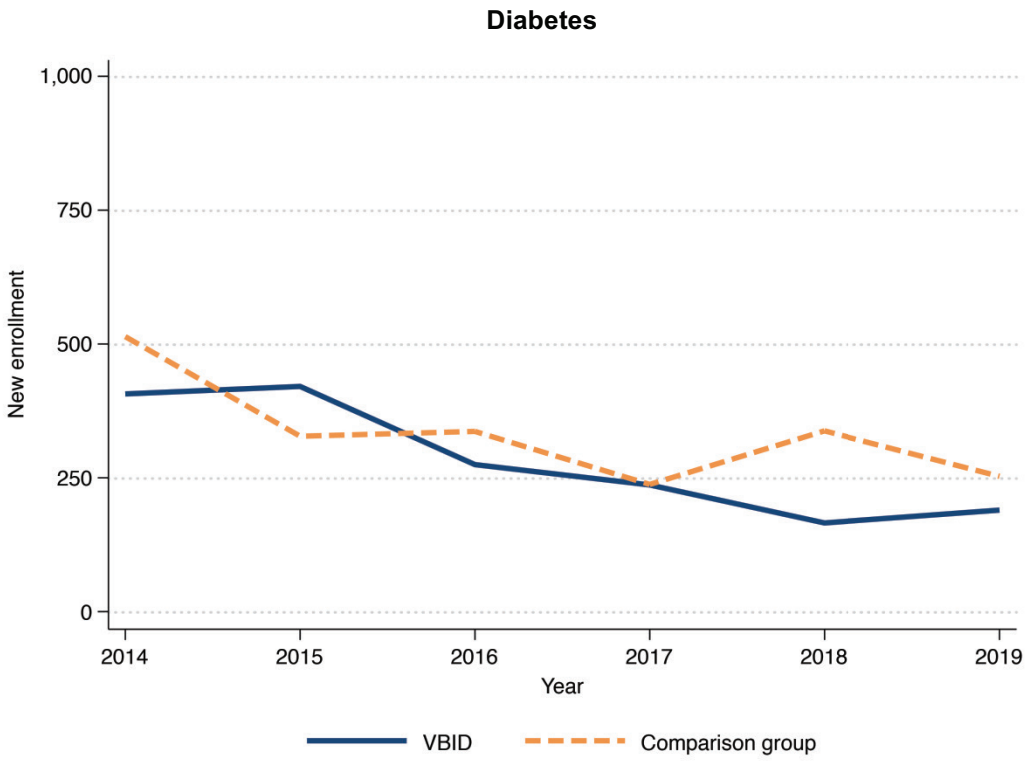

Hypertension
10,000
7,500
5,000
2,500
0
Total enrollment
2014
2015
2016
2017
2018
2019
Year
VBID
Comparison group

Figure F.8. Unadjusted Trends in PBP July New Enrollment, by Chronic Condition




Diabetes
1,000
750
500
250
0
New enrollment
2014
2015
2016
2017
2018
2019
Year
VBID
Comparison group

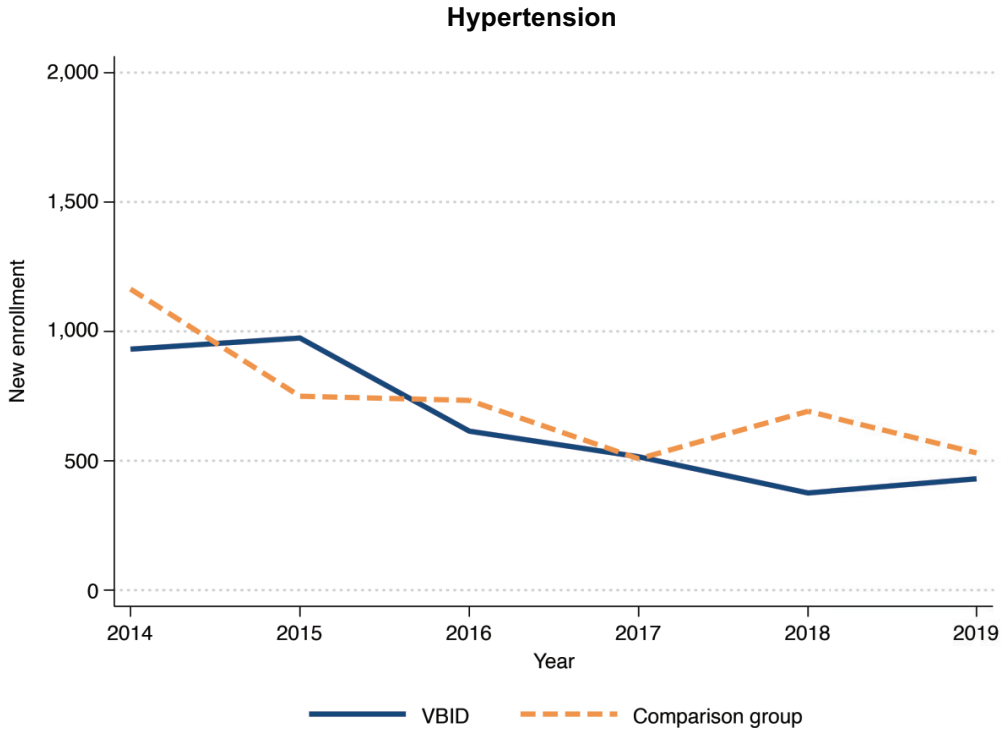

Hypertension
2,000
1,500
1,000
500
0
New enrollment
2014
2015
2016
2017
2018
2019
Year
VBID
Comparison group

Table F.3. Regression-Adjusted Trends in July Enrollment, by Chronic Condition

	COPD Enrollment		CHF Enrollment		Diabetes Enrollment		Hypertension Enrollment	
Enrollment Measure	**Total**	**New**	**Total**	**New**	**Total**	**New**	**Total**	**New**
VBID × 2017	−41	1	20	−3	27	8	157	18
	(140)	(52)	(92)	(30)	(221)	(86)	(501)	(199)
VBID × 2018	−145	−80	−46	−59	−169	−151	−226	−286
	(170)	(50)	(152)	(33)	(358)	(83)	(807)	(174)
VBID × 2019	−209	−28	−107	−24	−274	−56	−398	−96
	(276)	(94)	(217)	(63)	(463)	(173)	(991)	(378)
Observations								
R-squared								

NOTE: SEs are in parentheses. ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.

Summary

Based on our analyses of enrollment data for VBID-participating and matched comparison PBPs, we find no significant effects of VBID on any of our measures of enrollment. As described elsewhere in the report, this may be due to the inability of VBID-participating POs to market VBID to potential enrollees early in the model test.

Appendix G. Utilization Methods and Results

This appendix describes in greater detail the methods used for the utilization analysis presented in Chapter 6. The general approaches for the study design are described in Chapter 1, and the beneficiary matching process was described in Appendix D.

Data Sources

MA encounter data and Medicare Part D PDE data for 2014–2017.

Study Population

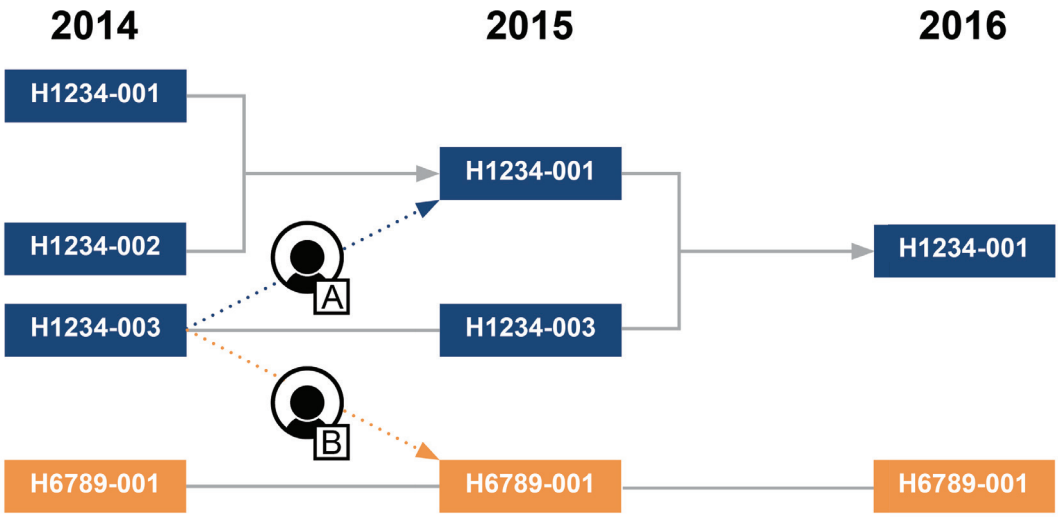
The unit of analysis is the beneficiary. The study population is the VBID-eligible beneficiaries in VBID-participating PBPs and their matched beneficiaries in matched comparison PBPs. For analyses of the VBID-targeted services, the sample consists of the VBID-eligible beneficiaries and their matched comparison beneficiaries. Beneficiaries could start to be observed throughout the study period for several reasons such as becoming eligible for Medicare or choosing to enroll in a MA plan. Beneficiaries could disappear from observation for a variety of reasons that include death, moving to FFS Medicare, or switching plans. We adjusted results for the length of time each beneficiary was observed.

Plan Switching

Tracking utilization in the "same" PBP over time for a specific beneficiary is complicated in MA. CMS allows POs offering MA PBPs to modify the contract and PBP numbers over time and move beneficiaries from one PBP to another. A common example is when one PO acquires another and then changes either the contracts or PBPs that are offered. CMS requires insurers to "cross-walk" PBPs so that administrators can follow the PBP, and beneficiaries in the PBP, over time. The PBPs can be cross-walked annually. For our VBID analyses, we treat a cross-walked PBP as the same PBP.

Figure G.1 illustrates when we consider a beneficiary as having switched PBPs and when we do not. The entity H1234-001 in 2016 represents the consolidation of several PBPs in previous years. Beneficiary A and Beneficiary B both started in H1234-003 in 2014. In 2015, Beneficiary A switched to H1234-001, whereas Beneficiary B switched to H6789-001. In this example, we do not consider Beneficiary A as having switched PBPs because they were switching between PBPs that had been cross-walked together as the same PBP in 2016. We consider Beneficiary B as having switched PBPs because that person's switch occurred between two PBPs that were not cross-walked.

Figure G.1. Example of PBP Cross-Walking and Beneficiary Switching



Outcome Measures

The utilization analyses have two categories of outcome measures. The first set of measures is the VBID-targeted services defined by each PO, and the second is the general utilization measures.

Value-Based Insurance Design-Targeted Services

The summary of each PO-specific outcome for Chapter 6 is in Table G.1. For some of the POs, we grouped several services together due to lower utilization of these services. For PO E, we grouped pulmonary rehab with oxygen therapy and sleep studies with CT chest scans to increase the utilization in both outcomes. The former occurs regularly, and the latter occur only occasionally. PO B did target some dental services, which would have generated encounter data, but there was not enough utilization of those services to analyze. The high-value PCP methodology could not be replicated in the comparison group, so high-value PCP visits are not included in this analysis. PO B also did not have enough utilization of transportation or DME services in the control group, and we analyzed the change over time in the VBID-only group for these two services. Finally, we did not include PO I in the targeted services analysis because PO I issued participating beneficiaries weight scales, blood pressure cuffs, and pulse oximeters, not all of which are tracked in the encounter data.

We did not list PO C in this table because they did not have specific benefits that we can track in the encounter data. PO C did not target any specific services as they issued rebates for any incurred Part C cost-sharing.

Each PO determined which Healthcare Common Procedure Code System/Current Procedural Terminology (HCPCS/CPT) codes to use to identify specific services eligible for reduced cost-sharing and to identify provider types such as cardiologists. As a result, the definition for

Table G.1. Summary of 2017 VBID-Targeted Benefits and the Change in Cost-Sharing Between 2016 and 2017

PO	Category	Service	2016 Cost-Sharing[a]	2017 Cost-Sharing
PO A	Laboratory tests	HbA1c test, Fasting lipid profile, urine test, Diabetic eye examination	N/A[b]	N/A
PO A	Specialist visits	Endocrinology, podiatry, ophthalmology	N/A	N/A
PO A	PCP visits	Primary care	N/A	N/A
PO B	Dental	Dental: periodontal, root planning	Not covered	$0
PO B	Supplemental	Transportation	$10 one way	$5 one way
PO B	PCP visits	Primary care	$10	$0
PO B	Specialist visits	Endocrinology, ophthalmology, nephrology, pulmonology, and podiatry (up to four)	$35	$10
PO B	DME	Diabetic supplies	20 percent	5 percent
PO D	Drugs	Selected tier 1–3 drugs	Multiple	$0
PO E	Specialist visits	Cardiology, pulmonology, palliative care, sleep specialists	$20/$35	$0
PO E	Laboratory tests	Pulmonary function test	$5	$0
PO E	Therapy/rehab	Pulmonary rehab	$10	$0
PO E	DME	Oxygen therapy treatment	20 percent	$0
PO E	Laboratory tests	Sleep studies and associated DME	$20/$35	$0
PO E	Laboratory tests	CT chest scans	$150/$225	$0
PO F	PCP visits	Primary care	Multiple	$0
PO F	Specialist visits	Cardiology and pulmonology	Multiple	$10/$20
PO G	PCP visits	Primary care	$5	$0
PO G	Specialist visits	Cardiology	$40	$0
PO G	Drugs	Selected tier 1 drug	$7	$0
PO H	Specialist visits	Endocrinology, cardiology	$40/$45	$10
PO H	Specialist visits	Podiatry	$40/$45	$5
PO I	Supplemental benefits	Weight scales, blood pressure cuffs, pulse oximeters	N/A	$0

NOTE: This table shows the change in cost-sharing for VBID-participating beneficiaries for each PO from 2016 (pre-VBID) to 2017 (first year of VBID implementation).
[a] Initial copays may depend on the PBP: Some POs entered multiple benefit designs that are indicated with "multiple" in parentheses, and others entered two benefit designs that are indicated with a slash ($X/$X).
[b] PO A offered rebates for incurred cost-sharing for specific services.

PCP visit, for example, varied across POs. These definitions are proprietary to the POs and as such, we do not reproduce these in detail here.

We asked POs to convert their provider taxonomy codes to either the Medicare or Healthcare Provider Taxonomy Code from the National Plan and Provider Enumeration System (NPPES) ("Health Care Provider Taxonomy Code Set," 2019; "Healthcare Provider Taxonomy Code Set," 2018). Some POs were not able to provide us with standard taxonomy codes, so with input from the clinician on the RAND evaluation team (Sherry), we selected taxonomy codes from NPPES to define both specialists and primary care providers. We describe this process more fully in the general utilization measures section, below.

For prescription drug utilization outcomes, we used lists of included National Drug Codes (NDCs) that were current in 2017, provided to us by the POs targeting prescription drugs in

their interventions. We identified utilization of these sets of drugs in the years prior to the implementation of VBID. NDCs within a particular therapeutic class, such as antihypertensives, can change over time, and POs can exclude certain NDCs from their formularies. Such changes and exclusions reduced the number of NDCs for these targeted services in past years of data. However, this trend was similar in the comparison group, so we do not think these processes affect our difference-in-difference results.

General Utilization Measures

The second category of utilization measures we used are general utilization measures, summarized in Table G.2. These categories allowed us to apply the same definition of primary care, for example, across participating and comparison PBPs, and also allowed us to capture effects on utilization not targeted in the specific VBID designs. We did not separately analyze hospital outpatient visits because these are captured under office visits. We also did not analyze supplemental benefits as an outcome because this would have required identifying and matching PBPs based on whether they offer the specific supplemental benefit types.

Table G.2. Summary of General Utilization Measures

Utilization Category	Definition (per Beneficiary per Month)
All-cause inpatient	Probability of at least one inpatient stay
ASC inpatient	Probability of at least one inpatient stay for ACS conditions
All-cause ED	Probability of at least one ED visit
ACS ED	Probability of at least one ED visit for ACS conditions
Office visits, primary care	Number of primary care visits
Office visits, specialists	Number of specialist visits
SNF	Probability of at least one SNF visit
Home health	Probability of at least one home health visit
Laboratory tests	Number of claims for HCPCS/CPT codes in 8000–8999 range
DME	Number of DME claims of any kind (claim type 4800)
Prescription drugs	Number of 30-day fills, all active ingredients combined

To develop these outcome measures, we used a grouping logic developed by another RAND team that recategorizes MA encounters into broad utilization categories of inpatient, outpatient, etc., using a combination of revenue center, bill type, claim type, and HCPCS codes (Mulcahy et al., 2019). Below we discuss the details of selected outcome measure categories that warrant further elaboration than Table G.2 provides.

Ambulatory Care Sensitive Conditions

There are many measures that could be used to capture whether VBID affects inpatient and ED use for conditions sensitive to whether a beneficiary has access to regular outpatient care. A challenge in selecting a measure for this evaluation is that each VBID PO selected a different condition or set of targeted conditions, meaning that many measures would not apply across the

full sample of VBID-participating POs. As such, we used the Acute Conditions Composite which is composed of three measures: bacterial pneumonia, urinary tract infection, and dehydration. This measure is part of the Agency for Healthcare Quality's Prevention Quality Indicators (AHRQ, webpage, undated).

Office Visits

Professional encounters can occur in a variety of settings and capture a range of information about the physician's services. Our goal was to count the number of unique visits per setting. Therefore, we only counted professional visits occurring in provider offices and hospital outpatient departments—encounters which were not captured in other care settings such as inpatient visits.

We subdivided professional encounters into primary care and specialist visits according to taxonomy codes, and for PCPs, HCPCS/CPT codes as well. From the full set of Health Care Provider Taxonomy codes, we identified the specialties likely to be associated with primary and specialty care, in conjunction with the RAND team's clinician. We did not use Medicare's classifications because there is a time lag in when this variable is populated in the IDR and it was not available at the time of our analysis.

Primary Care Visits

We considered primary care visits as professional encounters with an organizational or individual National Provider Identifier (NPI—unique codes to identify each provider billing Medicare) with selected taxonomy types that offer primary care, including general practice, family practice, internal medicine, geriatric medicine, pediatrics, obstetrics and gynecology, physician's assistants, and nurse practitioners in primary care specialties (Reid, Damberg, and Friedberg, 2019). The encounter must also have had at least one primary care focused HCPCS code such as an evaluation and management code. We also incorporated all encounters from federally qualified health centers and rural health centers as primary care visits (without restrictions on HCPCS).

Specialist Visits

As part of the process to identify PCPs and specialists, we identified specialists that were most likely to see patients in an office setting and as a result excluded certain specialties such as anesthesiology, pathology, and radiology. Visits to a provider with a PCP taxonomy code, but without one of the specific HCPCS codes, were grouped with specialty care visits.

Deduplicating Encounters

For all outcome measures in the general utilization category, encounters were deduplicated based on unique beneficiaries, dates of service, and service categories (e.g., inpatient or ED). For the professional encounters, we took one extra step to deduplicate based on whether the encounter was flagged as specialty or primary care. The deduplication method disregards

multiple encounters from the same NPI for the same beneficiary on the same date. In the following example, encounters 1 and 2 from the same NPI counted as one PCP visit:

- Encounter 1: NPI#1, professional office, PCP
- Encounter 2: NPI#1, professional HOPD, PCP
- Encounter 3: NPI#2, professional office, specialist.

For inpatient visits, inpatient encounters with the same billing NPI counted as one visit if they have consecutive or overlapping dates, using from and through dates. All other unique billing NPI-date combinations counted as separate visits. For example, if there were two ED encounters with consecutive days, this counted as two ED visits.

Control Variables

Although beneficiaries were matched on characteristics that include age, gender, race/ethnicity, dual eligibility, indicators for selected chronic conditions, and risk scores, we also controlled for these same characteristics in case there were any lingering differences between beneficiaries in VBID-participating versus comparison PBPs. We also included several PBP-level variables as potential controls because they may influence health care utilization in ways unrelated to VBID. These variables and their definitions are summarized in Table G.3. Some beneficiary-year observations were missing selected characteristic data: risk score, race/ethnicity, and disability status. For disability, every observation that was missing was given the same value for disability in all other years. For risk score, only a single year would be missing from each beneficiary, so

Table G.3. Summary of Included Control Variables

Characteristic	Coding
Gender	Female = 1, male = 0
Age	Continuous
Race/ethnicity	Probability that each beneficiary is a given race/ethnicity
Dual eligible	Dual = 1 if beneficiary has at least one month of dual status, zero otherwise; annual indicator
LIS	LIS eligible = 1 if beneficiary has at least one month of LIS eligibility and is NOT already coded as a dual, zero otherwise; annual indicator
Disabled	Disabled = 1 if reason for entitlement was disability, zero otherwise; annual indicator
Hierarchical condition category score	Final HCC score as constructed by CMS, continuous, using the measurement year; annual indicator
Rx HCC score	Final HCC score as constructed by CMS, continuous, using the measurement year; annual indicator
ESRD	ESRD = 1, zero otherwise. ESRD is defined as having either a reason for entitlement of ESRD or having an ESRD diagnosis code of ICD10: N18.5 and N18.6, ICD9: 585.6 and 585.6; annual indicator.
PO fixed effect	Unique PO
MA/MA-PD indicator	MAPD = 1 if PBP has Part D, zero otherwise; annual indicator
PBP premium	The MA and Part D premiums derived from the BENE_PRM_PRFL table in the IDR
PBP OOP maximum	OOP maximum for each PBP (variable name pbp_d_out_pocket_amt) (CMMI, 2018)

NOTE: MA without drug coverage and MA-PD with drug coverage.

we used within-beneficiary mean imputation. Race/ethnicity was missing in all years for some beneficiaries; therefore, we could not use a beneficiary's own data to impute the missing values. As such, we used a mean imputation using all available beneficiary data for the race/ethnicity field.

Analysis

This section describes the analysis steps for both the PO-specific outcomes and the general utilization outcomes. The unit of analysis was the beneficiary-year. Our regression approach for the utilization analyses followed the same basic model described in Appendix D. Let y_{cpit} be the outcome for beneficiary *i* in PO *c* in PBP *p* in year *t*. Then, the model is given by

$$g(E[y_{cpit}]) = \alpha + \alpha_t + \theta_c + \delta * VBID_{cp} + \gamma_t * VBID_{cp} + \beta^T X_{cpit,}$$

where

- g(x) = appropriate link function for the outcome, typically the log link for count data and the logit link for binary data
- α = overall intercept
- α_t = set of year fixed-effects (with $\alpha_{2014} = 0$)
- θ_c = PO fixed effect capturing time-invariant differences between POs
- δ = baseline difference between VBID and comparison PBPs
- γ_t = interaction effect between time and VBID-participating PBPs (with $\gamma_t = 0$ for $t \leq$ 2016) representing the difference-in-differences effect
- β = effect of the additional characteristics included in the model.

We used generalized estimating equations (GEE) that allow for a family and link function to be chosen for the specific outcomes and also adjust for correlations among beneficiaries over time. We examined the correlation structure of the outcomes over time and selected an autoregressive correlation structure of order 2 (AR2). As noted in the measures section, we included several beneficiary characteristics in the models: gender, age, low-income subsidy (LIS) status, dual eligibility, disabled; race/ethnicity; whether the beneficiary has end-stage renal disease (ESRD), risk score, plan OOP maximum, and premium in the year.

We assessed parallel trends both visually and in the same regression framework as the outcome regression equation as described in Appendix D. For some outcomes, parallel trends were not met either visually or statistically. In these cases, we weighted the comparison group using the entropy balancing approach described in Appendix D. Finally, there were two VBID-targeted services (transportation and DME for PO B) where the outcome was too rare in the comparison group, and parallel trends were not met even with weighting. In these cases, we present a model that assesses the change in utilization for the given outcome over time for the VBID group only.

Results

We now present specific results that were not included in Chapter 6.

Value-Based Insurance Design-Targeted Services

Parallel Trends

The parallel trends assumption can be assessed in two main ways: visually inspecting the parallel trends and testing whether the coefficients on the time trend are significantly different between the VBID group and comparison group. We first tested the parallel trends using an unweighted comparison group, by examining whether the trend over time between the VBID and comparison group was statistically different (at the 5-percent level) in 2016 from 2014 or 2015 (described in Appendix D). Not all outcomes met the parallel trends tests using the unweighted comparison group (columns 3–5 from the left in Table G.4). With weighting the control group, many of the parallel trends improved (columns 6–9 from the left in Table G.4). The last column shows which comparison group (weighted or unweighted) was used in the difference-in-

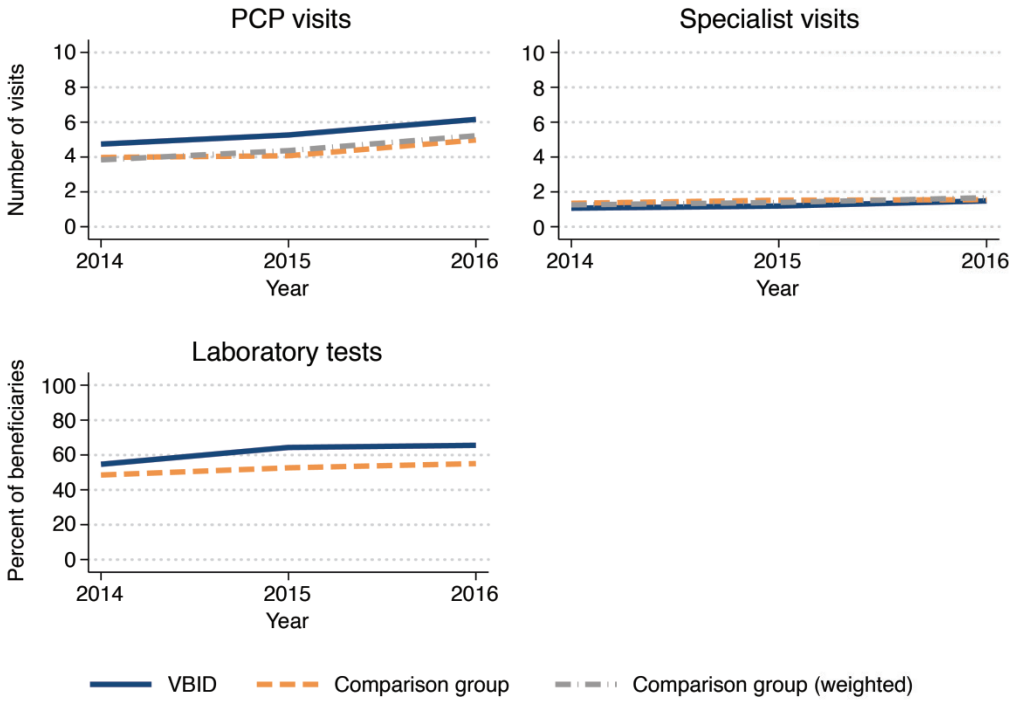
Table G.4. Summary of Parallel Trends Test, VBID-Targeted Services

		Unweighted Parallel Trends Tests			Weighted Parallel Trends Tests			
PO Name	Outcome	Chi-squared value	DF	p-Value	Chi-squared value	DF	p-Value	Model Used
PO A	Preventative services	4.806	2.000	0.090	0.125	2.000	0.939	Unweighted
PO A	PCP	8.332	2.000	0.016	1.378	2.000	0.502	Weighted
PO A	Specialists	16.130	2.000	0.000	1.854	2.000	0.396	Weighted
PO B	Number of HVP specialists	182.707	2.000	0.000	2.104	2.000	0.349	Weighted
PO B	Diabetic supplies	44.672	2.000	0.000	5.813	2.000	0.055	VBID only
PO B	Retinal photograph	1.853	2.000	0.396	1.119	2.000	0.571	Unweighted
PO B	Transportation	12.823	2.000	0.002	4.169	2.000	0.124	VBID only
PO D	Prescription drugs	18.501	2.000	0.000	18.359	2.000	0.000	Weighted
PO E	CT scans/sleep studies	4.799	2.000	0.091	1.901	2.000	0.387	Unweighted
PO E	Oxygen/pulmonary rehab/other DME	8.819	2.000	0.012	8.548	2.000	0.014	Weighted
PO E	Pulmonary function test	0.172	2.000	0.918	3.000	2.000	0.223	Unweighted
PO E	Specialists	0.616	2.000	0.735	1.519	2.000	0.468	Unweighted
PO F	PCP	62.963	2.000	0.000	0.579	2.000	0.749	Weighted
PO F	Specialists	28.687	2.000	0.000	50.614	2.000	0.000	Weighted
PO G	PCP	164.739	2.000	0.000	5.039	2.000	0.080	Weighted
PO G	Prescription drugs	0.544	2.000	0.762	0.021	2.000	0.990	Unweighted
PO G	Specialists	37.727	2.000	0.000	4.637	2.000	0.098	Weighted
PO H	Specialists	10.218	2.000	0.006	3.980	2.000	0.137	Weighted

Abbreviation: DF = degrees of freedom. HVP = high-value provider.

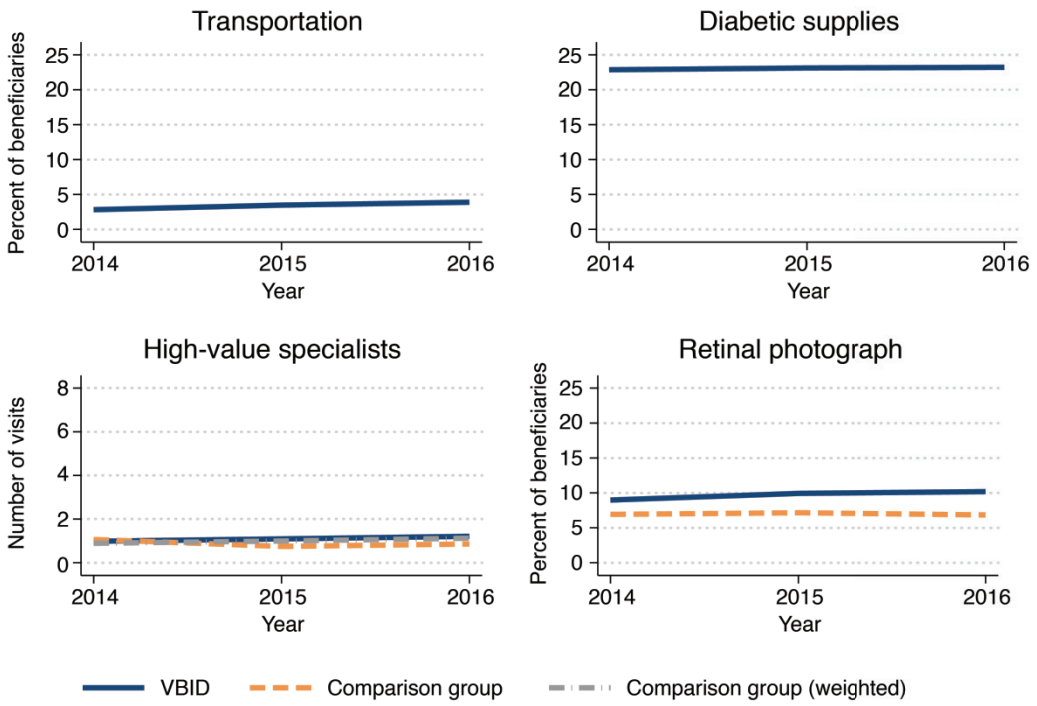
differences analysis. Figures G.2–G.8 show the average outcome value per year for the VBID group, the unweighted comparison and the weighted comparison group for outcomes in which we needed to weight the parallel trends.

Figure G.2. PO A, Average Utilization per Beneficiary per Year



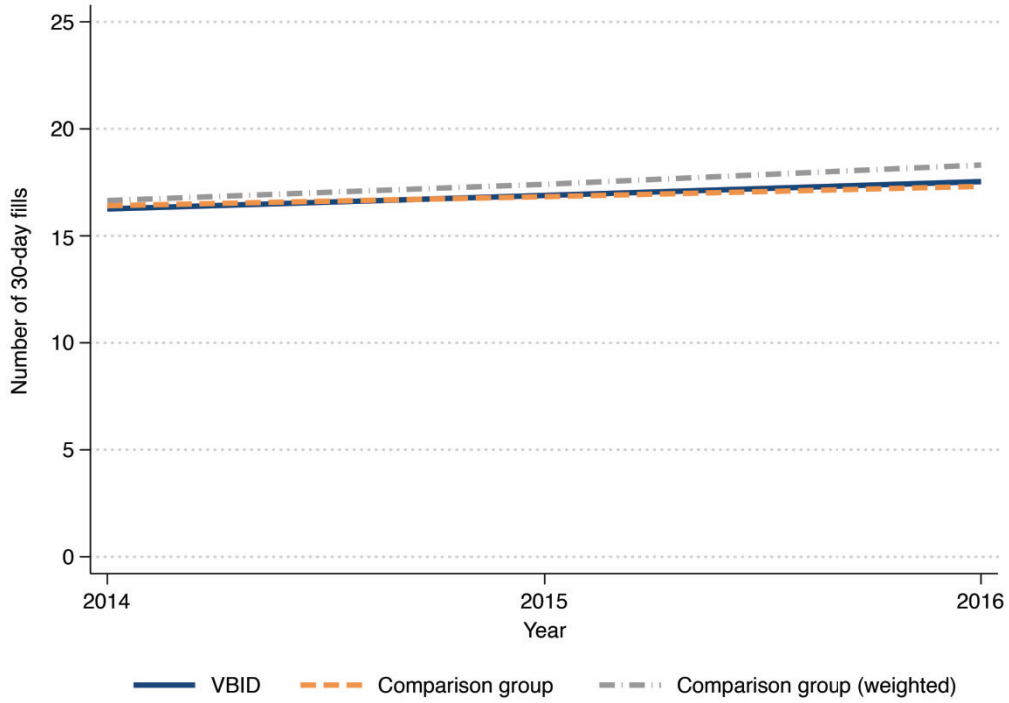
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for PCP visits was 0.02 before weighting and 0.50 after weighting. The p-value for the test of parallel trends for specialist visits was 0.00 before weighting and 0.40 after weighting. The p-value for the test of parallel trends for laboratory tests was 0.90.

Figure G.3. PO B, Average Utilization per Beneficiary per Year



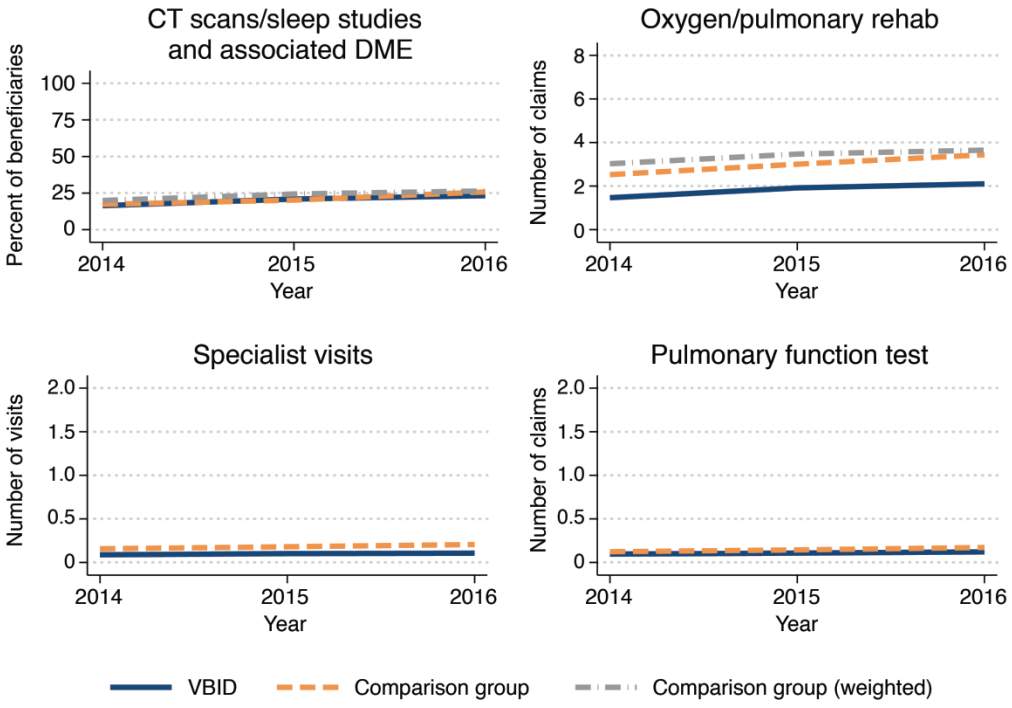
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for high-value specialist visits was 0.00 before weighting and 0.35 after weighting. The p-value for the test of parallel trends for retinal photography was 0.40. We did not test the trend for transportation or diabetic supplies because there were not enough observations in the comparison group.

Figure G.4. PO D, Average Number of 30-Day Fills per Beneficiary per Year



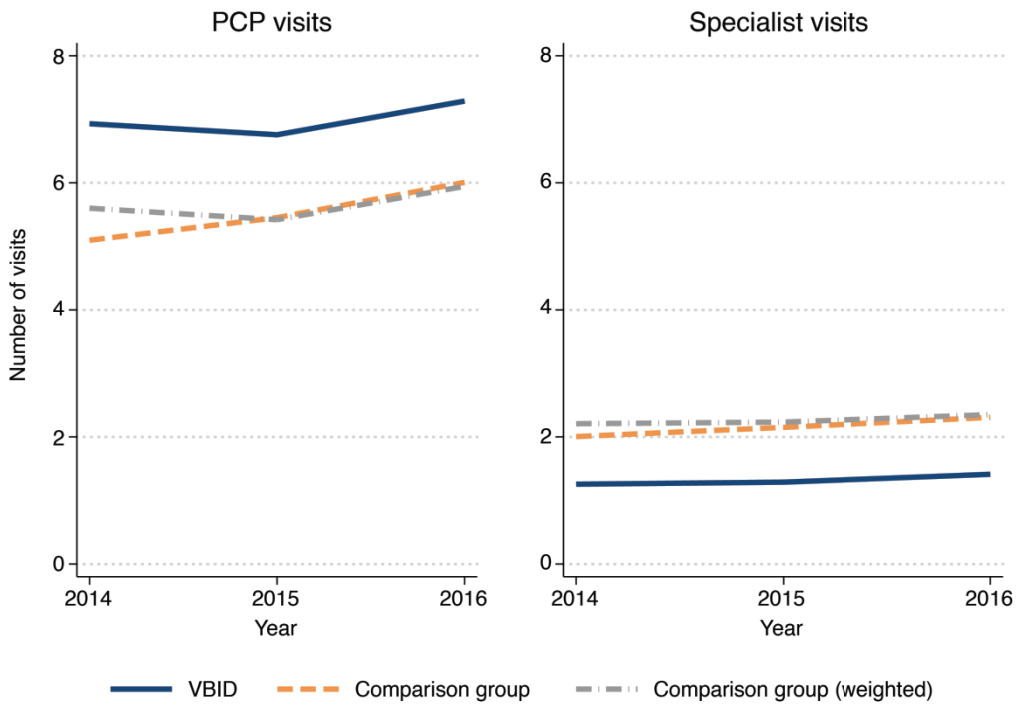
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for 30-day fills was 0.00 before weighting and 0.00 after weighting.

Figure G.5. PO E, Average Utilization per Beneficiary per Year



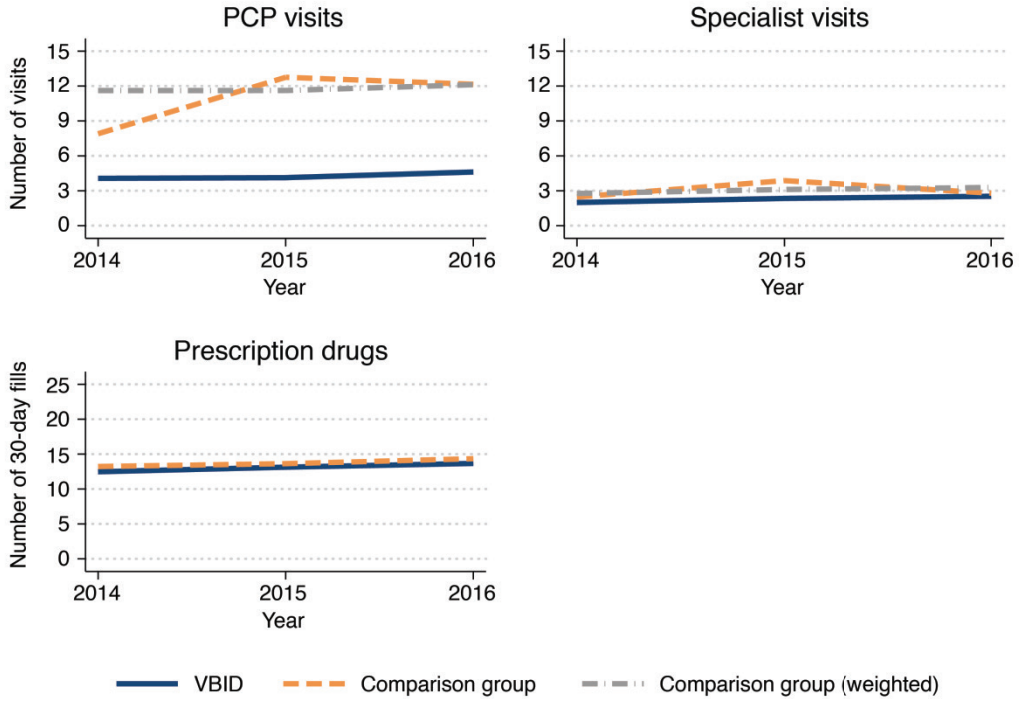
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for CT scans/sleep studies was 0.09. The p-value for the test of parallel trends for oxygen therapy/pulmonary rehabilitation was 0.01 before weighting and 0.01 after weighting. The p-value for the test of parallel trends for specialist visits was 0.74. The p-value for the test of parallel trends for pulmonary function tests was 0.92.

Figure G.6. PO F, Average Utilization per Beneficiary per Year



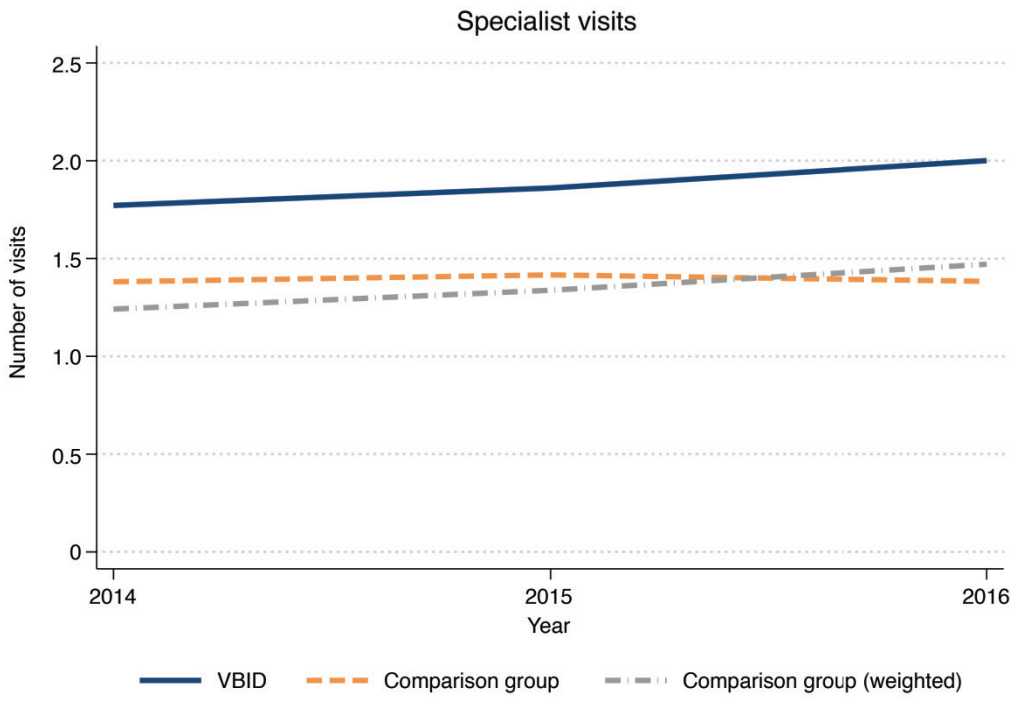
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for PCP visits was 0.00 before weighting and 0.75 after weighting. The p-value for the test of parallel trends for specialist visits was 0.00 before and after weighting.

Figure G.7. PO G, Average Utilization per Beneficiary per Year



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for PCP visits was 0.00 before weighting and 0.08 after weighting. The p-value for the test of parallel trends for specialist visits was 0.00 before and after 0.10 after weighting. The p-value for the test of parallel trends for prescription drugs was 0.76.

Figure G.8. PO H, Average Utilization per Beneficiary per Year



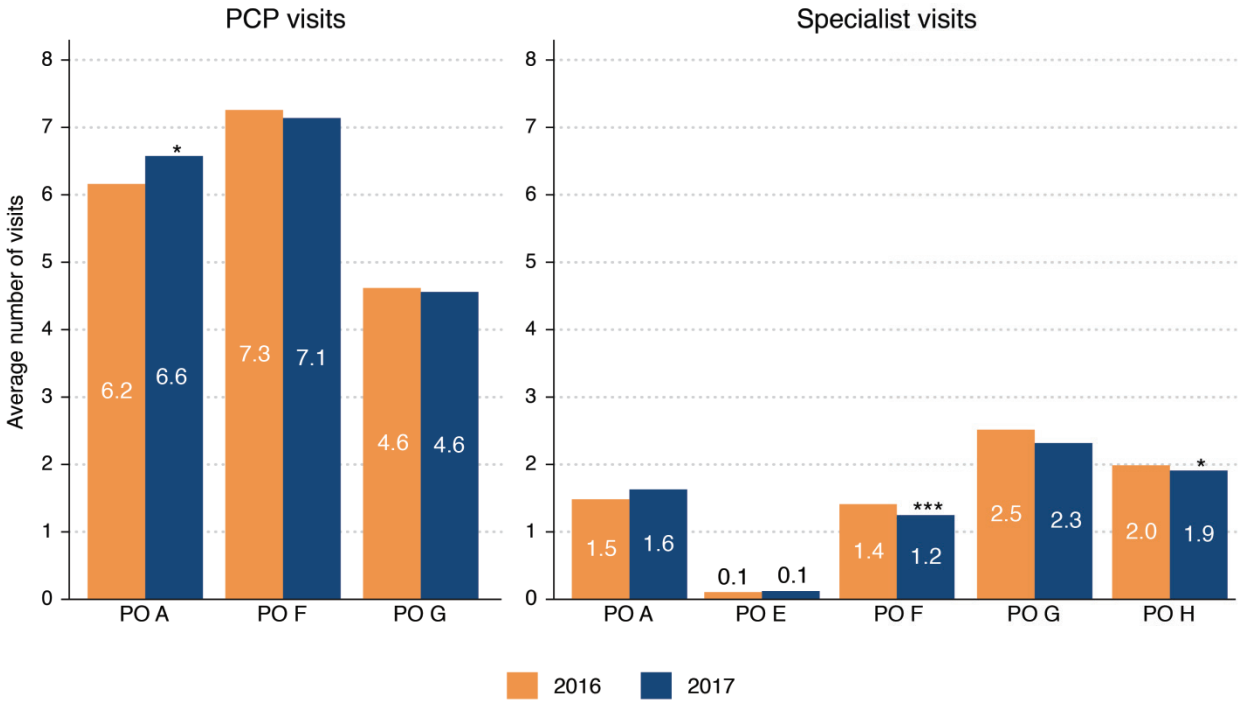
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for specialist visits was 0.01 before and 0.14 after weighting.

Descriptive Results

We begin our results by presenting descriptive changes in the use of VBID-targeted services among VBID-eligible beneficiaries between 2016 and 2017 (Figures G.9–G.13). These results are unadjusted and do not represent the causal effect of VBID on utilization. However, we provide them to show trends in service use for VBID-eligible beneficiaries. In general, the change in utilization of VBID-targeted services between 2016 and 2017 was small. In some cases, very small changes are statistically significant, a finding that may reflect large sample sizes for some POs.

Figure G.9 shows the changes in PCP and specialty care visits among POs targeting these services. There were no statistically significant changes in primary care visits among eligible beneficiaries in any of the three POs targeting these services. VBID-eligible beneficiaries in PO F experienced a small, statistically significant decline in the use of targeted specialist visits between 2016 and 2017.

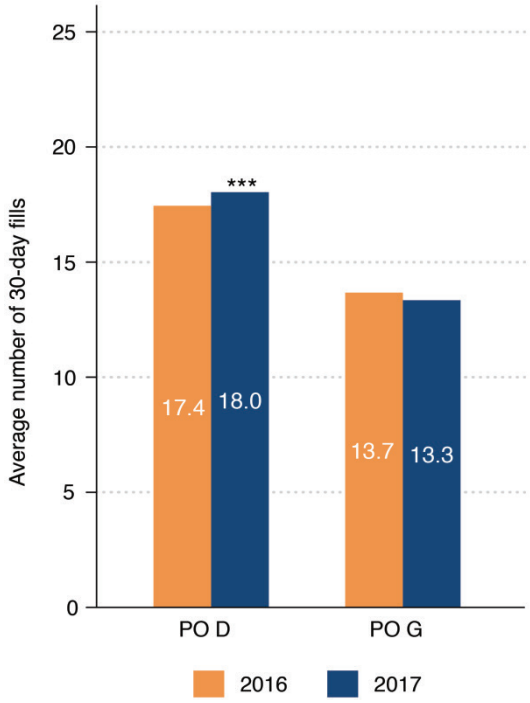
Figure G.9. Descriptive Results, Average Number of PCP, and Specialty Care Visits per Year per Eligible Beneficiary, 2016–2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample sizes are 1,142 for PO A, 1,931 for PO G, 5,224 for PO H, 23,541 for PO F, and 2,312 for PO E.

There was a slight, statistically significant increase in the utilization of targeted drugs between 2016 and 2017 for PO D, but no change in utilization for beneficiaries in PO G (Figure G.10).

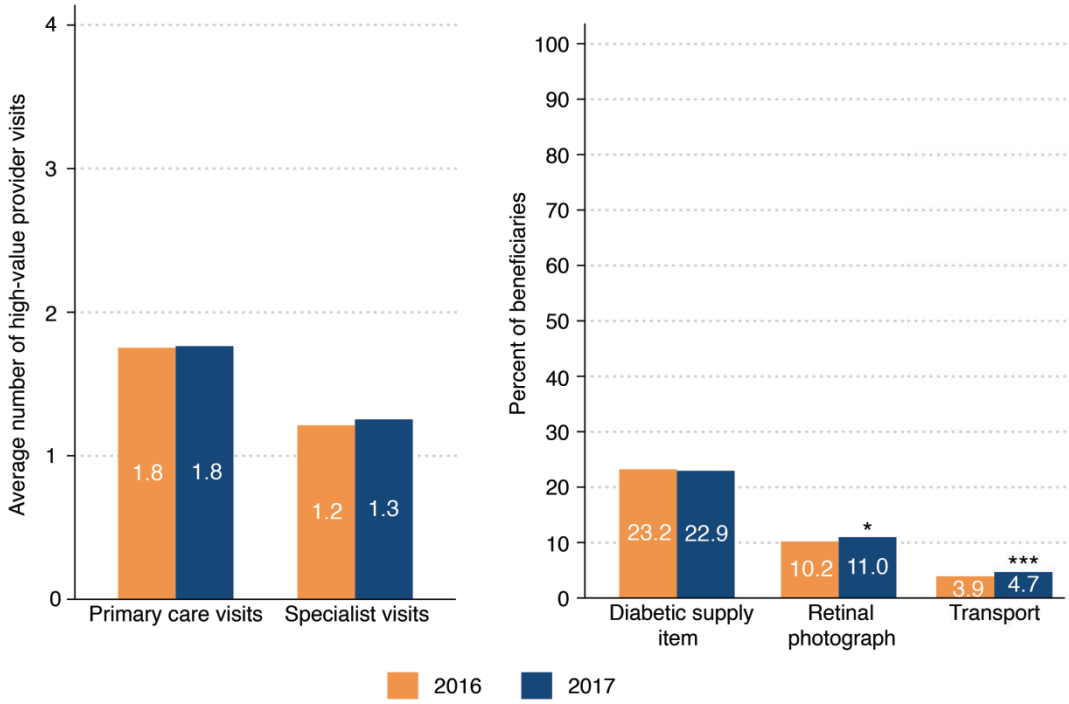
Figure G.10. Average Number of 30-Day Prescription Fills per Year per Eligible Beneficiary, 2016–2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample sizes are 23,541 for PO D and 1,931 for PO G.

PO B was the only PO to target reduced cost-sharing for beneficiaries who visited high-value providers, including high-value PCPs and high-value specialists in 2017. The methodology for how the PO identified high-value providers is described in Chapter 2. The utilization of high-value providers was unchanged from 2016 to 2017 (Figure G.11). PO B also reduced cost-sharing for several specific supplemental benefits. The proportion of beneficiaries in PO B using diabetic supplies and retinal photography was unchanged. The proportion using transportation increased by nearly 1-percentage point, but the overall proportion of eligible beneficiaries using these benefits was less than 5 percent.

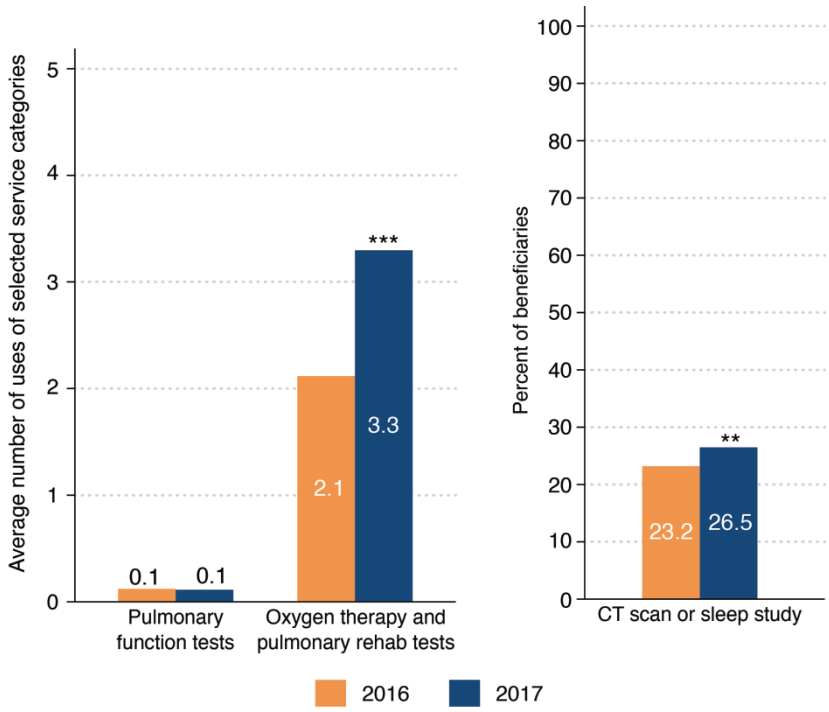
Figure G.11. Share of Eligible Beneficiaries with at Least One Encounter for Selected Service Categories, 2016–2017, PO B Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample size is 11,580 for PO B across all outcomes. The slight differences in bar heights with the same numbers in the bars are due to rounding.

PO E targeted several tests and DME associated with these tests (Figure G.12). The utilization of pulmonary function tests was unchanged from 2016 to 2017, whereas the use of oxygen and pulmonary rehabilitation services increased by 1.2 claims per beneficiary in 2017 (statistically significant at the 1-percent level). This figure also shows the proportion of beneficiaries in PO E who used at least one CT scan or sleep study, which increased 3.3-percentage points for VBID-eligible beneficiaries (statistically significant).

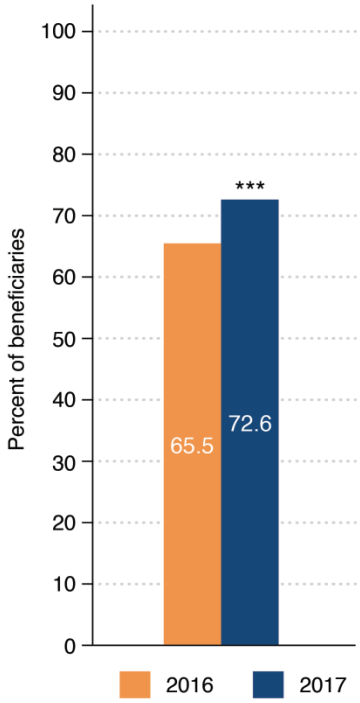
Figure G.12. Average Number of Uses of Selected Service Categories Among Eligible Beneficiaries, 2016–2017, PO E Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample size is 2,312 for PO E across all outcomes.

PO A targeted four preventive services (HbA1c test, urine analysis, foot check, and cholesterol check), and the proportion of eligible beneficiaries completing all four services increased by 7-percentage points (Figure G.13).

Figure G.13. Share of Eligible Beneficiaries with Scorecard Completion, 2016–2017, PO A Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample size is 1,931 for PO A.

Regression Results

Tables G.5–G.9 display the regression coefficients for the VBID-targeted services for the outcomes assessed with the difference-in-differences model and the outcomes analyzed with the VBID-eligible beneficiaries only. The significant changes in the outcomes were discussed in Chapter 6. Across models, most beneficiary characteristics had either no effect or no consistent directional association with the outcomes; however, beneficiary risk score is generally positively associated with the utilization outcomes.

Table G.5. Regression Coefficients, Difference-in-Difference Analysis, VBID-Targeted Services POs A, B, D, and E

Variable	PO A—PCP	SE	PO A—Specialist	SE	PO B—HVP	SE	PO D - RX	SE	PO E—Specialist	SE
2015 (2014 reference)	0.107***	0.019	0.112***	0.031	0.045**	0.020	0.053***	0.002	0.109	0.085
2016	0.214***	0.019	0.271***	0.040	0.094***	0.020	0.099***	0.003	0.143*	0.084
2017	0.254***	0.032	0.286***	0.047	−0.003	0.026	0.103***	0.004	0.036	0.095
Treatment beneficiary* year	−0.008	0.034	0.070	0.052	0.111***	0.025	0.014***	0.004	0.274**	0.122
Female	0.118***	0.024	0.148***	0.054	−0.051**	0.025	−0.044***	0.006	−0.003	0.065
Age	−0.004**	0.002	0.009**	0.004	−0.002	0.002	0.010***	0.000	−0.045***	0.005
LIS	0.142***	0.049	–0.347***	0.102	0.012	0.052	0.053***	0.013	−0.166	0.112
Dual	−0.025	0.105	0.201	0.124	−0.247***	0.063	−0.026*	0.014	0.049	0.132
Disabled	0.027	0.038	0.055	0.075	−0.007	0.036	0.029**	0.011	−0.064	0.090
Black (white, reference)	0.358	0.259	0.021	0.348	0.157	0.099	0.235***	0.023	1.068**	0.467
Hispanic	0.112	0.166	0.380	0.314	0.047	0.455	−0.060	0.037	−0.807	1.016
Asian/Pacific Islander	−0.076	0.174	0.195	0.269	−0.638***	0.223	−0.073***	0.022	0.660	0.949
American Indian/Alaskan Native	9.288	5.730	11.147	11.720	−0.682	6.460	0.372	0.457	−1.547	7.647
Multiple	−8.191	5.669	7.059	7.190	−1.417	2.193	−0.062	0.283	−7.807	6.601
ESRD	0.565***	0.180	0.128	0.142	0.376***	0.084	−0.019	0.027	0.527***	0.195
Plan OOP max, 2016	0.000	0.000	0.000**	0.000	−0.000*	0.000	−0.000***	0.000	−0.000	0.000
Plan premium, 2016	−0.000	0.000	-0.000	0.000	0.000**	0.000	0.000***	0.000	0.000	0.000
Risk score	0.260***	0.011	0.103***	0.021	0.274***	0.008	0.033***	0.002	0.274***	0.016
Constant	1.251***	0.216	−1.366***	0.390	0.652	0.575	2.195***	0.043	1.342***	0.513
PO fixed effect	Yes		Yes		Yes		Yes		Yes	
Beneficiaries, treatment + control	2,292		2,292		23,176		50,870		4,694	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs. HVP = high-value provider.

Table G.6. Regression Coefficients, Difference-in-Difference Analysis, VBID-Targeted Services, POs E and F

Variable	PO E—Pulm rehab/O_2	SE	PO E—PFT	SE	PO F—PCP	SE	PO F—Specialist	SE
2015 (2014 reference)	0.167***	0.037	0.148**	0.070	−0.047***	0.008	−0.032**	0.013
2016	0.230***	0.042	0.268***	0.067	0.002	0.009	−0.057***	0.015
2017	0.270***	0.051	0.009	0.084	−0.019*	0.011	−0.199***	0.020
Treatment beneficiary*year	0.423***	0.042	0.211**	0.100	0.011	0.011	0.058***	0.021
Female	0.117**	0.057	0.023	0.055	0.029**	0.012	−0.126***	0.020
Age	0.003	0.004	−0.040***	0.004	−0.006***	0.001	−0.004***	0.001
LIS	−0.012	0.080	−0.242**	0.095	−0.057**	0.023	−0.148***	0.049
Dual	−0.073	0.078	0.229**	0.116	−0.102***	0.027	−0.022	0.052
Disabled	0.183***	0.070	−0.321***	0.077	0.009	0.017	−0.075**	0.031
Black (white, reference)	−1.619*	0.836	1.168**	0.512	−0.136**	0.055	−0.138	0.121
Hispanic	0.426	0.775	−1.111	0.986	0.064	0.081	0.175	0.132
Asian/Pacific Islander	−2.888	2.836	0.776	0.959	−0.121*	0.070	−0.063	0.101
American Indian/Alaskan Native	−3.468	3.557	−36.511	42.300	0.172	1.132	−1.460	2.690
Multiple	8.809	5.491	−6.210	10.051	0.233	0.794	−1.157	2.037
ESRD	0.136	0.114	−0.312	0.196	0.193***	0.030	0.460***	0.048
Plan OOP max, 2016	0.000	0.000	−0.000	0.000	0.000***	0.000	−0.000	0.000
Plan premium, 2016	0.000**	0.000	0.000	0.000	−0.000	0.000	0.000***	0.000
Risk score	0.130***	0.011	0.208***	0.014	0.216***	0.003	0.341***	0.004
Constant	0.262	0.446	1.124**	0.445	1.228***	0.092	0.456***	0.161
PO fixed effect	Yes		Yes		Yes		Yes	
Beneficiaries, treatment + control	4,694		4,694		26,168		26,168	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs.
Abbreviation: PFT = pulmonary function test.

Table G.7. Regression Coefficients, Difference-in-Difference Analysis, VBID-Only Analysis, VBID-Targeted Services, POs G and H

Variable	PO G—PCP	SE	PO G—Specialist	SE	PO G—Rx	SE	PO H—Specialist	SE
2015 (2014 reference)	0.034	0.048	0.012	0.057	0.042***	0.008	0.038***	0.014
2016	0.066	0.069	0.027	0.071	0.085***	0.011	0.106***	0.018
2017	−0.054	0.060	−0.296***	0.091	0.061***	0.015	0.056**	0.024
Treatment beneficiary*year	0.140***	0.028	0.294***	0.072	0.002	0.015	0.025	0.023
Female	0.043	0.049	−0.295***	0.070	−0.029	0.021	−0.117***	0.027
Age	−0.002	0.002	−0.015***	0.005	−0.005***	0.002	−0.006***	0.002
LIS	0.010	0.055	−0.084	0.123	0.033	0.028	−0.045	0.047
Dual	−0.187*	0.101	0.036	0.151	−0.033	0.031	−0.130**	0.063
Disabled	0.167**	0.081	−0.062	0.115	−0.023	0.031	−0.046	0.042
Black (white, reference)	−0.525	0.330	0.151	0.330	−0.028	0.169	−0.115*	0.067
Hispanic	0.413	0.750	−0.344	1.493	0.018	1.006	0.026	0.134
Asian/Pacific Islander	−1.183	0.794	1.482	0.943	−0.438	0.371	−0.113	0.129
American Indian/Alaskan Native	−19.831	21.316	−18.554	29.962	−9.963	9.813	−1.664	5.712
Multiple	12.569*	7.228	−11.517	7.513	1.333	3.401	0.505	2.081
ESRD	0.174***	0.052	0.237	0.207	−0.041	0.034	0.038	0.056
Plan OOP max, 2016	−0.000***	0.000	0.000*	0.000	0.000	0.000	0.000***	0.000
Plan premium, 2016	0.000	0.000	0.000	0.000	0.000	0.000	0.000***	0.000
Risk score	0.156***	0.015	0.192***	0.020	0.013***	0.004	0.130***	0.005
Constant	2.929***	0.679	−0.179	1.141	2.187***	0.574	0.082	0.191
PO fixed effect	Yes		Yes		Yes		Yes	
Beneficiaries, treatment + control	3,902		3,902		3,902		10,642	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs.

Table G.8. Regression Coefficients, VBID-Targeted Services, Logit Models

Variable	PO A - Scorecard	SE	PO B - Retinal Photo	SE	PO E - CT scans/Sleep studies	SE
2015 (2014, reference)	0.294***	0.056	0.060*	0.036	0.027***	0.008
2016	0.384***	0.057	0.063*	0.037	0.045***	0.008
2017	0.437***	0.075	0.150***	0.048	0.044***	0.011
Treatment beneficiary*year	0.318***	0.092	−0.014	0.047	0.031**	0.013
Female	−0.043	0.060	0.080**	0.038	−0.009	0.008
Age	−0.021***	0.005	0.009***	0.003	−0.006***	0.001
LIS	−0.040	0.135	−0.088	0.073	−0.032***	0.012
Dual	−0.039	0.163	−0.008	0.093	0.003	0.016
Disabled	0.004	0.088	−0.170***	0.061	−0.050***	0.011
Black (white, reference)	1.249**	0.635	−0.034	0.180	0.207*	0.112
Hispanic	−0.151	0.412	−1.000	1.130	−0.018	0.071
Asian/Pacific Islander	−0.212	0.258	−0.136	0.359	−0.008	0.131
American Indian/Alaskan Native	−0.194	16.087	21.022**	9.735	0.622	0.611
Multiple	−14.779	10.102	−7.885*	4.324	−1.137	0.942
ESRD	−0.051	0.261	0.089	0.137	0.089**	0.036
Plan OOP max, 2016	−0.000*	0.000	−0.000	0.000	−0.000	0.000
Plan premium, 2016	0.000	0.000	0.000*	0.000	0.000	0.000
Risk score	0.044	0.029	0.027**	0.013	0.103***	0.003
Observed, 10–11 mo (all year, reference)	−0.572***	0.221	−0.086	0.104	−0.001	0.024
Observed, 7–9 mo	−0.754***	0.189	−0.310***	0.111	0.006	0.022
Observed, 4–6 mo	−1.872***	0.268	−1.108***	0.176	−0.073***	0.022
Observed, 1–3 mo	−2.574***	0.356	−2.038***	0.289	−0.119***	0.020
Constant	2.210***	0.482	−2.898***	1.105	0.513***	0.063
PO fixed effect	Yes		Yes		Yes	
	2,292		50,870		4,694	

NOTE: The logit models include the offset as a categorical variable in the regression. ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs. The CT scan outcome for PO E was run as a linear probability model due to convergence issues with the logit model. The “scorecard” is the proportion of beneficiaries completing all four diabetes-related preventative services.

Table G.9. Regression Coefficients, VBID-Only Analysis, VBID-Targeted Services, PO B Only

Variable	PO B - DME	SE	PO B - Transportation	SE
Year	0.015	0.012	−0.020	0.043
Treatment beneficiary*year	−0.041*	0.022	0.091	0.088
Female	0.168***	0.041	0.351***	0.074
Age	−0.012***	0.003	0.067***	0.005
LIS	0.143**	0.065	0.025	0.135
Dual	−0.195**	0.076	0.782***	0.163
Disabled	−0.073	0.060	0.436***	0.101
Black (white, reference)	0.424**	0.171	−0.125	0.375
Hispanic	0.430	0.917	−2.028	2.090
Asian/Pacific Islander	0.558	0.357	−0.616	0.807
American Indian/Alaskan Native	4.678	9.553	−19.216	44.334
Multiple	−4.535	3.801	−5.519	10.559
ESRD	−0.024	0.117	1.616***	0.200
Plan OOP max, 2016	0.000	0.000	−0.001***	0.000
Plan premium, 2016	−0.000	0.000	0.001	0.001
Risk score	0.089***	0.010	0.715***	0.022
Observed, 10–11 mo (all year, reference)	−0.207***	0.074	0.572***	0.173
Observed, 7–9 mo	−0.152**	0.075	0.280	0.190
Observed, 4–6 mo	−0.665***	0.094	0.248	0.242
Observed, 1–3 mo	−1.222***	0.156	−1.349***	0.501
Constant	−31.163	23.482	36.501	86.471
PO fixed effect	No		No	
	11,588		11,588	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs.

General Utilization

We now turn to discussing the results for the general utilization services.

Parallel Trends

The statistical tests for whether parallel trends in the preperiod failed for some outcomes using the unweighted comparison group (Table G.10). For these outcomes, we then weighted the comparison group. The last column shows which comparison group (weighted or unweighted) was used in the difference-in-differences analysis.

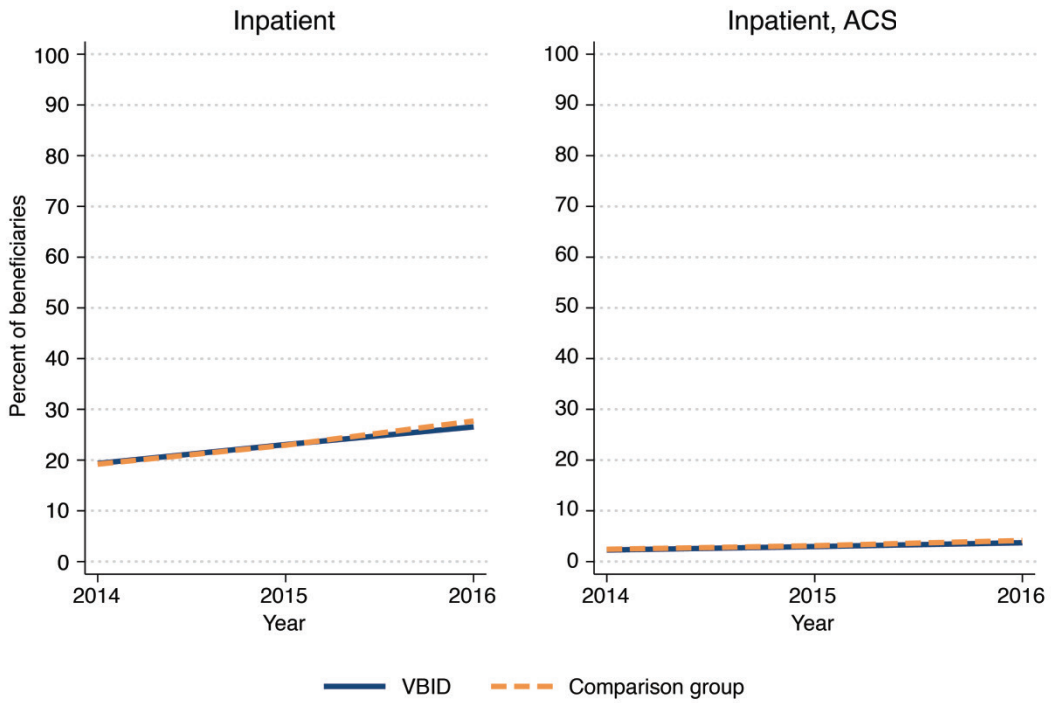
The visual inspection of the trends of beneficiaries with at least one inpatient stay showed minor evidence of nonparallel trends for the comparison group in ACS inpatient stays (Figure G.14) that became more parallel on weighting the comparison group. The trends for ED visits looked visually parallel (Figure G.15).

Table G.10. Summary of Parallel Trends Tests Results

Outcome	Chi-Squared Value	DF	p-Value	Chi-Squared Value	DF	p-Value	Model Used
Inpatient, all-cause	3.559	2.000	0.169	22.901	2.000	0.000	Unweighted
Inpatient, ACS	1.364	2.000	0.506	8.212	2.000	0.016	Unweighted
ED, all-cause	1.090	2.000	0.580	16.586	2.000	0.000	Unweighted
ED, ACS	0.016	2.000	0.992	3.922	2.000	0.141	Unweighted
Office visits, primary care	42.099	2.000	0.000	5.247	2.000	0.073	Weighted
Office visits, specialists	325.611	2.000	0.000	52.866	2.000	0.000	Weighted
SNF	29.187	2.000	0.000	8.696	2.000	0.013	Weighted
Home health	16.518	2.000	0.000	6.133	2.000	0.047	Weighted
Laboratory tests	96.973	2.000	0.000	22.614	2.000	0.000	Weighted
DME	112.526	2.000	0.000	11.696	2.000	0.003	Weighted
Prescription drug	106.254	2.000	0.000	36.212	2.000	0.000	Weighted

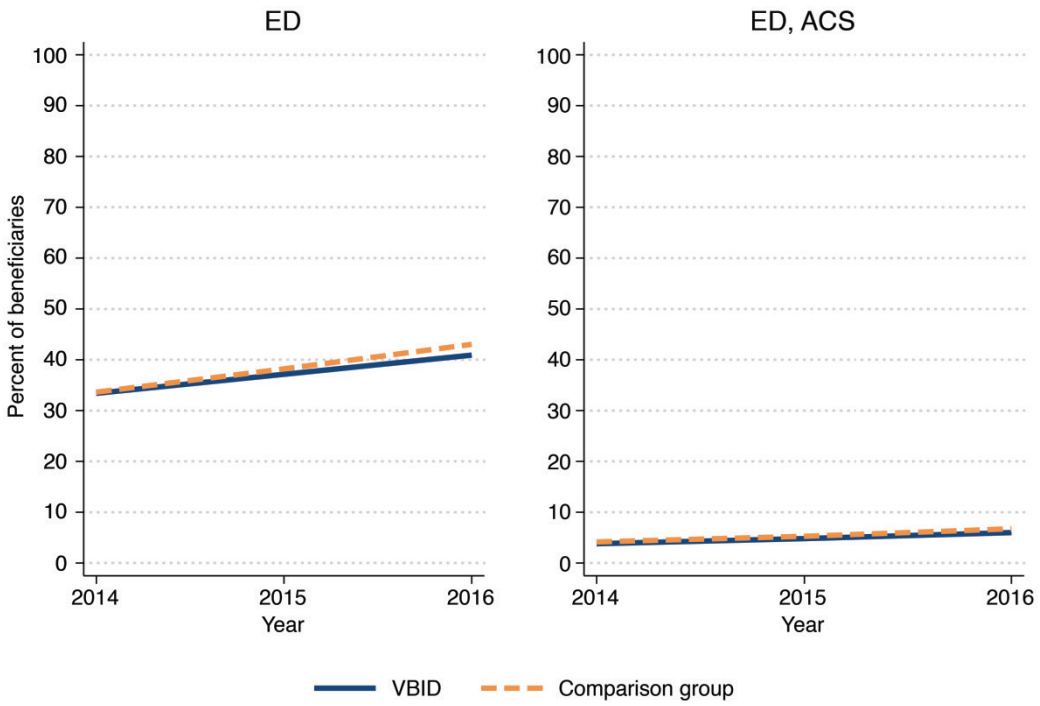
Abbreviation: DF = degrees of freedom.

Figure G.14. Trends in Proportion of Beneficiaries with at Least One Inpatient or Inpatient ACS Stay



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for inpatient visits was 0.17. The p-value for the test of parallel trends for inpatient ACS visits was 0.51.

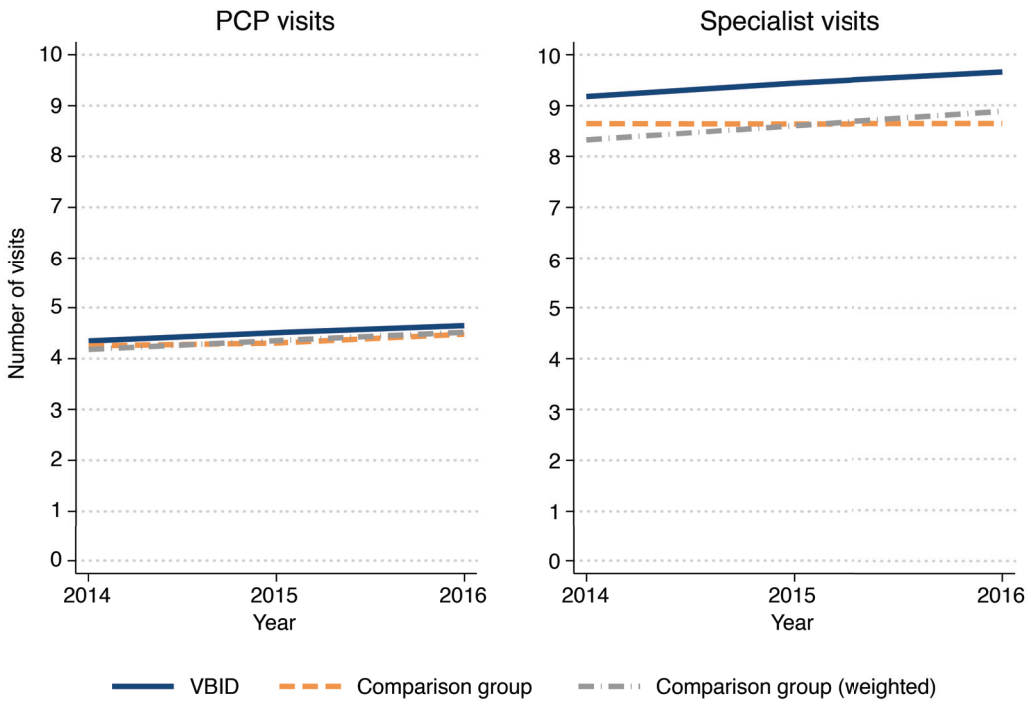
Figure G.15. Trends in Proportion of Beneficiaries with at Least One ED or ED ACS Visit



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for ED visits was 0.58. The p-value for the test of parallel trends for ED ACS visits was 0.99.

The trends for PCP visits and specialist visits showed clear nonparallel trends in the comparison group without weights (Figure G.16), but improved with entropy-balanced weights.

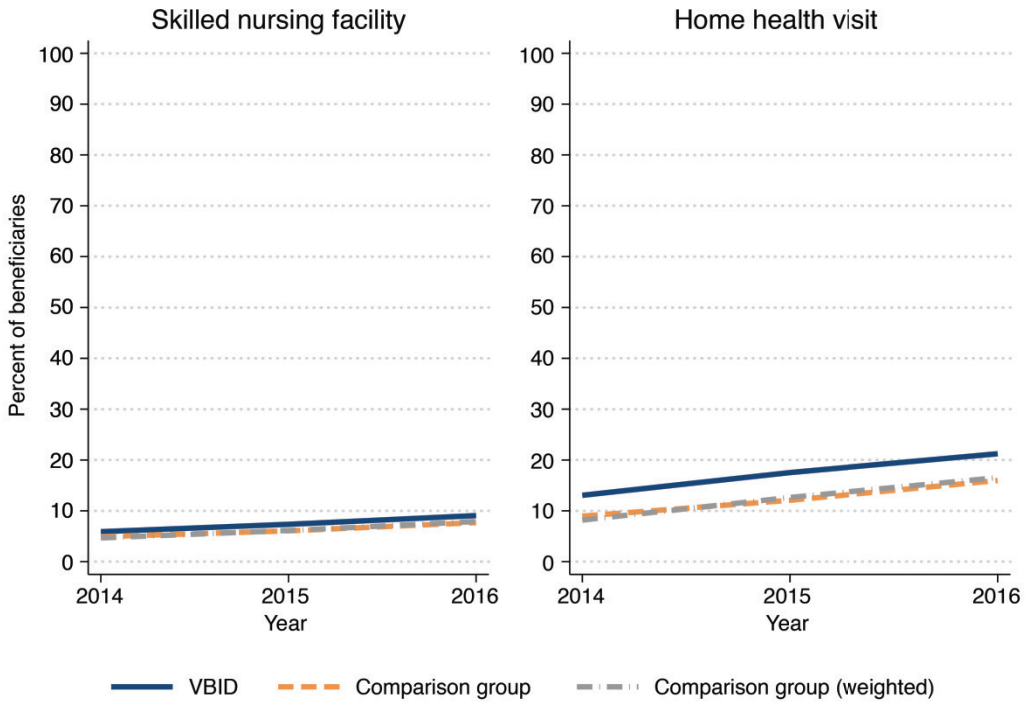
Figure G.16. Trends in the Number of PCP and Specialist Visits



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for PCP visits was 0.00 before weighting and 0.07 after weighting. The p-value for the test of parallel trends for specialist visits was 0.00 before and after weighting.

In general, the utilization of SNF and home health services was low. The proportion of beneficiaries with at least one SNF or home health visit trend improved with entropy weighting (Figure G.17).

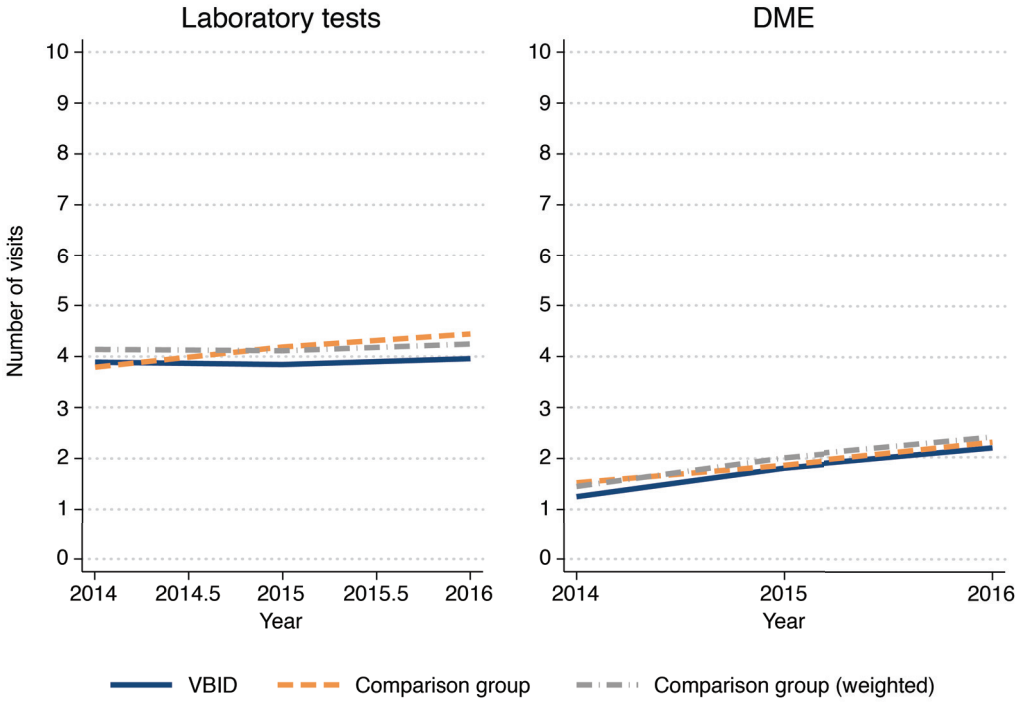
Figure G.17. Trends in the Proportion of Beneficiaries with At Least One SNF and Home Health Visit



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for at least one SNF visit was 0.00 before weighting and 0.01 after weighting. The p-value for the test of parallel trends for at least one home health visit was 0.00 before and 0.05 after weighting.

The laboratory tests and DME outcomes showed visual signs of not being parallel for the comparison group without entropy weighting even though the statistical test showed DME as parallel (Figure G.18).

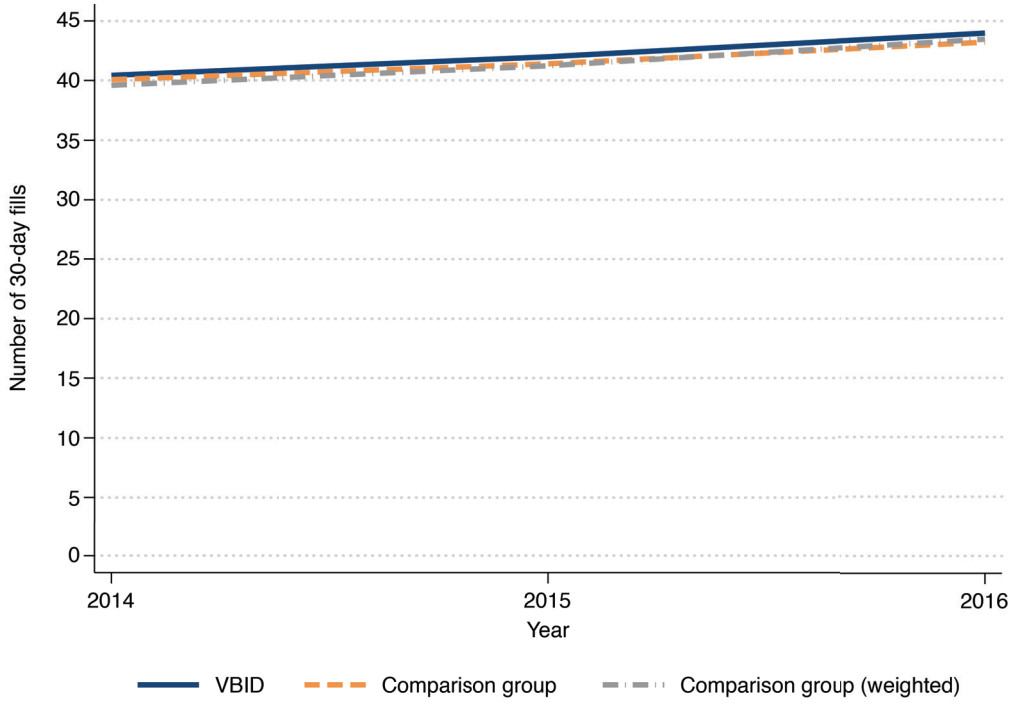
Figure G.18. Trends in the Number of DME and Laboratory Tests



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for laboratory claims was 0.00 before and after weighting. The p-value for the test of parallel trends for DME claims was 0.00 before and after weighting.

While trends in drug fills diverged for the treatment and comparison group, they became parallel after applying entropy-balanced weights (Figure G.19).

Figure G.19. Trends in the Number of 30-Day Drug Fills



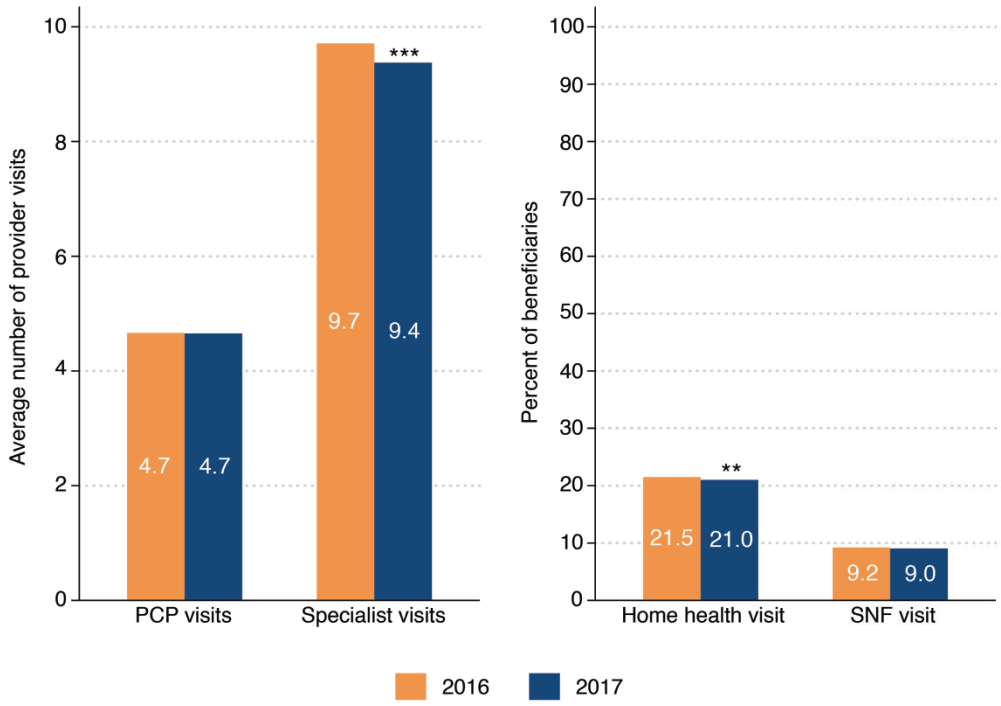
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends for number of 30-day fills was 0.00 before and after weighting.

Descriptive Results

We present the means of the general utilization categories for the VBID-eligible beneficiaries only from 2016 to 2017. These means are unadjusted and cannot be interpreted as demonstrating the causal effect of VBID on outcomes.

VBID-eligible beneficiaries visited a PCP approximately 4.7 times per year in both 2016 and 2017 (Figure G.20). VBID-eligible beneficiaries visited a specialist 9.7 times in 2016 and 9.4 times in 2017, which is a statistically significant decline at the 1-percent level. The proportion of VBID-eligible beneficiaries using SNF services was small and did not change markedly between 2016 and 2017. Approximately, 21 percent of beneficiaries used home health services in each year, although the decline in the proportion of beneficiaries with at least one home health visit is statistically significant.

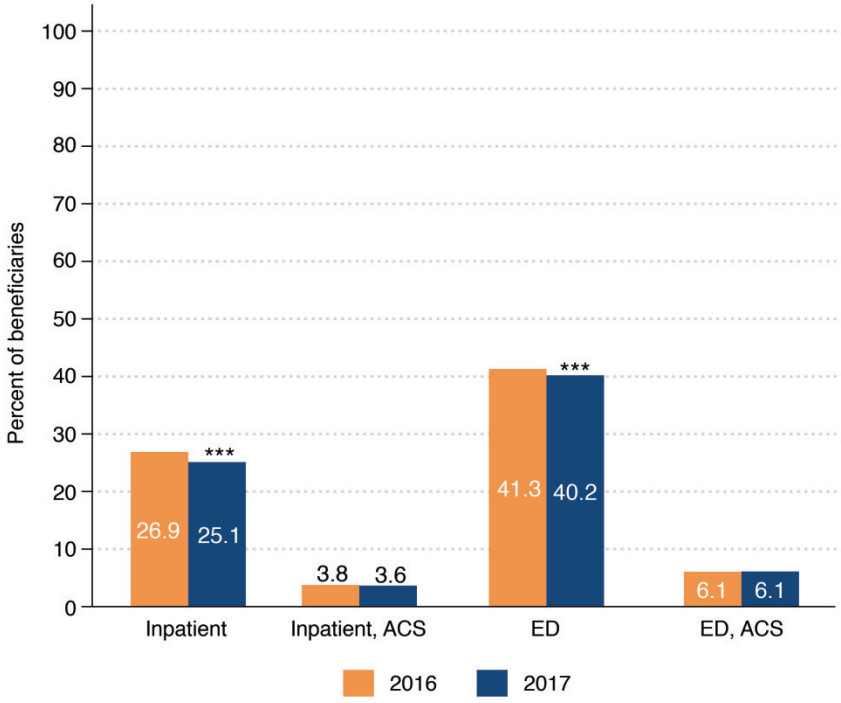
Figure G.20. Descriptive Statistics, Utilization of PCPs and Specialists, SNF and Home Health Among VBID-Eligible Beneficiaries, 2016 and 2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample size is 70,710 across all outcomes.

Figure G.21 shows that about 27 percent of VBID-eligible beneficiaries used the inpatient setting in 2016, and less than 5 percent of VBID-eligible beneficiaries had visits for ACS conditions (conditions for which hospital admissions or ED visits could be reduced or avoided with improved ambulatory care). Just over 40 percent of VBID-eligible beneficiaries had an ED visit in 2016, and just over 6 percent of those visits were for ACS conditions. In 2017, there was a slight decrease in the proportion of VBID-eligible beneficiaries visiting the ED (1.1-percentage points) or being admitted for an inpatient hospital stay (1.8-percentage points) for any reason, both statistically significant at the 1-percent level. There were no changes in the utilization of inpatient services for ACS conditions.

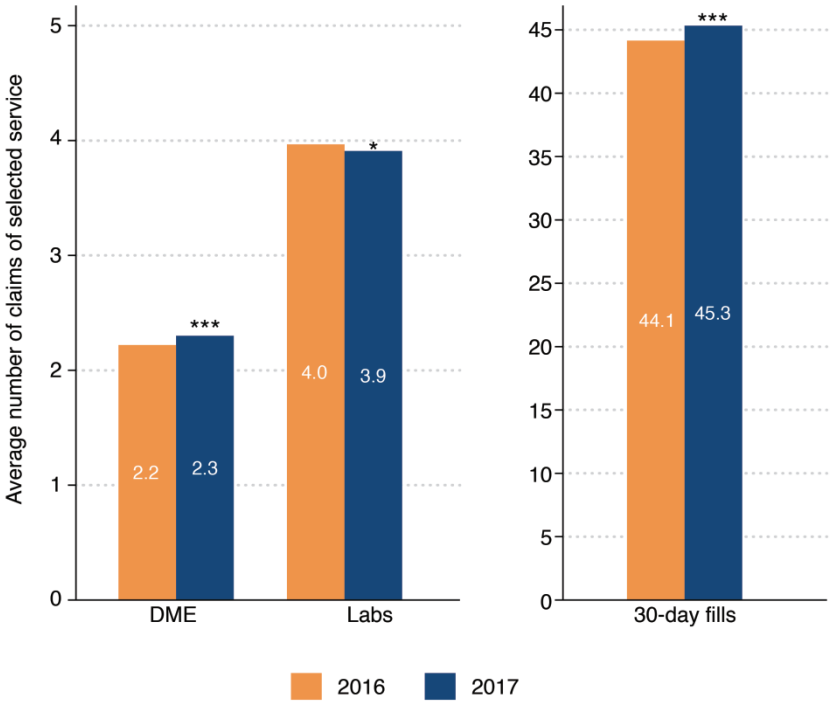
Figure G.21. Descriptive Statistics, Utilization of Inpatient and ED Among VBID-Eligible Beneficiaries, 2016 and 2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample size is 70,710 across all outcomes.

VBID-eligible beneficiaries had approximately 2.3 DME claims and approximately four laboratory tests on average in both 2016 and 2017 (Figure G.22). The utilization changed less than one claim for both services between 2016 and 2017, though the change for DME was a significant increase at the 1-percent level. Across all classes of drugs, VBID-eligible beneficiaries used 44.1 30-day fills in 2016, which increased to 45.3 fills in 2017 (statistically significant).

Figure G.22. Descriptive Statistics, Utilization of DME, Laboratory Tests, and Prescription Drugs Among VBID-Eligible Beneficiaries, 2016–2017



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing beneficiaries in 2016 with 2017, on the sample with four years of data. Sample size is 70,710 for laboratory tests and DME claims and 67,624 for prescription drugs.

Regression Results

In Tables G.11–G.13, we display the full regression results for the general utilization outcomes discussed in Chapter 6. There were 134,397 beneficiaries (treatment plus comparison) in these regression models and 128,875 for the prescription drug regression which was limited to the PBPs offering Part D. Table G.11 contains the regression coefficients for inpatient and ED utilization outcomes. Table G.12 shows the regression coefficients for primary care, specialist visits, as well as for the proportion of beneficiaries with at least one SNF stay or home health visit. Table G.13 lists the regression coefficients for laboratory tests, DME, and prescription drugs.

The marginal effects for the effect of being in the VBID group in the postperiod were discussed in Chapter 6. The process of matching beneficiaries was designed to balance characteristics between the VBID participants and comparison beneficiaries. The coefficients for age, LIS, disabled, race/ethnicity, and dual status vary in sign and magnitude across outcomes and thus do not have a clear effect on utilization. Across outcomes, being female generally has a positive and significant effect on utilization (except for labs), as does having ESRD. Risk scores are also positively associated with all outcomes.

Table G.11. Regression Coefficients, Difference-in-Difference Analysis, Inpatient and ED Utilization

Variable	Inpatient, Weighted	SE	Inpatient, ACSC	SE	ED	SE	ED, ACSC	SE
2015 (2014 reference)	0.163***	0.012	0.152***	0.027	0.128***	0.009	0.006***	0.001
2016	0.279***	0.012	0.266***	0.026	0.226***	0.009	0.011***	0.001
2017	0.133***	0.015	0.021	0.032	0.163***	0.011	0.004***	0.001
Treatment beneficiary*year	0.029*	0.017	0.181***	0.036	0.014	0.014	0.007***	0.001
Female	0.155***	0.009	0.207***	0.019	0.228***	0.008	0.016***	0.001
Age	−0.005***	0.001	0.035***	0.001	0.010***	0.001	0.001***	0.000
LIS	0.016	0.021	0.134***	0.039	0.191***	0.018	0.011***	0.002
Dual	−0.142***	0.025	0.017	0.045	−0.079***	0.022	−0.001	0.002
Disabled	−0.248***	0.014	0.202***	0.026	0.099***	0.012	0.000	0.001
Black (white, reference)	−0.194***	0.032	−0.144**	0.063	−0.002	0.034	−0.008***	0.002
Hispanic	0.040	0.071	0.013	0.143	0.313***	0.063	0.003	0.005
Asian/Pacific Islander	−0.130***	0.049	−0.270**	0.122	−0.449***	0.042	−0.006**	0.002
American Indian/Alaskan Native	1.151*	0.591	1.881	1.342	0.272	0.767	0.087	0.065
Multiple	−0.603	0.579	−1.830	1.346	3.067***	0.781	−0.005	0.033
ESRD	1.621***	0.039	1.729***	0.040	1.155***	0.036	0.140***	0.005
Plan OOP max, 2016	0.000***	0.000	−0.000	0.000	−0.000***	0.000	0.000*	0.000
Plan premium, 2016	0.000	0.000	0.000***	0.000	−0.000***	0.000	0.000	0.000
Risk score	1.253***	0.006	0.585***	0.005	0.961***	0.005	0.042***	0.000
Observed, 10−11 mo (all year, reference)	0.724***	0.031	0.586***	0.058	0.431***	0.027	0.031***	0.003
Observed, 7−9 mo	0.638***	0.029	0.337***	0.056	0.288***	0.024	0.012***	0.002
Observed, 4−6 mo	0.296***	0.033	0.063	0.066	−0.132***	0.028	−0.001	0.002
Observed, 1−3 mo	−0.526***	0.047	−0.525***	0.086	−1.035***	0.038	−0.027***	0.002
Constant	−3.464***	0.062	−7.714***	0.125	−3.051***	0.056	−0.091***	0.005
PO fixed effect	Yes		Yes		Yes		Yes	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs.

Table G.12. Regression Coefficients, Difference-in-Difference Analysis, Primary Care, SNF, and Home Health

Variable	Primary Care Visits	SE	Specialty Visits	SE	SNF, Weighted	SE	Home Health, Weighted	SE
2015 (2014 reference)	0.028***	0.003	0.024***	0.003	0.149***	0.020	0.374***	0.014
2016	0.039***	0.003	0.022***	0.003	0.235***	0.020	0.549***	0.014
2017	0.007	0.004	−0.025***	0.004	0.216***	0.027	0.457***	0.020
Treatment beneficiary*year	0.031***	0.004	0.021***	0.004	−0.058**	0.028	0.018	0.020
Female	0.084***	0.005	−0.017***	0.004	0.480***	0.017	0.425***	0.013
Age	−0.002***	0.000	−0.001**	0.000	0.060***	0.001	0.048***	0.001
LIS	0.005	0.009	−0.074***	0.009	−0.081**	0.036	0.123***	0.027
Dual	−0.151***	0.010	−0.096***	0.010	0.514***	0.042	−0.223***	0.032
Disabled	0.083***	0.007	0.057***	0.007	0.104***	0.025	0.178***	0.019
Black (white, reference)	−0.107***	0.014	−0.076***	0.014	−0.109	0.068	−0.020	0.036
Hispanic	0.037	0.035	−0.068**	0.030	−0.540***	0.154	−0.134	0.099
Asian/Pacific Islander	−0.138***	0.023	−0.233***	0.023	−0.318***	0.113	−0.403***	0.074
American Indian/Alaskan Native	−0.373	0.320	−0.174	0.284	4.717***	1.606	−0.608	0.680
Multiple	0.917***	0.249	0.073	0.223	−7.060***	1.770	1.942***	0.537
ESRD	0.092***	0.013	0.188***	0.013	1.174***	0.050	1.040***	0.042
Plan OOP max, 2016	0.000	0.000	0.000***	0.000	−0.000	0.000	0.000	0.000
Plan premium, 2016	0.000***	0.000	0.000***	0.000	0.000***	0.000	0.000	0.000
Risk score	0.152***	0.001	0.198***	0.001	0.808***	0.006	0.837***	0.006
Observed, 10–11 mo (all year, reference)					0.736***	0.047	0.420***	0.039
Observed, 7–9 mo					0.671***	0.045	0.231***	0.037
Observed, 4–6 mo					0.319***	0.054	−0.140***	0.044
Observed, 1–3 mo					−0.408***	0.075	−0.986***	0.065
Constant	1.218***	0.029	1.708***	0.028	−9.706***	0.118	−8.313***	0.083
PO fixed effect	Yes		Yes		Yes		Yes	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs.

Table G.13. Regression Coefficients, Difference-in-Difference Analysis, Laboratory Tests, DME, and Rx

Variable	Laboratory tests, Weighted	SE	DME, Weighted	SE	Prescription Drugs	SE
2015 (2014 reference)	−0.008**	0.004	0.297***	0.011	0.052***	0.001
2016	−0.005	0.005	0.455***	0.012	0.106***	0.001
2017	0.038***	0.006	0.449***	0.012	0.130***	0.002
Treatment beneficiary*year	−0.050***	0.006	0.062***	0.012	0.004**	0.002
Female	−0.066***	0.007	0.104***	0.014	0.045***	0.003
Age	0.010***	0.000	0.001	0.001	0.005***	0.000
LIS	−0.020	0.014	0.169***	0.026	0.118***	0.005
Dual	0.126***	0.017	−0.061**	0.027	−0.011*	0.005
Disabled	0.059***	0.011	0.380***	0.017	0.188***	0.005
Black (white, reference)	−0.034	0.022	−0.137***	0.036	−0.081***	0.011
Hispanic	−0.029	0.046	−0.286***	0.099	0.016	0.023
Asian/Pacific Islander	−0.209***	0.029	−0.626***	0.086	−0.182***	0.016
American Indian/Alaskan Native	0.667	0.646	0.580	0.481	−0.494***	0.187
Multiple	−0.679*	0.395	2.639***	0.464	1.247***	0.186
ESRD	0.326***	0.022	0.308***	0.035	0.010	0.007
Plan OOP max, 2016	0.000***	0.000	−0.000***	0.000	−0.000***	0.000
Plan premium, 2016	0.000***	0.000	0.000***	0.000	0.000***	0.000
Risk score	0.204***	0.002	0.268***	0.004	0.039***	0.001
Constant	−0.090**	0.045	0.062	0.081	3.279***	0.021
PO fixed effect	Yes		Yes		Yes	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from the difference-in-differences models comparing VBID-participating PBPs to matched comparison PBPs.

Limitations

The analysis of the utilization outcomes relies primarily on the MA encounter data for medical services (Part C). There are two limitations of this data source that warrant caveats on this analysis. The first limitation is that there have been documented concerns on whether the encounter data are "complete," (Creighton, Duddy-Tenbrunsel, and Michel, 2019; GAO, 2017; HHS, OIG, 2018; Medicare Payment Advisory Commission, 2019) meaning that POs submit all of the encounter records for each beneficiary. The Medicare encounter data record services for beneficiaries in MA plans, but not transactions for payment, as claims do in Medicare fee for service. Because FFS claims represent requests for payment from providers to CMS, it is expected that claims are an accurate reflection of services rendered. Encounters are records of services between providers and the POs, regardless of the type of payment structure—some providers may be on capitated payments with POs for a group of beneficiaries or set of services.

Prior to 2012, MA POs did not have to submit encounter records for medical services, because they are paid a capitated amount per beneficiary per year. In 2012, CMS began requesting that MA POs submit encounters for multiple purposes that include risk adjustment and research (CMS, 2018). Encounter data first became available to researchers using the IDR in 2014 and was released more recently for the broader research community. In a 2019 report on the encounter data, MedPAC identified several issues with some service categories such as home health or SNF that were underreported among plans, which may be due in part to the fact that only inpatient, outpatient, and professional encounters are used for risk adjustment purposes. MedPAC still found that 97 percent of contracts were reporting physician encounters and 96 percent were reporting inpatient encounters (Medicare Payment Advisory Commission, 2019), though the report highlighted there may be some lack of completeness in particular service categories among reporting contracts.

We assessed completeness by reviewing utilization across all POs for the sample of VBID-eligible beneficiaries and their matched comparison beneficiaries to identify whether there are POs with no utilization reported in specific categories. Comparing the level of utilization across POs is problematic in our context since each PO targeted a different set of conditions, some of which could be expected to result in more utilization (e.g., CHF and diabetes) than others (e.g., hypertension). We found one control PO with data in 2014 that appeared to be incomplete. We assessed the sensitivity of our results to dropping this PO and their matched VBID beneficiaries from the analysis and found that the results were unchanged.

The Part D event data do not suffer from the same limitations as the encounter data, because POs were required to submit data on all prescription drug events to CMS for payment reconciliation processes from the inception of the program.

The second potential caveat for this analysis is that there are many interpretations on how to assign a particular encounter to a service category, such as primary care services. We have assigned encounters to service categories using an algorithm developed at RAND under contract with CMS's Medicare Plan Payment Group (Mulcahy et al., 2019). There may be additional alternative ways to group services together for various purposes, which may lead to differences in findings across studies.

Appendix H. Quality/Adherence and Health Status Methods and Analytic Results

This appendix describes the details of the methods used for the analyses of health care quality/adherence and health outcomes, as a supplement to Chapter 7, and provides the detailed results. Table H.1 summarizes our approach across *all* quality/adherence and health outcome measures. In the sections below, we discuss our approach to each category of measures in greater detail.

Table H.1. Health Care Quality/Adherence and Health Status Analytic Approaches

Outcome Category	VBID Group	Comparison Group	Years Observed	Model	Covariates
Quality/adherence (contract level)	Contracts with participating PBPs[a]	Contracts with matched comparison PBPs	2014–2017	Difference-in-differences	PO and year fixed effects
Quality/adherence (beneficiary level)	VBID-eligible beneficiaries in participating PBPs	Matched comparison beneficiaries	2014–2018	Difference-in-differences with GEE	Beneficiary and plan characteristics, PO and year fixed effects
Health status—self-reported	Cohort 18 HOS respondents in participating PBPs	Any matched comparison beneficiaries who were cohort 18 HOS respondents	2015, 2017	Difference-in-differences; one balanced panel model, one unbalanced panel model with a beneficiary random effect	Beneficiary and plan characteristics, PO and year fixed effects
Health status—HCC/RxHCC	VBID-eligible beneficiaries in participating PBPs	Matched comparison beneficiaries	2014–2018	Difference-in-differences with GEE	Beneficiary and plan characteristics, PO and year fixed effects
Health status—mortality	VBID-eligible beneficiaries in participating PBPs	Matched comparison beneficiaries	2017–2018	Kaplan-Meier survival curves, stratified by sex	Beneficiary characteristics

NOTE: [a] Not every PBP under a given contract participated in the VBID model test. Among the contracts in our sample, the average share of their PBPs participating in VBID was 42.9 percent, and across individual contracts the share of VBID PBPs ranged from 5.3 percent to 100 percent.

Health Care Quality and Adherence: Contract-Level Analyses

Study Population and Outcome Measures

We compare changes in the outcomes of interest in contracts containing VBID-participating PBPs ("VBID contracts") and contracts containing matched comparison PBPs ("comparison contracts"), observed during measurement years 2014–2017.

Table H.2 summarizes our contract-level quality/adherence measures, including the Star Rating component measures included in each of the three composite indexes (VBID-targeted condition index, diabetes index, and general medical care index).

Table H.2. Quality/Adherence Measures for MAO-Contract Level Analyses

Rating/Index	Component Measures	Measure Source
Overall Star Rating	All Star Ratings measures used to calculate the overall Star Rating in a given year	—
VBID-targeted condition index[a]	Controlling blood pressure	HEDIS
	Medication adherence for hypertension (RAS antagonists)	PDE data
	Medication adherence for cholesterol (statins)	PDE data
Diabetes index	Diabetes care—eye examination	HEDIS
	Diabetes care—kidney disease monitoring	HEDIS
	Diabetes care—blood sugar controlled	HEDIS
	Medication adherence for diabetes medications	PDE data
General medical care index	Breast cancer screening	HEDIS
	Colorectal cancer screening	HEDIS
	Monitoring physical activity	HEDIS
	Adult BMI assessment	HEDIS
	Osteoporosis management in women who had a fracture	HEDIS
	Rheumatoid arthritis management	HEDIS

NOTE: [a] Includes measures relevant to hypertension, diabetes, and CHF. Abbreviations: RAS = renin angiotensin system; BMI = body mass index.

We constructed the three composite indexes following the approach of Kling, Liebman, and Katz (2007), in which individual measures are first normalized by subtracting the mean and dividing by the standard deviation of the comparison group observations, after which the index is calculated as the simple average of the normalized component measures. This is a common strategy used to limit multiple hypothesis testing (i.e., in which analyzing a large number of distinct outcome measures increases the risk of falsely rejecting the null hypothesis that VBID had no effect on quality/adherence). The strategy has the added advantage of reducing measurement error by averaging across outcomes. Each composite index is scaled to have a value of 0 to 100, where higher values of the index indicate better performance. Missing values

for individual component measures were imputed as the mean value of the component measure for comparison contracts in the same year.

Note that, by convention, Star Rating measures are named according to the year in which they are reported, but there is a two-year lag between when quality/adherence is actually measured and when it is reported. For example, the 2019 Star Rating is reported in 2019, but is (largely) based on the quality of care delivered in 2017. For simplicity and ease of interpretation, in our analyses we relabel the year of the Star Rating measures to correspond to the year in which quality was actually measured, that is, the 2019 Star Rating is relabeled as the 2017 measurement year Star Rating.

Difference-in-Differences Models

Note that this section will use different indexing from other sections of this report due to the outcome being measured at the contract level. Let y_{qrt} be the outcome for contract *r* in PO q at year *t*, and let $VBID_{qr}$ be an indicator that the *r*th contract in PO q is a VBID-participating contract. Our difference-in-differences model for contract-level analysis is given by

$$y_{qrt} = \alpha + \alpha_t + \theta_q + \gamma_t * VBID_{qr} + \varepsilon_{qrt}, \quad \text{(H.1)}$$

where

- α = overall intercept
- α_t = year fixed effect (with $\alpha_{2014} = 0$) that captures the trend over time
- θ_q = PO fixed effect capturing time-invariant differences between POs
- γ_t = interaction effect between time and VBID-participating contracts (with γ_t = 0 for $t \leq 2016$) that represents the difference-in-differences estimate.

We used linear models for all outcome measures except for the overall Star Rating; because this measure can only take discrete variables, we used an ordered logistic regression. Parallel trends were assessed using the strategy described in Appendix D, but using a model consistent with Equation (H.1). In addition, the propensity score weighting to improve parallel trends was implemented as described in Appendix D.

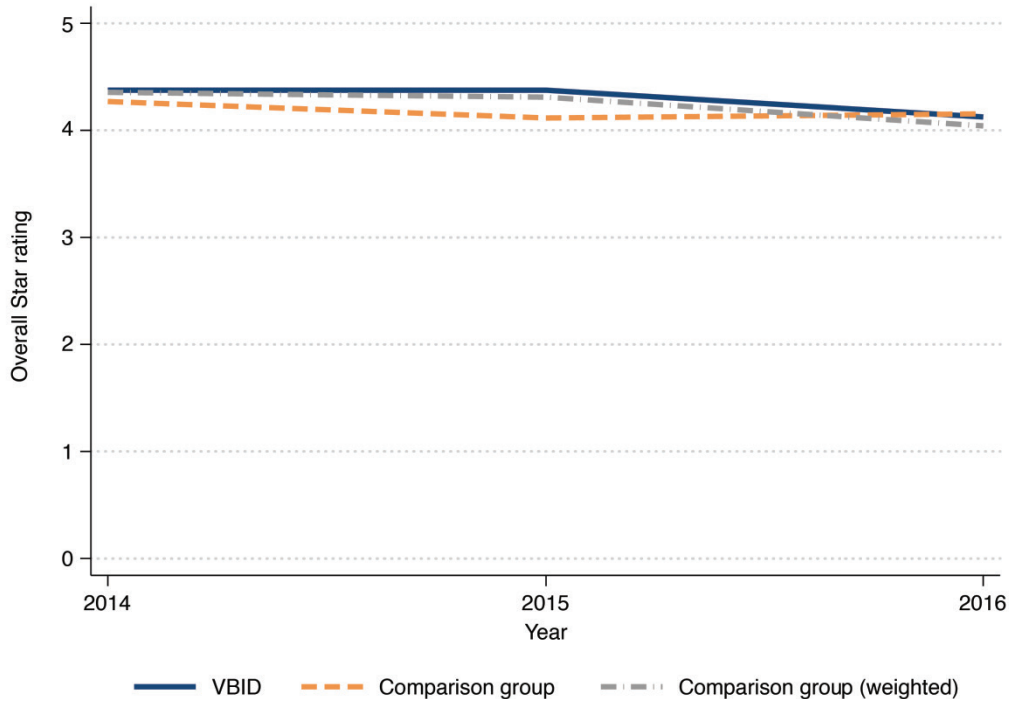
Results for the tests of the parallel trends assumption are shown in Table H.3. For each outcome measure, Table H.3 shows the difference in slopes between VBID and comparison contracts and the corresponding p-value. For the overall Star Rating and diabetes index, the pre-VBID trends differed significantly between the VBID and comparison groups when tested in a regression framework. For all other measures, even though there was no significant difference in pre-VBID trends when tested in a regression framework, visually the plots of unadjusted average measures did not appear parallel (Figures H.1–H.4). We therefore reweighted each outcome measure using propensity scores as described in Appendix D.

Table H.3. Contract-Level Quality/Adherence Measures: Summary of Parallel Trends Tests Results

Measure	Trend Years	Unweighted Parallel Trends Test Coefficient	p	Weighted Parallel Trends Test Coefficient	p	Model Used
Overall star rating	2014–2015	1.89	0.03	0.77	0.84	Weighted
	2015–2016	−3.79		−0.05		
VBID-targeted condition index	2014–2015	1.39	0.68	−0.34	0.94	Weighted
	2015–2016	2.02		−0.82		
Diabetes index	2014–2015	−0.35	0.05	0.84	0.73	Weighted
	2015–2016	−10.55		−3.97		
General medical care index	2014–2015	−1.75	0.48	0.20	0.98	Weighted
	2015–2016	−4.38		−0.65		

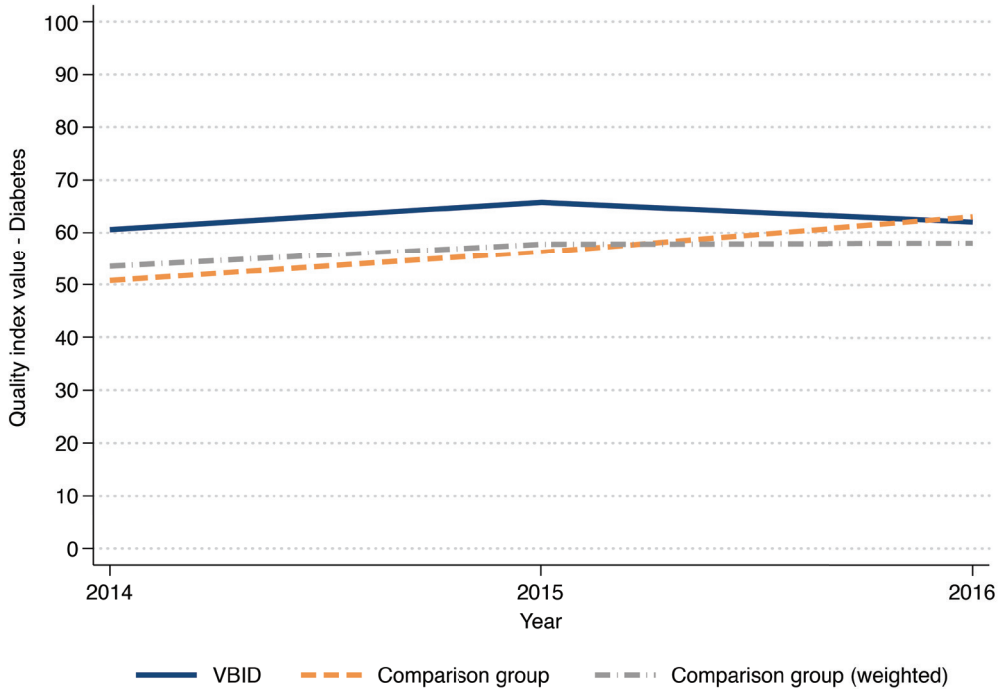
NOTE: p-values correspond to a joint test of significance of the 2014–2015 and 2015–2016 trend coefficients. For the overall Star Rating, p-values were derived from a chi-square test. For all other outcome measures, p-values were derived from an F test.

Figure H.1. Pre-VBID Trends for Overall Star Rating, Unweighted and Weighted



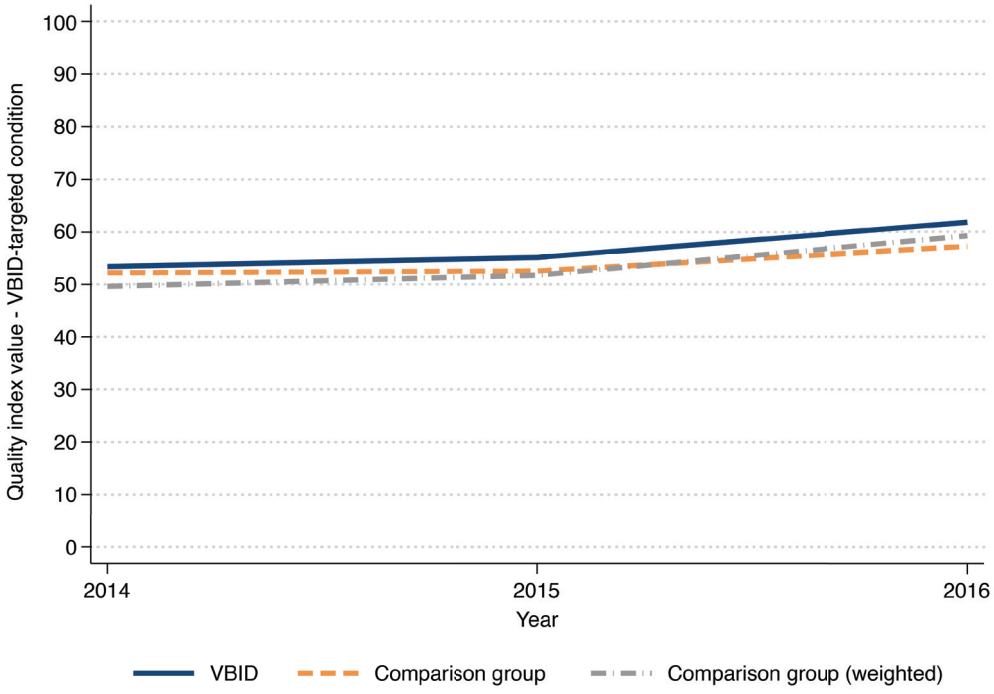
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.03 before weighting and 0.84 after weighting.

Figure H.2. Pre-VBID Trends for Diabetes Index, Unweighted and Weighted



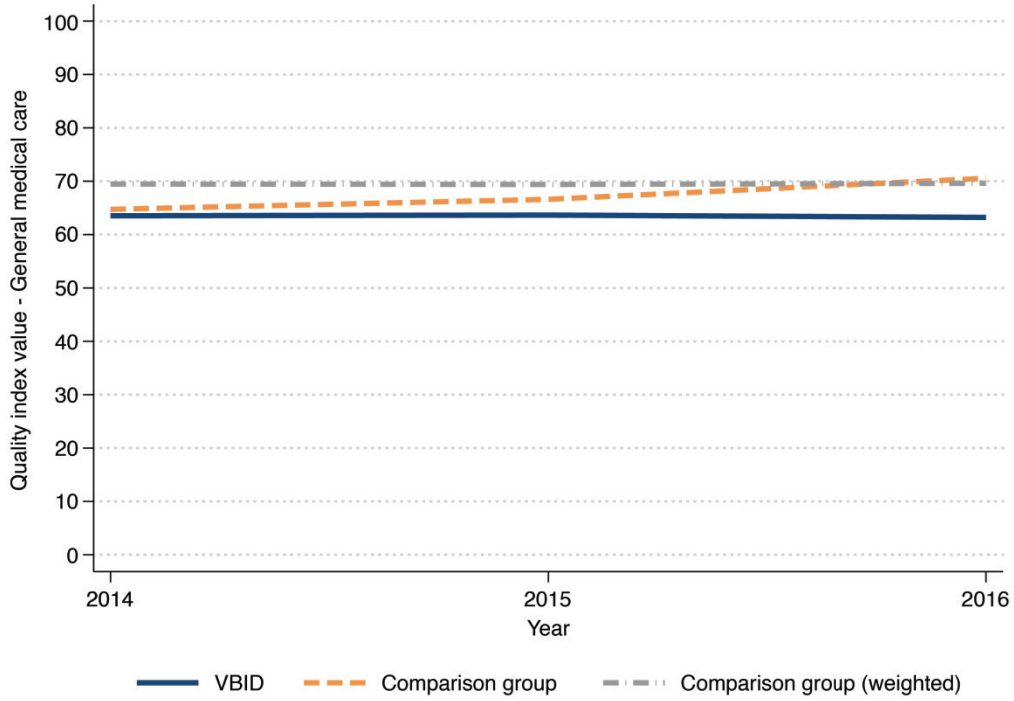
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.05 before weighting and 0.73 after weighting.

Figure H.3. Pre-VBID Trends for Targeted Condition Index, Unweighted and Weighted



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.68 before weighting and 0.94 after weighting.

Figure H.4. Pre-VBID Trends for General Medical Care Index, Unweighted and Weighted



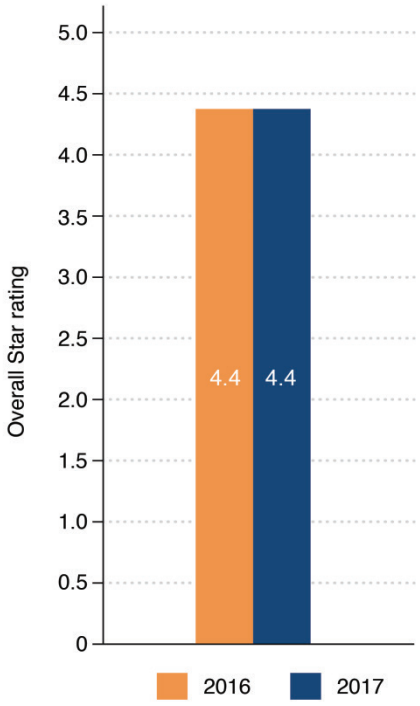
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.48 before weighting and 0.98 after weighting.

Table H.3 shows that, after reweighting the pre-VBID trends using propensity scores, there were no significant differences in the slopes of the pre-VBID trends for VBID and comparison contracts (including for the overall Star Rating and the diabetes index, for which trends previously differed significantly). Figures H.1–H.4, which show unweighted and weighted pre-VBID trends on the same plots, show that propensity score weighting achieved similar slopes for these trends for VBID and comparison contracts. We therefore used the propensity scores as weights in our difference-in-differences models.

Descriptive Statistics

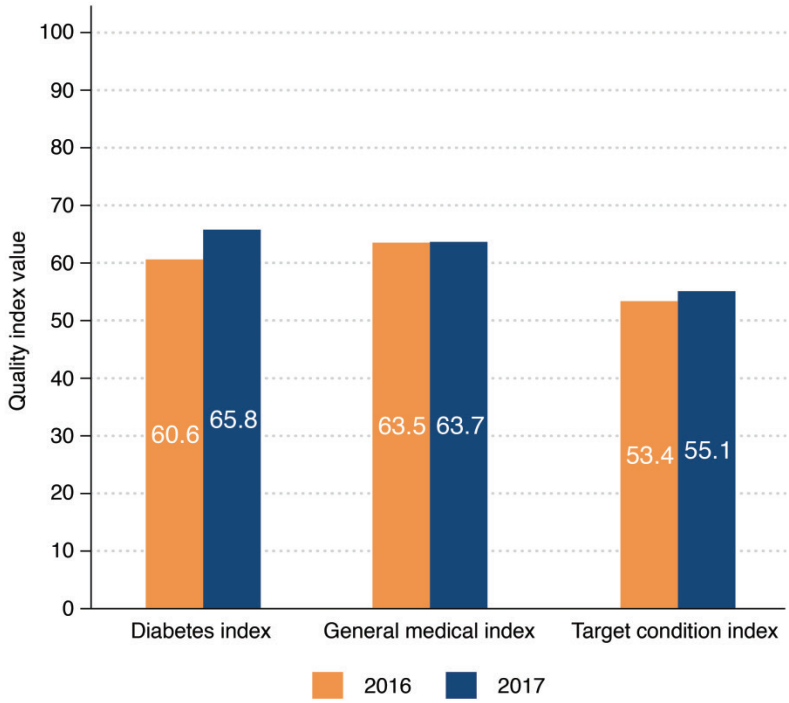
In Figures H.5 and H.6, we first present descriptive findings illustrating how the contract-level quality/adherence measures changed before and after the implementation of VBID among VBID-participating contracts. These estimates are unadjusted for covariates and do not necessarily reflect the causal effect of VBID. Figures H.5 and H.6 show that from 2016 to 2017, there were no statistically significant changes in the overall Star Rating, the quality of care for diabetes or other VBID-targeted conditions, or for recommended preventive care (i.e., the general medical care index).

Figure H.5. Descriptive Statistics for Overall Star Ratings, VBID Contracts Only



NOTE: Sample size is 24.

Figure H.6. Descriptive Statistics for Quality Indexes, VBID Contracts Only



NOTE: Differences are not statistically significant at conventional levels based on a paired t-test comparing beneficiaries in 2016 with 2017. Sample size is 24 across all outcomes.

These descriptive statistics highlight changes in quality of care and medication adherence for VBID contracts between 2016 and 2018, but they do not tell us whether VBID might have contributed to these changes.

Results

Regression results are shown in Table H.4. VBID had no significant effect on any contract-level quality/adherence measures.

Table H.4. Contract-Level Quality/Adherence Analyses: Difference-in-Differences Regression Coefficients

Variable	Overall Star Rating	SE	VBID-Targeted Condition Index	SE	Diabetes Index	SE	General Medical Care Index	SE
Treatment contract × year (γt)	**0.75**	**1.33**	**−4.71**	**3.06**	**−4.86**	**3.79**	**−0.87**	**2.98**
2015 (2014 reference)	−0.21	0.65	1.90	1.64	4.77	3.00	0.02	1.35
2016	−3.23***	1.11	9.03***	2.46	2.82	3.68	−0.08	2.41
2017	−3.49***	0.83	19.92***	3.86	14.24***	3.91	4.94**	2.23
Parent org. fixed effects	Yes	—	Yes	—	Yes	—	Yes	—
Constant	N/A	—	49.85***	1.85	47.46***	2.41	66.86***	1.21
Mean	4.19		58.27		61.23		67.59	
SD	0.40	—	17.72	—	17.94	—	13.94	—
Range	3–5		0–100		0–100		0–100	
N	152		152		152		152	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. All models weighted to satisfy parallel trends. SEs clustered by parent organization. N refers to number of contract-year observations.

Limitations and Sensitivity Analyses

Measurement Years of Overall Star Rating

A limitation of the overall Star Rating is that for any given measurement year *t*, the overall Star Rating measure reflects a contract's performance on HEDIS measures in the same year *t*, but also performance on CAHPS measures in the following year *t + 1*. Of the 47 Star Rating component measures, 11 are CAHPS measures. In our analyses, this is only consequential in measurement year 2016, the final year before VBID was introduced: In this year, the overall Star Rating mostly reflects the quality of care delivered in 2016 but will also reflect some 2017 (VBID Year 1) CAHPS measures. This could bias our estimated effects of VBID toward the null. Table H.5 shows the results of a sensitivity analysis excluding the measurement year 2016 overall Star Rating; the estimated effect of VBID (i.e., the coefficient on the treatment contract × year interaction term in Table H.5) is not statistically significantly different from zero. The results of our sensitivity analysis are therefore consistent with the results of our main analysis: that VBID had no significant effect on the Overall Star Rating of participating contracts.

Table H.5. Contract-Level Quality/Adherence Analyses: Difference-in-Differences Regression Coefficients

Variable	Overall Star Rating	SE
Treatment contract × year (γt)	**−0.94**	**1.64**
2015 (2014 reference)	−0.04	0.70
2017	−1.72*	0.99
Parent org. fixed effects	Yes	—
N	114	—

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. All models weighted to satisfy parallel trends. SEs clustered by parent organization. N refers to number of contract-year observations.

Contract Merges

In 2016, two comparison contracts in our sample merged, which were under the same PO. To ensure that our analysis included stable organizational entities over time, we considered these two combined contracts to represent a single entity throughout our observation period. Therefore, from 2014 through 2016, this entity was assigned the average value of each outcome measure across the two contracts, weighted by contract enrollment in that year. Since the overall Star Rating can only take discrete values, for this outcome measure the average was rounded to the nearest possible Star Rating level.

Other Limitations

Our sample size was small due to the limited number of VBID and comparison contracts; hence, we had limited statistical power to detect VBID's impacts. In addition, our follow-up period was only one year, which might have been an insufficient period of time in which to detect changes in quality/adherence. It is also worth noting that, in a given VBID contract, VBID was typically implemented by some PBPs but not others, so even if quality improvements were

achieved by a subset of PBPs, they might not have been sufficient to change overall contract-level quality. Last, the difference-in-differences parallel trends assumption was not satisfied for the overall Star Rating and the diabetes index, and although trends were not statistically different for other outcomes, they did not appear to be parallel based on visual inspection. However, we were able to achieve parallel trends—both statistically and visually—through reweighting.

Health Care Quality/Adherence: Beneficiary-Level Analyses

Study Population and Outcome Measures

Outcome measures are summarized in Table H.6. For analyses of breast cancer screening, our sample included female VBID-eligible beneficiaries in VBID-participating PBPs ("VBID beneficiaries") and their matched comparisons ("comparison beneficiaries"), observed during measurement years 2014–2017. For analyses of all other beneficiary-level quality measures, we restricted our sample to VBID beneficiaries who (1) had the medical condition targeted by the measure, (2) were in a VBID-participating PBP targeting that same medical condition, and (3) were in a MA-PD PBP (PDE data are required for the medication adherence measures). We included these individuals' matched comparison beneficiaries. For these analyses, we used measurement years 2014–2018.

With the exception of percent of eligible measures satisfied, all measures are binary (1 if beneficiary was eligible for the measure and satisfied it; 0 if the beneficiary was eligible for the measure and did not satisfy it). Percent of eligible measures satisfied is a continuous measure with a value from 0 to 1. The denominator for this measure is the total number of the three beneficiary-level medication adherence measures for which a given beneficiary is eligible, and the numerator is the number of these measures satisfied. For example, if a beneficiary with diabetes was eligible for both the medication adherence for diabetes medications and medication adherence for cholesterol measures, but only satisfied the former, his or her score would be 0.5.

Table H.6. Beneficiary-Level Quality/Adherence Measures

Measure	Related Conditions	Measure Source
Medication adherence for diabetes medications	Diabetes	PDE data
Medication adherence for hypertension (RAS antagonists)	Hypertension, CHF	PDE data
Medication adherence for cholesterol (statins)	Diabetes, CHF	PDE data
Percent of eligible measures satisfied	Diabetes, hypertension, CHF	PDE data
Breast cancer screening	General	HEDIS data

Abbreviation: RAS = renin angiotensin system.

Difference-in-Differences Models

Let y_{cpit} be the outcome for individual i in PBP p in PO c at year t, let $VBID_{cp}$ be an indicator that the p-th PBP in PO c is a VBID-participating PBP, and let X_{cpit} be a set of

additional characteristics to be controlled for in the analysis, including the PBP OOP maximum and the PBP premium and beneficiary demographic characteristics. Our difference-in-differences model is given by

$$y_{cpit} = \alpha + \alpha_t + \theta_c + \gamma_t * VBID_{cp} + \beta^T X_{cpit} + \varepsilon_{cpit}, \quad \text{(H.2)}$$

where

- α = overall intercept
- α_t = year fixed effect (with $\alpha_{2014} = 0$) that captures the trend over time in the comparison PBPs
- θ_c = PO fixed effect capturing time-invariant differences between POs
- γ_t = interaction effect between time and VBID-participating PBPs (with $\gamma_t = 0$ for $t \leq 2016$) that represents the difference-in-differences estimate of interest
- β = effect of the additional characteristics included in the model.

We used linear models to test for the outcome percent of eligible measures satisfied and a comparable logistic regression for all other quality/adherence measures. To adjust for serial correlation in the setting of repeated observations on the same individuals, we estimated these models using GEE with an AR-2 correlation structure.[1] Parallel trends were tested as described in Appendix D, but using the model formulation shown here.

Results for the tests of the parallel trends assumption are shown in Table H.7, and Figures H.7–H.11 show unweighted and weighted pre-VBID trends.

Table H.7. Beneficiary-Level Quality/Adherence Measures: Summary of Parallel Trends Tests Results

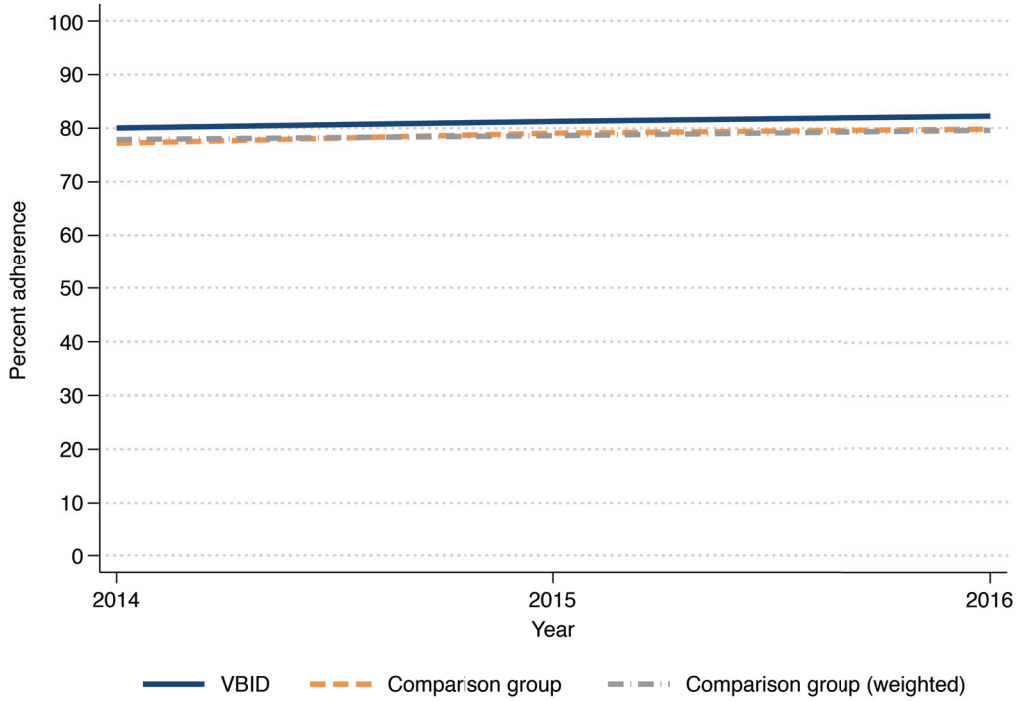
		Unweighted Parallel Trends Test		Weighted Parallel Trends Test		
Measure	**Trend Years**	**Coefficient**	**p**	**Coefficient**	**p**	**Model Used**
Medication adherence: diabetes	2014–2015 2015–2016	−0.014 0.196	0.03	0.025 0.113	0.23	Weighted
Medication adherence: hypertension	2014–2015 2015–2016	0.031 0.003	0.71	0.016 −0.045	0.55	Weighted
Medication adherence: cholesterol	2014–2015 2015–2016	−0.016 0.009	0.93	0.050 −0.005	0.49	Unweighted
Percent of eligible measures satisfied	2014–2015 2015–2016	0.004 0.004	0.05	–0.001 −0.001	0.93	Weighted
Breast cancer screening	2014–2015 2015–2016	−0.079 −0.069	0.02	0.045 0.013	0.53	Weighted

NOTE: p-values correspond to a joint test of significance of the 2014–2015 and 2015–2016 trend coefficients and were derived from a chi-square test.

[1] Choice of the AR-2 correlation structure was based on our examination of the correlation in residuals over time.

For the medication adherence for cholesterol measure, the pre-VBID parallel trends assumption was satisfied based on the unweighted regression results (Table H.7), and plotted trends were not improved by weighting (Figure H.7); therefore, we used the unweighted outcome measures in our analysis.

Figure H.7. Pre-VBID Trends for Cholesterol Medication Adherence, Unweighted and Weighted



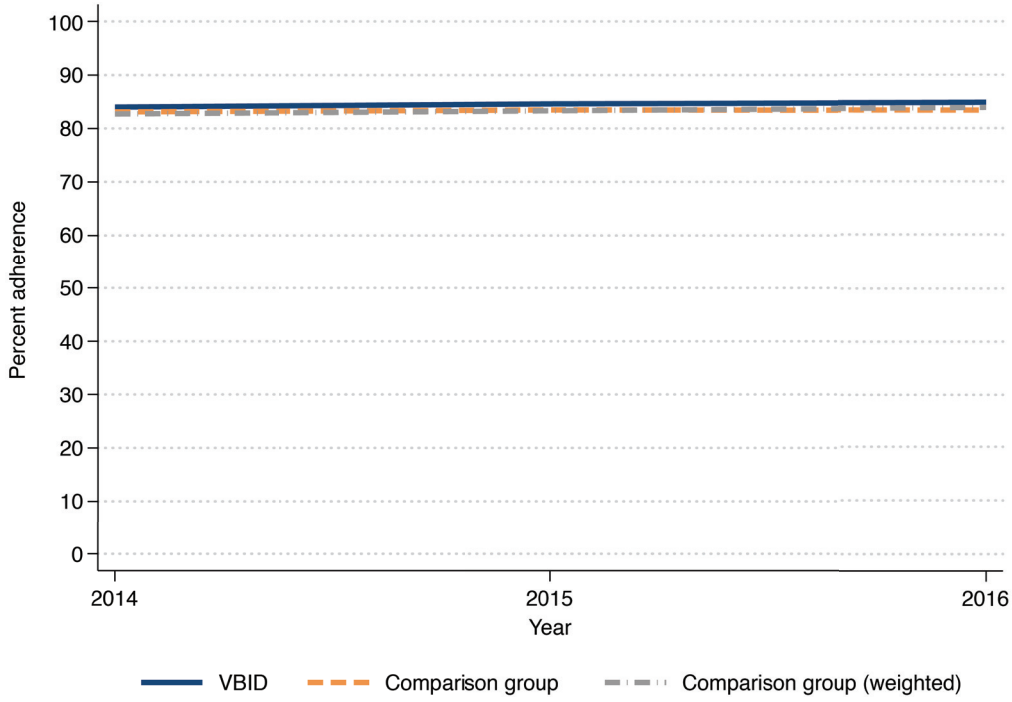
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.93 before weighting and 0.49 after weighting.

For the medication adherence for hypertension measure, the pre-VBID parallel trends assumption was satisfied based on the unweighted regression results (Table H.7), but unweighted trends did not appear parallel by visual inspection (Figure H.8). Weighting improved the plotted trends; therefore, we used the weighted outcome measures in our analysis.

For all other outcome measures, weighting was needed to satisfy the parallel trends regression test, and therefore weighted outcome measures were used in our analyses.

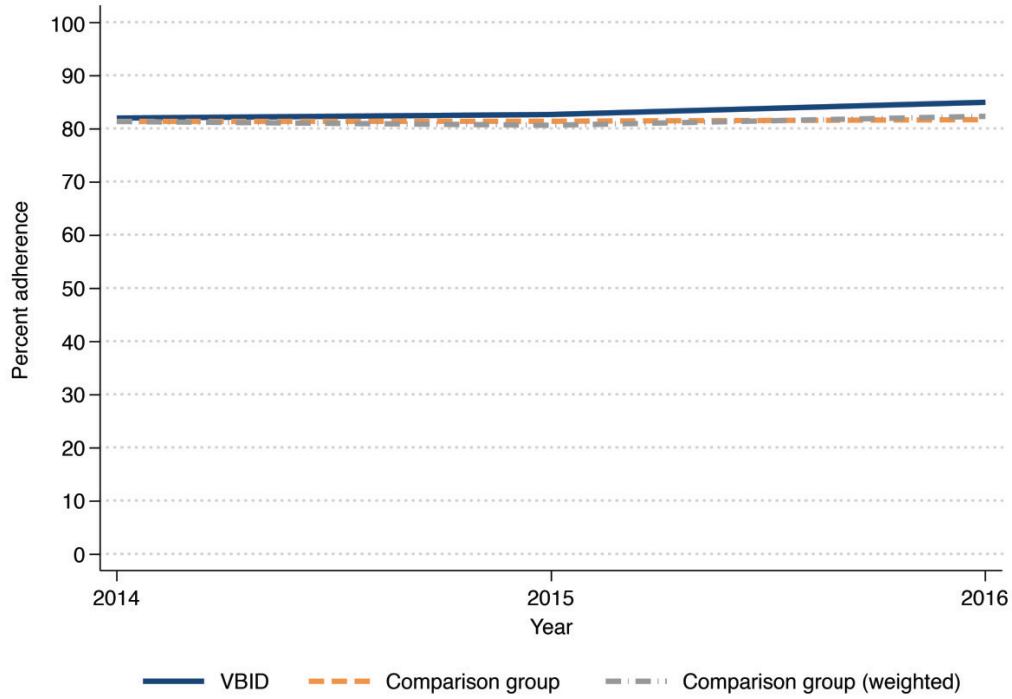
For the breast cancer screening measure, weighting introduced divergence in the pre-VBID trends by visual inspection (Figure H.11), even though it was necessary to satisfy the parallel trends assumption in the regression framework. The discrepancy is likely due to the fact that our plots are unadjusted for covariates, whereas we adjust for covariates when testing for trends in a regression framework. We used the weighted breast cancer screening measures in our main analyses and confirmed that results were unchanged in a sensitivity analysis (see Table H.9).

Figure H.8. Pre-VBID Trends for Hypertension Medication Adherence, Unweighted and Weighted



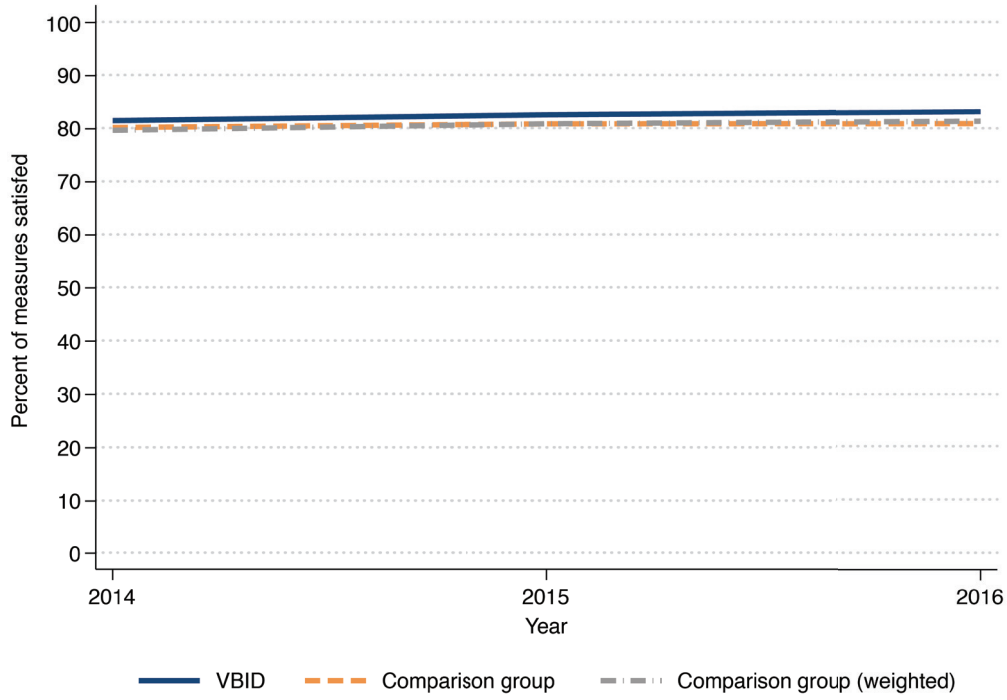
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.71 before weighting and 0.55 after weighting.

Figure H.9. Pre-VBID Trends for Diabetes Medication Adherence, Unweighted and Weighted



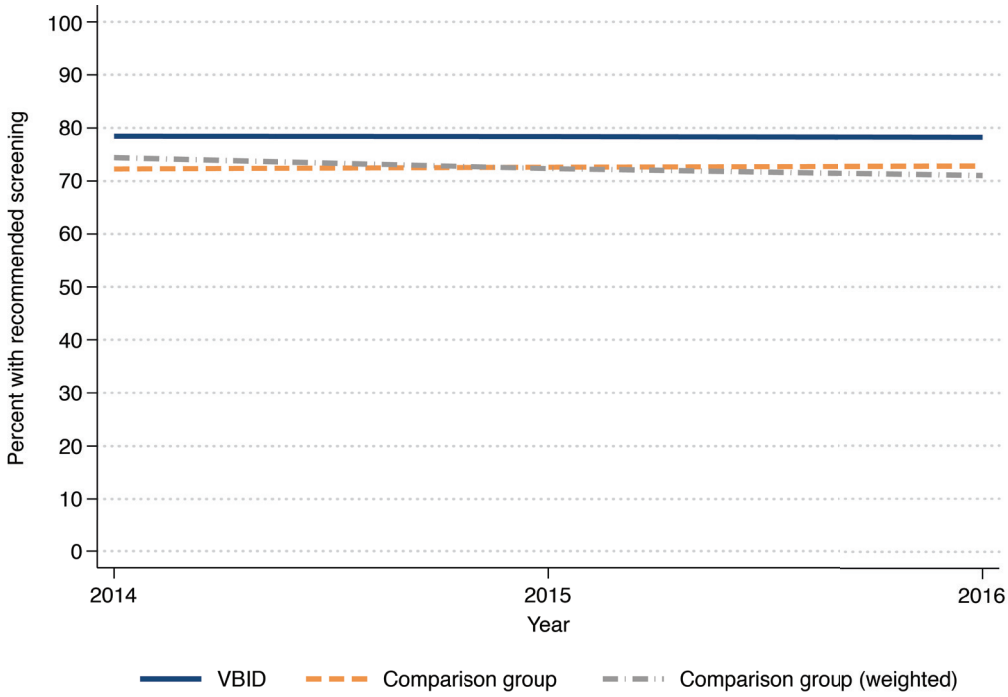
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.03 before weighting and 0.23 after weighting.

Figure H.10. Pre-VBID Trends for Percent Eligible Measures Satisfied, Unweighted and Weighted



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.05 before weighting and 0.93 after weighting.

Figure H.11. Pre-VBID Trends for Breast Cancer Screening, Unweighted and Weighted



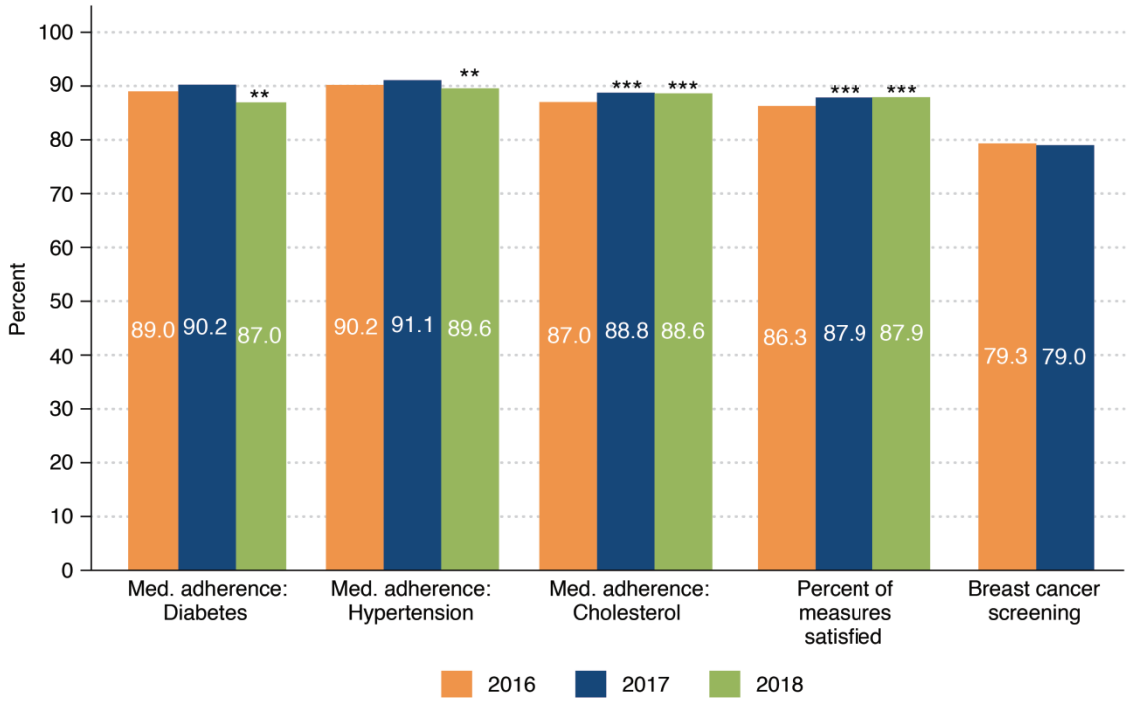
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.02 before weighting and 0.53 after weighting.

In summary, analyses using the medication adherence for cholesterol measure were unweighted, whereas all other analyses were weighted to satisfy the parallel trends assumption.

Descriptive Statistics

Figures H.12 presents descriptive findings illustrating how the beneficiary-level quality/adherence measures changed before and after the implementation of VBID among VBID beneficiaries. These estimates are unadjusted for covariates and do not necessarily reflect the causal effect of VBID. Figure H.12 shows that between 2016 and 2018, among VBID beneficiaries, medication adherence for diabetes and hypertension medications decreased significantly, though these changes were small in magnitude (2-percentage point decrease for diabetes medications, 0.6-percentage point decrease for hypertension medications). Medication adherence for cholesterol-lowering medications increased significantly, as did the overall percentage of medication adherence measures that beneficiaries satisfied, though again the magnitude of these changes was small (1.6-percentage points each). There was no significant change in breast cancer screening rates.

Figure H.12. Descriptive Statistics for Beneficiary-Level Quality/Adherence Measure Indexes, VBID Beneficiaries Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively from paired t-tests comparing each year with 2016. Sample sizes for paired t-tests range from 1,466 to 22,523 unique beneficiaries depending on the outcome measure. See Appendix H for details.

Results

Regression results are shown in Table H.8. VBID had no significant effect on any of the beneficiary-level quality/adherence measures, as evidenced by the fact that all coefficients on the Treatment Contract x Year interaction terms in Table H.8 (i.e., the estimated effect of VBID) were not statistically significantly different from zero. Results of a sensitivity analysis comparing weighted and unweighted regression results for the breast cancer screening measure are shown in Table H.9. In both models, VBID had no significant effect on breast cancer screening.

Table H.8. Beneficiary-Level Quality/Adherence Analyses: Difference-in-Differences Regression Coefficients

Variable	Medication Adherence: Diabetes	SE	Medication Adherence: Hypertension	SE	Medication Adherence: Cholesterol	SE	Percent Eligible Measures Satisfied	SE	Breast Cancer Screening	SE
Treatment contract × year (γt)	**0.04**	**0.07**	**0.01**	**0.04**	**–0.00**	**0.04**	**–0.00**	**0.00**	**0.00**	**0.04**
2015 (2014 reference)	0.01	0.04	0.02	0.02	0.08***	0.02	0.01***	0.00	−0.04*	0.02
2016	0.15***	0.04	0.07**	0.02	0.15***	0.02	0.02***	0.00	−0.07**	0.03
2017	0.24***	0.06	0.17***	0.03	0.36***	0.03	0.04***	0.00	−0.04	0.04
2018	0.23***	0.06	0.29***	0.03	0.48***	0.04	0.05***	0.00	–	
HCC	−0.14***	0.02	−0.25***	0.01	−0.10***	0.01	−0.03***	0.00	−0.03**	0.01
Female	−0.14***	0.05	−0.02	0.02	−0.19***	0.02	−0.01***	0.00	–	
Age	−0.00	0.00	0.00	0.00	0.01***	0.00	0.00***	0.00	−0.01**	0.01
LIS	0.15	0.10	0.02	0.05	0.06	0.05	0.009**	0.00	−0.11	0.08
Dual	−0.05	0.12	0.03	0.06	0.12**	0.06	0.01	0.00	−0.16*	0.09
Disabled	−0.09	0.07	−0.11***	0.03	−0.12***	0.03	−0.02***	0.00	−0.36***	0.05
Black (white, reference)	−0.70***	0.12	−0.36***	0.06	−0.61***	0.07	−0.09***	0.01	0.71***	0.14
Hispanic	−0.31	0.57	−0.60***	0.12	−0.60***	0.17	−0.08***	0.01	−0.32	0.30
Asian/Pacific Islander	0.10	0.27	−0.25***	0.08	0.01	0.15	−0.02**	0.01	−0.08	0.21
American Indian/ Alaskan Native	−21.61**	9.30	−0.50	1.16	1.21	1.98	−0.12	0.17	12.97	10.47
Multiple	−1.82	2.53	−1.41	1.00	−1.23	1.66	−0.19*	0.11	−1.88	3.59
ESRD	0.25	0.22	−0.39***	0.09	−0.07**	0.06	−0.03***	0.01	−0.17	0.13
Plan OOP max	0.00	0.00	−0.00***	0.00	−0.00	0.00	−0.00***	0.00	−0.00**	0.00
Plan premium	0.00	0.00	0.00***	0.00	0.00**	0.00	0.00***	0.00	0.00	0.00
Parent org. fixed effects	Yes	—	Yes	—	Yes	—	Yes	—	Yes	—
Constant	1.83***	0.34	2.17***	0.15	0.99***	0.17	0.80***	0.01	2.25***	0.43
Mean	0.83		0.85		0.82		0.83		0.75	
SD	0.37	—	0.36	—	0.39	—	0.33	—	0.43	—
Range	0–1		0–1		0–1		0–1		0–1	
N	23,522		104,604		78,006		302,406		33,466	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. See Table H.7 for weighting strategy.

Table H.9. Sensitivity Analysis for Breast Cancer Screening: Weighted and Unweighted Regression Coefficients

Variable	Breast Cancer Screening: Unweighted	SE	Breast Cancer Screening: Weighted	SE
Treatment contract × year (γt)	**0.02**	**0.04**	**0.00**	**0.04**
2015 (2014 reference)	0.02	0.02	−0.04*	0.02
2016	0.04	0.03	−0.07**	0.03
2017	0.04	0.04	−0.04	0.04
HCC	−0.03**	0.01	−0.03**	0.01
Age	−0.01**	0.01	−0.01**	0.01
LIS	−0.11	0.08	−0.11	0.08
Dual	−0.17*	0.09	−0.16*	0.09
Disabled	−0.36***	0.05	−0.36***	0.05
Black (white, reference)	0.70***	0.14	0.71***	0.14
Hispanic	−0.32	0.30	−0.32	0.30
Asian/Pacific Islander	−0.06	0.21	−0.08	0.21
American Indian/ Alaskan Native	12.96	10.47	12.97	10.47
Multiple	−1.72	3.58	−1.88	3.59
ESRD	−0.19	0.12	−0.17	0.13
Plan OOP Max	−0.00**	0.00	−0.00**	0.00
Plan premium	−0.00	0.00	0.00	0.00
Parent org. fixed effects	Yes	—	Yes	—
Constant	2.18***	0.43	2.25***	0.43
Mean	0.76		0.75	
SD	0.43	—	0.43	—
Range	0–1		0–1	
N	33,466		33,466	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.

Limitations

Missing Data

For each of the quality/adherence measures studied, there are specific criteria for measure eligibility (i.e., inclusion in the measure denominator). Therefore, it is not unexpected that some of our VBID beneficiaries might be in the denominator for a given measure one year but not the following year, or that their matched comparison beneficiary might not be in the denominator of the same measure in the same year that they are. We required that, for a given VBID beneficiary to be included in our analysis in a specific year, both the VBID beneficiary and his or her matched comparison beneficiary had to be in the denominator of the outcome measure of interest (i.e., both had to have nonmissing data). Note that for the percent eligible measures satisfied,

matched VBID-comparison beneficiary pairs were included as long as both members of the pair were in the denominator for at least one of the medication adherence measures. Therefore, the sample size for this measure exceeds that of the other medication adherence measures.

These missing data patterns also had implications for reweighting parallel trends (where necessary). Since our weighting method relied on calculating differences in outcome measures for the same beneficiaries in adjacent years, this created some challenges for beneficiaries with missing quality/adherence measures in some years. To address this, we imputed values of outcome measures in years where they were missing using the values of the same outcome measures in other years before the start of the VBID model test (i.e., 2014–2016). This allowed us to calculate propensity scores to reweight trends.

Other Limitations

For breast cancer screening, we were only able to assess effects of VBID in 2017 because insufficient data had been reported for 2018 at the time of this report. Also, we were only able to examine performance on the medication adherence measures for MA-PD PBPs, because these measures were not reported for beneficiaries enrolled in MA-only PBPs.

Health Status: Self-Reported Health Status

Study Population and Outcome Measures

Our measures for the self-reported health status analyses were drawn from the Medicare HOS. All Medicare MAOs are required to participate in the HOS, a longitudinal survey that is fielded annually to a random sample of MA beneficiaries in an overlapping cohort design. Each spring a random sample of MA beneficiaries is drawn from participating MAOs with a minimum of 500 enrollees and surveyed; the same cohort receives a follow-up survey two years later.

Our sample consisted of VBID-eligible beneficiaries in VBID-participating PBPs who were also cohort 18 HOS respondents ("VBID beneficiaries") and matched comparison beneficiaries who were also cohort 18 HOS respondents ("comparison beneficiaries"). Cohort 18 was selected because respondents received their baseline survey in spring 2015 (pre-VBID) and their follow-up survey in spring 2017 (shortly after VBID implementation).

Given the relatively small number of HOS respondents in our beneficiary sample, we included *any* VBID-eligible, VBID-participating respondents and *any* matched comparison respondents in our sample. That is, in contrast to the other beneficiary-level analyses, we did *not* restrict our sample only to VBID and comparison beneficiaries who were matched with each other.

There was substantial attrition among all HOS respondents between the baseline and follow-up surveys. Only 60 percent of individuals who responded to the baseline survey also responded to the follow-up survey. We therefore created two analytic samples for our analyses: a balanced panel (i.e., beneficiaries who provided both baseline and follow-up responses for at least one of

our HOS measures of interest)[2] and an unbalanced panel (i.e., beneficiaries who provided a response to at least one of our HOS measures of interest in any wave of the survey). We report the balanced panel results in Chapter 7 as our main results, but because beneficiaries responding to both survey waves could differ significantly from the overall HOS sample in ways that might affect the external validity of our results, we also describe the unbalanced panel analyses and results here as a sensitivity analysis.

We used the following outcome measures derived from the HOS:

1. The Veterans RAND 12-Item Health Survey (VR-12) Physical Component Score (PCS)
2. VR-12 Mental Component Score (MCS)
3. Limitations in Activities of Daily Living (ADLs)[3]
4. Limitations in Independent Activities of Daily Living (IADLs)[4]
5. Number of days of poor physical health
6. Number of days where health limited usual activities
7. Smoking status.[5]

The VR-12 PCS and MCS are validated self-reported measures of physical and mental health status, respectively. Because the ADL and IADL measures are composites of individual items, we imputed the value of a missing item as the mean value of that item for comparison beneficiaries in the same wave of the survey. If a beneficiary had more than three items missing for a given composite measure (i.e., ADL or IADL), we assigned a missing value for the composite measure as a whole.

Propensity Score Weighting

Because VBID and comparison beneficiaries in our HOS sample are not necessarily matched with each other, and because we have only a single pre-VBID data point for each beneficiary and cannot test parallel trends, we used propensity score weighting to ensure covariate balance across our VBID and comparison beneficiaries. Propensity score weights were estimated using a

---

[2] In our balanced panel, item non-response was uncommon −86 percent of the panel responded to all questions in 2015 and 88 percent of the panel responded to all questions in 2017.

[3] Limitations in ADLs is a composite measure that includes items 10a–10f on the Medicare HOS 2015 3.0 survey (i.e. difficulties with bathing, dressing, eating, getting in or out of chairs, walking, using the toilet). Each item has a value of 1–3 (1 indicates no difficulty, 2 indicates some difficulty, 3 indicates respondent is unable to do this activity), therefore the total value of the composite can range from 6 to 18.

[4] Limitations in IADLs is a composite measure that includes items 11a–11c on the Medicare HOS 2015 3.0 survey (i.e. difficulties with preparing meals, managing money, taking medications as prescribed). Each item has a value of 1–3 (1 indicates no difficulty, 2 indicates some difficulty, 3 indicates respondent is unable to do this activity); therefore, the total value of the composite can range from 3 to 9.

[5] In the HOS, the raw smoking variable has 3 possible values: "Not at all," "Some days," and "Every day". We recoded this as a binary variable, such that "Not at all" was coded 0 and all other values were coded as 1.

logistic regression predicting whether a beneficiary belonged to a VBID-participating PBP using the following beneficiary-level covariates: sex, age, dual eligibility status, LIS eligibility status, disabled status, race/ethnicity, and diagnosis of ESRD. This was done separately for our balanced and unbalanced panel samples. Table H.10 presents the covariate balance for our samples before and after propensity score weighting. The table shows that, even before propensity score weighting, covariates were well-balanced across our VBID and comparison beneficiaries, with an average absolute standardized difference of only 0.065 for our balanced panel and 0.060 for our unbalanced panel. Propensity score weighting decreased this difference further to 0.042.

Table H.10. Standardized Differences of HOS Respondent Characteristics Before and After Propensity Score Weighting

Measure	Balanced Panel		Unbalanced Panel	
	Before	After	Before	After
Age	−0.092	0.033	−0.060	0.045
Percent Female	0.067	0.038	0.086	0.050
Race/Native American	−0.100	−0.056	−0.191	0.031
Race/Asian Pacific Islander	0.072	0.059	0.016	−0.150
Race/black	−0.057	−0.058	−0.052	−0.001
Race/Hispanic	0.025	0.058	−0.027	0.031
Multiple races	−0.209	−0.028	−0.125	0.028
White	0.013	0.004	0.049	0.059
ESRD	−0.099	−0.013	−0.030	0.020
Dually eligible for Medicare and Medicaid	−0.002	−0.055	0.010	−0.027
Low income subsidy	−0.015	−0.020	−0.043	−0.004
Disabled	−0.033	0.085	−0.030	0.059
Average absolute standardized difference, beneficiary characteristics	0.065	0.042	0.060	0.042

Difference-in-Differences Models

The difference-in-differences models for the self-reported health status outcomes have the same form as the beneficiary-level quality/adherence models, described in Equation (H.2), with two exceptions. First, these models were propensity score weighted, as described in the previous section. Second, because we have only two years of data for our self-reported health outcome measures, our year fixed-effects collapse to a single indicator variable for the year 2017.

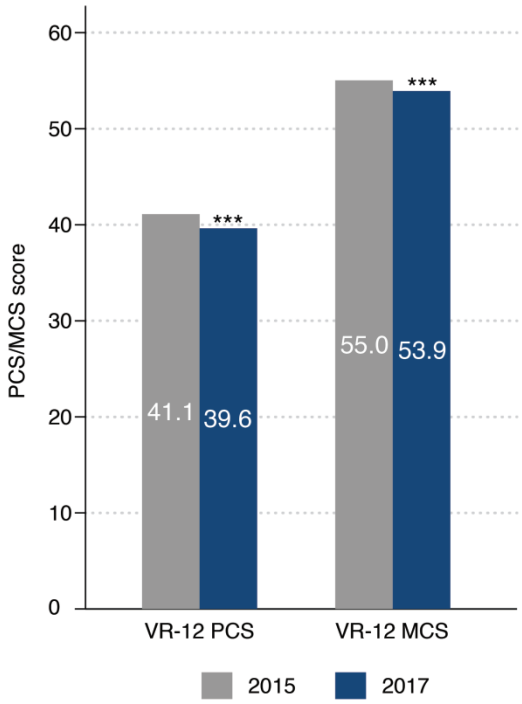
Our unbalanced panel analyses were identical but also included a beneficiary-level random effect. The beneficiary-level random effect was included for the following reasons. First, given we are capturing repeated measurements on individual beneficiaries, rather than treating these

measurements as independent, this approach allows us to account for beneficiary-level effects that are constant within an individual, but differ across individuals. Given the beneficiaries included in this analysis constitute a sample of individuals drawn from a larger population, it is appropriate to model these beneficiary-level effects as random effects rather than fixed effects. One exception to the use of beneficiary-level random effects was our unbalanced panel analysis of the effect of VBID on the probability of smoking: For this outcome measure, we were unable to achieve model convergence when including a beneficiary-level random effect, so this model is estimated without a random effect.

Descriptive Statistics

Figures H.13–H.15 present descriptive findings for changes in self-reported health measures during the study period, unadjusted for covariates. These estimates do not necessarily reflect the causal effect of VBID. From 2015 to 2017, among VBID beneficiaries there was a statistically significant reduction in self-reported physical health status (VR-12 PCS) by 1.5 units and mental health status (VR-12 MCS) by 1.1 units (Figure H.13). For reference, in the Medical Expenditure

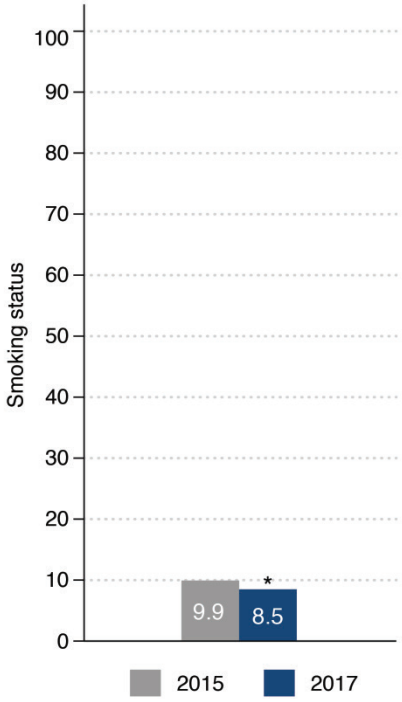
Figure H.13. Descriptive Statistics for HOS Self-Reported Physical and Mental Health (VR-12 PCS, MCS; VBID Beneficiaries Only)



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively from paired t-tests. To be included in paired t-test observations, beneficiaries must have observed data in 2015 and 2017. Sample sizes are 508 for VR-12 PCS and 506 for VR-12 MCS. In this sample, the range for the VR-12 PCS score is approximately 0–63 with a standard deviation of 12, and the range for the VR-12 MCS score is approximately 8–71 with a standard deviation of 10 (see Appendix H).

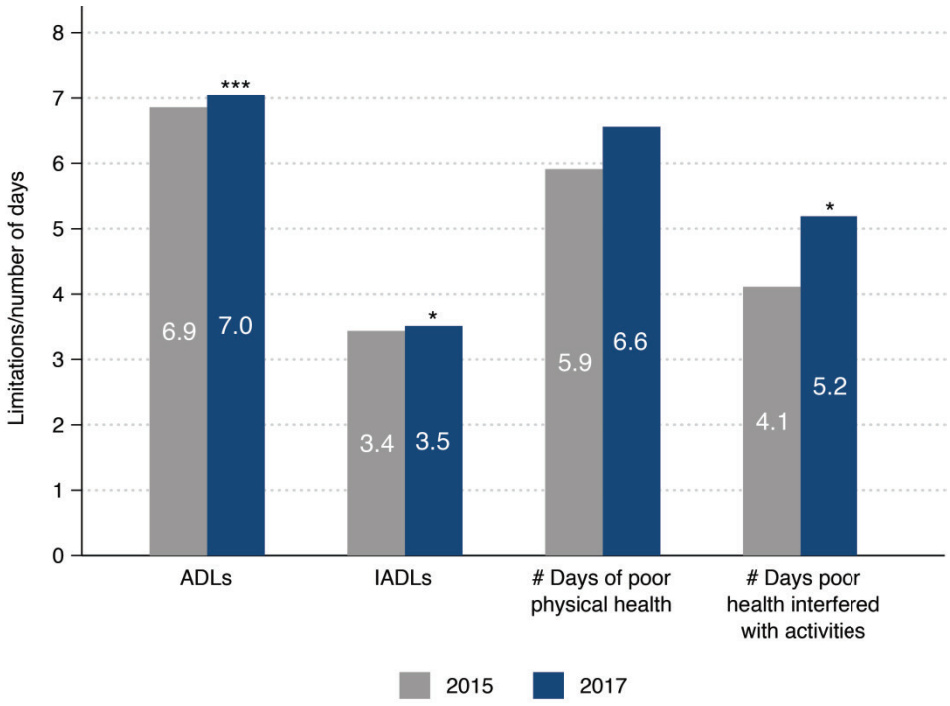
Panel Survey, a nationally representative survey of the general U.S. population, diabetes was associated with a decrease in the VR-12 PCS of 13.7 units relative to having no medical conditions and heart disease with a decrease in the VR-12 PCS of 14.2 units; depression was associated with a decrease in the VR-12 MCS by 15.9 units relative to having no medical conditions (Selim et al., 2009). There was also a statistically significant decrease in the probability of smoking, from 9.9 to 8.5 percent, though this change was only significant at the 10 percent level (Figure H.14). There was a statistically significant increase in ADL limitations, though this was very small in magnitude (Figure H.15). Finally, there were statistically significant increases in IADL limitations and the number of days in which poor health interfered with activities, though only at the 10-percent level (Figure H.15). There were no statistically significant changes in the number of days of poor physical health (Figures H.15). The observed reductions in self-reported physical and mental health status, and the observed increases in ADL and IADL limitations and days in which poor health interfered with activities, could reflect aging of the survey cohort. These descriptive trends alone do not tell us whether VBID was responsible for these changes.

Figure H.14. Descriptive Statistics for Smoking Status, VBID Beneficiaries Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from paired t-tests. Sample size is 484.

Figure H.15. Descriptive Statistics for Other HOS Self-Reported Health Measures, VBID Beneficiaries Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from paired t-tests. To be included in the paired t-test, observations must have observed data in 2015 and 2017. Sample sizes range from 421 to 502.

Results

Tables H.11 and H.12 present estimates of the effect of VBID on beneficiary self-reported health measures from the HOS in the balanced panel model. Tables H.13 and H.14 present estimates for the unbalanced panel model. VBID had no significant effects on any of the self-reported health status measures studied, for both the balanced and unbalanced panels.

Table H.11. HOS Self-Reported Health Analyses: Difference-in-Differences Regression Coefficients, Balanced Panel (Part 1)

Variable	VR-12 PCS	SE	VR-12 MCS	SE	ADL Limits	SE	IADL Limits	SE
Treatment beneficiary × year (γt)	**0.63**	**0.45**	**−0.06**	**0.54**	**−0.11**	**0.07**	**−0.03**	**0.06**
2017 (2015 reference)	−1.37***	0.35	−0.99**	0.38	0.20***	0.04	0.03	0.05
Female	−1.06**	0.50	−0.79	0.56	0.15**	0.07	−0.17*	0.09
Age	−0.31***	0.06	−0.03	0.03	0.05***	0.01	0.04***	0.004
LIS	−5.69***	1.61	−1.27	1.22	0.36	0.21	0.03	0.11
Dual	1.26	2.06	0.45	1.56	0.23	0.29	0.42**	0.18
Disabled	−10.17***	1.24	−5.41***	1.34	1.00***	0.12	0.53***	0.09
Black (white, reference)	4.21	2.46	4.61	3.13	−0.75	0.54	−0.25	0.47
Hispanic	−13.80***	1.59	−22.38***	6.36	2.18**	1.01	0.40*	0.19
Asian/Pacific Islander	5.35*	2.82	0.89	2.52	−0.52**	0.22	−0.06	0.19
American Indian/Alaskan Native	−320.21	314.76	−52.24	149.77	−3.38	34.73	15.63	19.14
Multiple	−30.03	63.83	−236.61***	69.19	18.37	13.96	8.27	14.24
ESRD	−7.57**	3.16	0.41	1.84	0.48	0.35	0.18	0.20
Plan OOP max	0.001	0.001	0.001***	0.000	−0.0002***	0.000	−0.0001***	0.000
Plan premium	0.001	0.002	0.005**	0.002	−0.0004*	0.000	−0.0003*	0.000
Parent org. fixed effects	Yes	—	Yes	—	Yes	—	Yes	—
Constant	62.88***	5.97	56.46***	2.27	3.45***	0.73	1.20**	0.50
Mean	38.44	—	54.24		7.01		3.51	
SD	12.47		10.00	—	1.72	—	1.14	—
Range	3.45–63.16		8.37–71.41		6–18		3–9	
N	2,218		2,206		2,164		2,156	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. All models propensity score weighted to improve covariate balance. N reflects the number of beneficiary-year observations.

Table H.12. HOS Self-Reported Health Analyses: Difference-in-Differences Regression Coefficients, Balanced Panel (Part 2)

Variable	No. of Days Poor Physical Health	SE	No. of Days Health Limited Activities	SE	Smoking	SE
Treatment beneficiary × year (γt)	**−0.29**	**0.71**	**0.67**	**0.62**	**0.005**	**0.08**
2017 (2015 reference)	0.84	0.49	0.32	0.31	−0.03	0.08
Female	1.13***	0.35	−0.02	0.48	−0.15	0.24
Age	0.04	0.04	0.03	0.04	−0.08**	0.03
LIS	3.62***	1.26	3.13**	1.17	1.02*	0.53
Dual	−0.10	1.66	0.43	1.83	−0.24	0.69
Disabled	6.42***	0.69	5.92***	0.97	0.59**	0.28
Black (white, reference)	−3.12	2.57	−2.64	3.58	0.23	0.99
Hispanic	11.07**	5.17	5.86	6.49	−6.08	5.49
Asian/Pacific Islander	−3.55**	1.32	−2.01	1.37	−2.00**	0.79
American Indian/Alaskan Native	329.25	196.95	−126.08	128.05	−34.27	146.41
Multiple	22.81	77.55	75.95	74.66	−4.10	24.15
ESRD	2.22	3.18	3.27	3.36	−0.60	0.90
Plan OOP max	−0.0004	0.000	−0.001**	0.000	−0.0003**	0.000
Plan premium	0.001	0.001	−0.0004	0.001	−0.0005	0.000
Parent org. fixed effects	Yes	—	Yes	—	Yes	—
Constant	2.02	3.37	3.85	3.81	4.53*	2.65
Mean	6.89		5.00		0.09	
SD	10.83	—	9.81	—	0.29	—
Range	0–88		0–88		0–1	
N	2,048		1,858		2,032	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. All models propensity score weighted to improve covariate balance. N reflects the number of beneficiary-year observations.

Table H.13. HOS Self-Reported Health Analyses: Difference-in-Differences Regression Coefficients, Unbalanced Panel (Part 1)

Variable	VR-12 PCS	SE	VR-12 MCS	SE	ADL Limits	SE	IADL Limits	SE
Treatment beneficiary × year (γt)	**0.59**	**0.38**	**–0.13**	**0.59**	**–0.05**	**0.07**	**–0.01**	**0.06**
2017 (2015 reference)	−0.96***	0.32	−0.70	0.43	0.15***	0.03	0.01	0.05
Female	−0.92***	0.30	−0.57	0.47	0.14**	0.06	−0.13**	0.06
Age	−0.37***	0.05	−0.05**	0.02	0.05***	0.01	0.03***	0.004
LIS	−2.31*	1.33	−2.83	2.65	0.24	0.15	0.04	0.11
Dual	−0.34	1.07	0.03	2.67	0.47***	0.11	0.43**	0.09
Disabled	−11.13***	0.98	−6.70***	0.78	1.17***	0.14	0.55***	0.08
Black (white, reference)	5.40**	2.38	5.30*	2.90	−0.65	0.45	−0.39	0.29
Hispanic	−7.09**	3.08	−14.11**	7.05	1.00**	0.49	0.24	0.33
Asian/Pacific Islander	0.19	1.20	0.30	2.35	−0.62***	0.20	−0.31***	0.08
American Indian/Alaskan Native	−240.83	191.16	−133.81	180.59	13.12	18.36	18.47	15.57
Multiple	−94.55*	51.51	−178.22***	50.40	13.77	9.38	8.34	5.53
ESRD	−5.56***	1.53	0.42	0.94	0.19	0.18	0.07	0.21
Plan OOP max	0.001**	0.000	0.001***	0.000	−0.000***	0.000	−0.000***	0.000
Plan premium	0.003*	0.002	0.003***	0.001	−0.000***	0.000	−0.000	0.000
Parent org. fixed effects	Yes	—	Yes	—	Yes	—	Yes	—
Constant	66.08***	5.24	59.30***	2.04	3.27***	0.65	0.95***	0.35
Mean	38.09		53.93		7.02		3.53	
SD	12.53	—	10.38	—	1.73	—	1.15	—
Range	−0.56 to 63.16		8.37–71.52		6–18		3–9	
N	3,065		3,048		2,965		2,957	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. All models propensity score weighted to improve covariate balance. N reflects the number of beneficiary-year observations.

Table H.14. HOS Self-Reported Health Analyses: Differences-in-Differences Regression Coefficients, Unbalanced Panel (Part 2)

Variable	No. of Days Poor Physical Health	SE	No. of Days Health Limited Activities	SE	Smoking	SE
Treatment beneficiary × year (γt)	**−0.04**	**0.70**	**0.66**	**0.52**	**0.03**	**0.10**
2017 (2015 reference)	0.51	0.45	0.06	0.31	−0.11	0.09
Female	0.86**	0.37	0.01*	0.47	−0.04	0.20
Age	0.08***	0.02	0.05	0.03	−0.07***	0.02
LIS	2.43***	0.82	3.12***	1.06	0.77**	0.38
Dual	1.45	1.03	0.77	2.00	−0.19	0.46
Disabled	7.04***	0.66	5.87***	0.78	0.55*	0.29
Black (white, reference)	−2.03	2.47	−4.30	2.67	−0.14	0.83
Hispanic	7.97	5.52	8.08	6.49	−0.75	1.12
Asian/Pacific Islander	−0.21	1.43	−1.35	1.43	−3.16**	1.45
American Indian/Alaskan Native	63.76	233.84	−19.87	83.62	−40.42	95.65
Multiple	51.24	76.91	89.65*	47.50	−3.06	13.50
ESRD	1.02	1.84	2.75*	1.59	−0.57	0.74
Plan OOP max	−0.001***	0.000	−0.000**	0.000	−0.000**	0.000
Plan premium	−0.002	0.001	−0.001	0.001	−0.001**	0.000
Parent org. fixed effects	Yes	—	Yes	—	Yes	—
Constant	0.84	2.53	0.89	2.84	4.13**	1.64
Mean	6.96		5.17		0.10	
SD	10.77	—	9.74	—	0.30	—
Range	0–88		0–88		0–1	
N	2,820		2,575		2,867	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. All models propensity score weighted to improve covariate balance. N reflects the number of beneficiary-year observations.

Limitations

Because the HOS follow-up survey was done in the spring of 2017, just several months after the implementation of the VBID model test, it is possible that this was too short an exposure to VBID to produce significant changes in chronic DM and functional status. As noted above, in most cases we were unable to compare specific VBID beneficiaries to their designated matched comparison beneficiaries because often only one member of a pair was surveyed. Finally, cohort 18 completed only one survey wave in the pre-VBID period; therefore, we were unable to assess pre-VBID trends in the outcomes of interest. To address these last two concerns, we used propensity score weighting to balance observable characteristics across groups and mitigate confounding, as described above.

Health Status: Hierarchical Condition Category and Prescription Drug Hierarchical Condition Category

Study Population and Outcome Measures

Additional health status outcome measures included beneficiary HCC and RxHCC scores. As with previously described beneficiary-level outcome measures, our sample included VBID beneficiaries and comparison beneficiaries, observed during measurement years 2014–2018. For a given year *t*, we restricted our sample to beneficiaries who were enrolled for all of CY *t-1*, because beneficiaries who were enrolled for a shorter duration were assigned a "new enrollee" risk score that did not reflect their actual health care utilization during that time period. To account for missing HCC/RxHCC scores, we applied the same sample restrictions as in our beneficiary-level quality/adherence analyses.

Note that, by convention, risk scores are named according to the year in which they are reported, but there is a one-year lag between when the diagnoses that form the basis of the risk scores are coded, and when the risk scores are reported. For example, the 2018 final HCC score is reported in 2018, but is based on diagnoses coded during CY 2017. For simplicity and ease of interpretation, in our analyses we relabel the year of the risk score measures to correspond to the year in which diagnoses were recorded; i.e., the 2018 HCC score is relabeled as the 2017 measurement year HCC score. At the time of this analysis, 2019 final HCC scores (i.e., final 2018 measurement year scores) were not yet available, so we used midyear scores. Midyear HCC scores for year *t* are typically made available in July of year *t* and are based on preliminary diagnosis data from the measurement year *t-1*.

Difference-in-Differences Models

The difference-in-differences models for health status are the same as those for the beneficiary-level quality/adherence analyses, described in Equation (H.2).

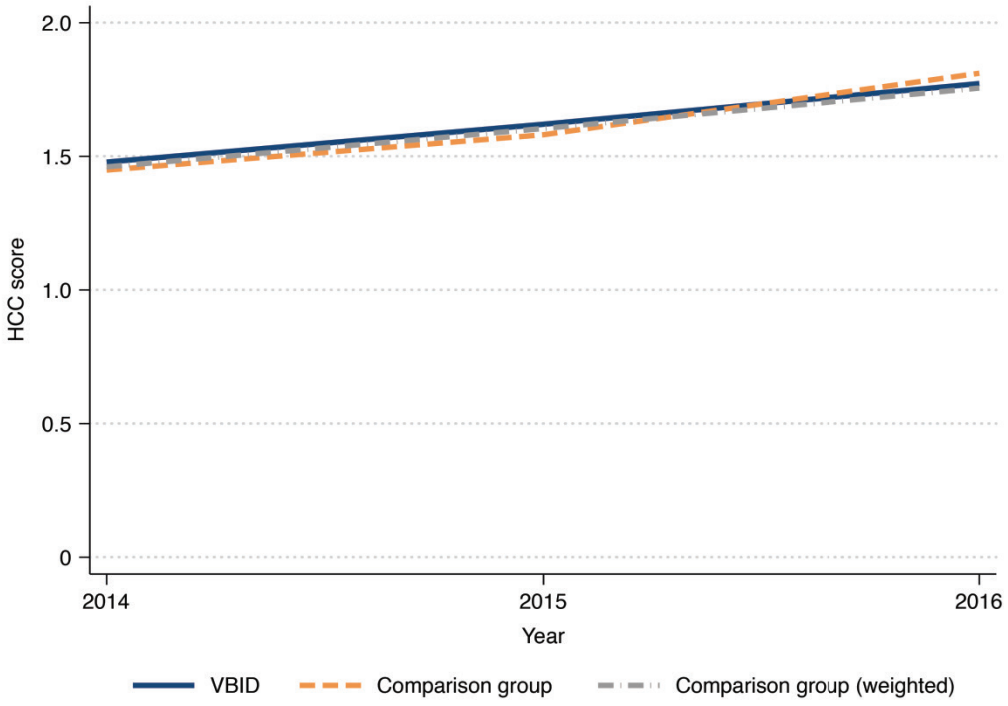
Results for the tests of the parallel trends assumption are shown in Table H.15. For both HCC and RxHCC scores, the slopes of the pre-VBID trends of VBID beneficiaries differed significantly from those of comparison beneficiaries when tested in a regression framework, even though the magnitude of the difference was small. This was likely driven by the large available sample size for each of these measures. We corrected these trends using the propensity score weighting approach described in Appendix D. Figures H.16 and H.17 show the plots of the pre-VBID trends before and after weighting.

Table H.15. HCC/RxHCC: Summary of Parallel Trends Tests Results

Measure	Trend Years	Unweighted Parallel Trends Test Coefficient	Unweighted Parallel Trends Test P	Weighted Parallel Trends Test Coefficient	Weighted Parallel Trends Test P	Model Used
HCC	2014–2015	0.01	0.00	0.00	0.55	Weighted
	2015–2016	−0.07		0.01		
RxHCC	2014–2015	0.02	0.00	−0.00	0.16	Weighted
	2015–2016	−0.02		0.01		

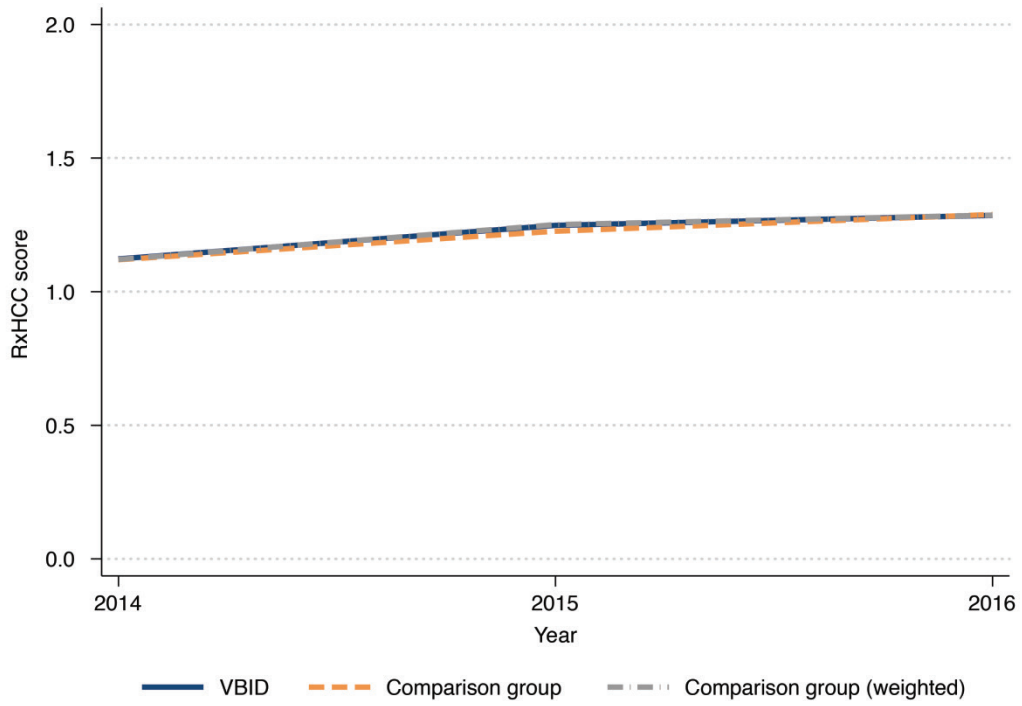
NOTE: p-values correspond to a joint test of significance of the 2014–2015 and 2015–2016 trend coefficients and were derived from a chi-square test.

Figure H.16. Pre-VBID Trends in HCC Scores, Unweighted and Weighted



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.00 before weighting and 0.55 after weighting.

Figure H.17. Pre-VBID Trends in RxHCC Scores, Unweighted and Weighted Using Modified Approach

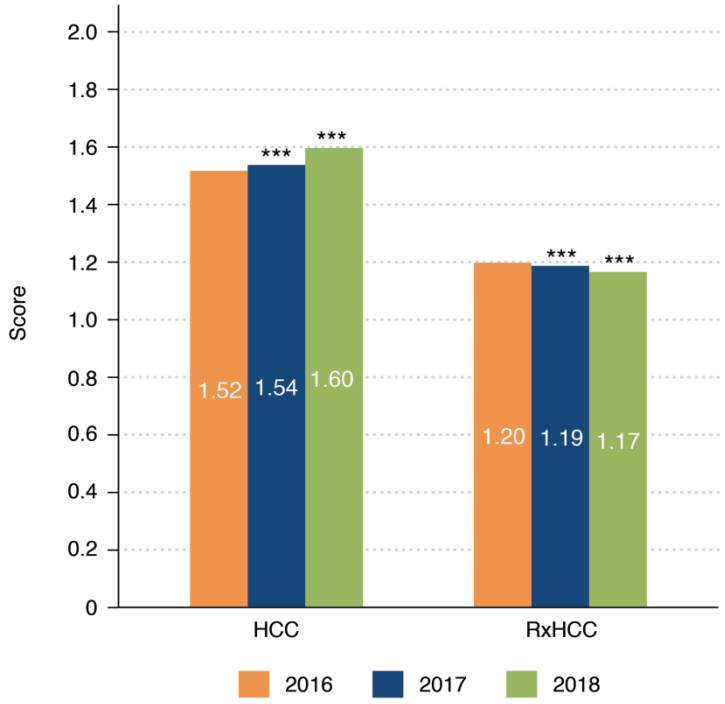


NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.00 before weighting and 0.16 after weighting.

Descriptive Statistics

Figure H.18 presents descriptive findings for changes in HCC and RxHCC scores, during the study period, unadjusted for covariates. These estimates do not necessarily reflect the causal effect of VBID. Figure H.18 shows that from 2016 to 2018, among VBID beneficiaries there was a small but statistically significant increase in HCC scores, and a small but statistically significant decline in RxHCC scores. These changes suggest shifts in the overall medical complexity of VBID beneficiaries during the study period. Specifically, expected overall medical spending increased for VBID beneficiaries increased, but expected prescription drug spending decreased. Because the changes were small in magnitude, and in opposite directions for the HCC and RxHCC scores, it is unclear whether they are clinically meaningful. These descriptive results also do not tell us whether VBID might have contributed to these changes.

Figure H.18. Descriptive Statistics for HCC/RxHCC, VBID Beneficiaries Only



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from paired t-tests comparing each year with 2016. Sample sizes for paired t-tests are 47,015 unique beneficiaries. See Appendix H for details.

Results

Regression results are shown in Table H.16. No significant effect of VBID was found for either HCC or RxHCC scores.

Table H.16. HCC/RxHCC Analyses: Difference-in-Differences Regression Coefficients

Variable	HCC	SE	RxHCC	SE
Treatment beneficiary × year (γt)	**0.00**	**0.00**	**−0.00**	**0.00**
2015 (2014 reference)	0.11***	0.00	0.12***	0.00
2016	0.23***	0.00	0.15***	0.00
2017	0.21***	0.00	0.14***	0.00
2018	0.17***	0.00	0.10***	0.00
Female	−0.15***	0.00	0.03***	0.00
Age	0.03***	0.00	0.00***	0.00
LIS	0.17***	0.01	0.27***	0.01
Dual	0.34***	0.01	0.12***	0.01
Disabled	0.66***	0.01	0.32***	0.00
Black (white, reference)	0.06***	0.02	0.07***	0.02
Hispanic	−0.21***	0.04	0.04*	0.02
Asian/Pacific Islander	−0.37***	0.02	−0.14***	0.01
American Indian/Alaskan Native	0.57	0.47	−0.13	0.33

Variable	HCC	SE	RxHCC	SE
Multiple	0.07	0.27	1.61***	0.40
ESRD	0.48***	0.02	0.44***	0.01
Plan OOP max	−0.00***	0.00	−0.00***	0.00
Plan premium	0.00***	0.00	0.00***	0.00
Parent org. fixed effects	Yes	—	Yes	—
Constant	−0.94***	0.03	0.99***	0.02
Mean	1.63		1.20	
SD	1.22	—	0.68	—
Range	0.11–19.45		0.05–16.14	
N	588,252		588,252	

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Models weighted as in Table H.15. N refers to the number of beneficiary-year observations.

Health Status: Mortality

Study Population and Outcome Measures

Our sample consisted of VBID-eligible beneficiaries age 65 and older in VBID-participating PBPs ("VBID beneficiaries") and their matched comparison beneficiaries (i.e., VBID-eligible beneficiaries age 65 and older in matched comparison plans: "comparison beneficiaries"). We restrict our sample to VBID-matched comparison pairs in which both members of the pair were alive as of January 1, 2017. The period of observation includes years 2017 and 2018, meaning we assessed differences in mortality only *after* the start of VBID and did not consider differences in mortality pre-VBID. Our principal outcome measure was the age at death (measured in years), and beneficiaries were censored at the end of follow-up or if they left our sample due to plan switching or disenrollment (also measured in years).

Analysis

Ideally, we would estimate the effect of VBID on mortality using models for survival analysis, such as the Cox proportional hazard model. A key assumption underlying these models, however, is the proportional hazard assumption: that the effect of any explanatory variable—such as VBID—on mortality must be constant over time. This assumption was examined by plotting Kaplan-Meier curves that show the probability of survival for VBID beneficiaries and their matched comparison beneficiaries over time, stratified by sex and adjusted for the following predictors: race, ESRD status, dual/LIS status, disabled status, and HCC score all as of January 1, 2017. Control-group beneficiaries were frequency weighted to account for the fact that a single comparison beneficiary could be matched to multiple VBID beneficiaries.

Descriptive Statistics

Table H.17 summarizes the unadjusted mortality rate, expressed as the number of deaths per 1,000 beneficiaries, for VBID and comparison beneficiaries in 2017 and 2018. In both years, unadjusted death rates were lower for VBID beneficiaries than comparison beneficiaries;

Table H.17. Number of Deaths per 1,000 VBID or Comparison Beneficiaries, 2017–2018

Year	VBID	Comparison
2017	56.57	83.82
2018	66.66	68.03

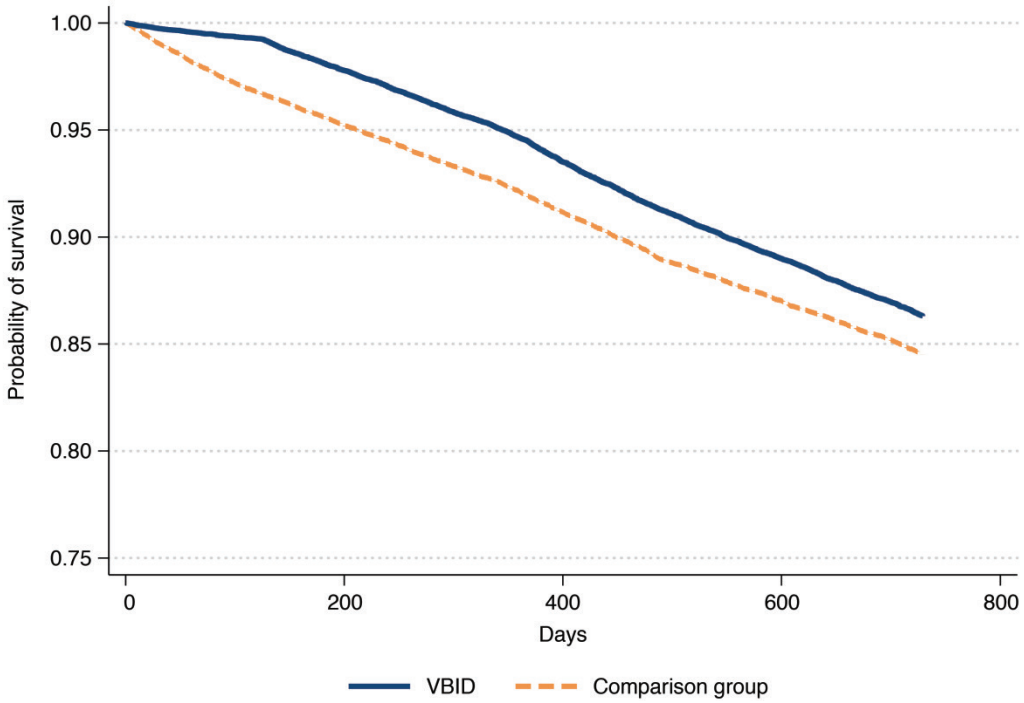
NOTE: Mortality rates are unadjusted for covariates and are calculated as the number of deaths per 1,000 VBID or comparison beneficiaries who were participating in the VBID Model Test as of January 1, 2017. Sample size is 66,454 unique VBID beneficiaries and 55,546 unique comparison beneficiaries.

however, the unadjusted death rate among VBID beneficiaries increased from 2017 to 2018, whereas for comparison beneficiaries the unadjusted death rate decreased. The reasons for these trends in death rates are unclear. These descriptive results do not tell us whether VBID contributed to these changes.

Results

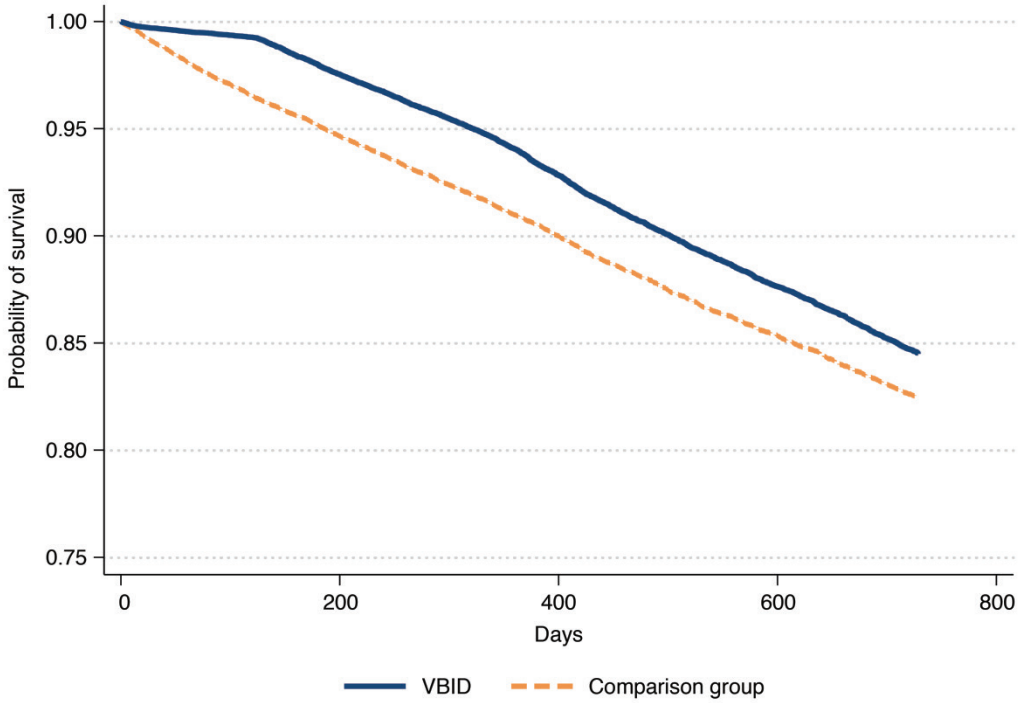
Figures H.19 and H.20 show the Kaplan-Meier curves assessing the validity of the proportional hazard assumption, separately for male and female beneficiaries. The figures show that for both sexes, the effect of VBID is not constant over time, in violation of the proportional hazard assumption. We are therefore unable to obtain valid estimates of the effect of VBID on mortality using standard survival analyses.

Figure H.19. Adjusted Probability of Survival, Female Beneficiaries



NOTE: Kaplan-Meier curve plotting average probability of survival across ages against number of days exposed to the VBID model test, where day 0 is January 1, 2017.

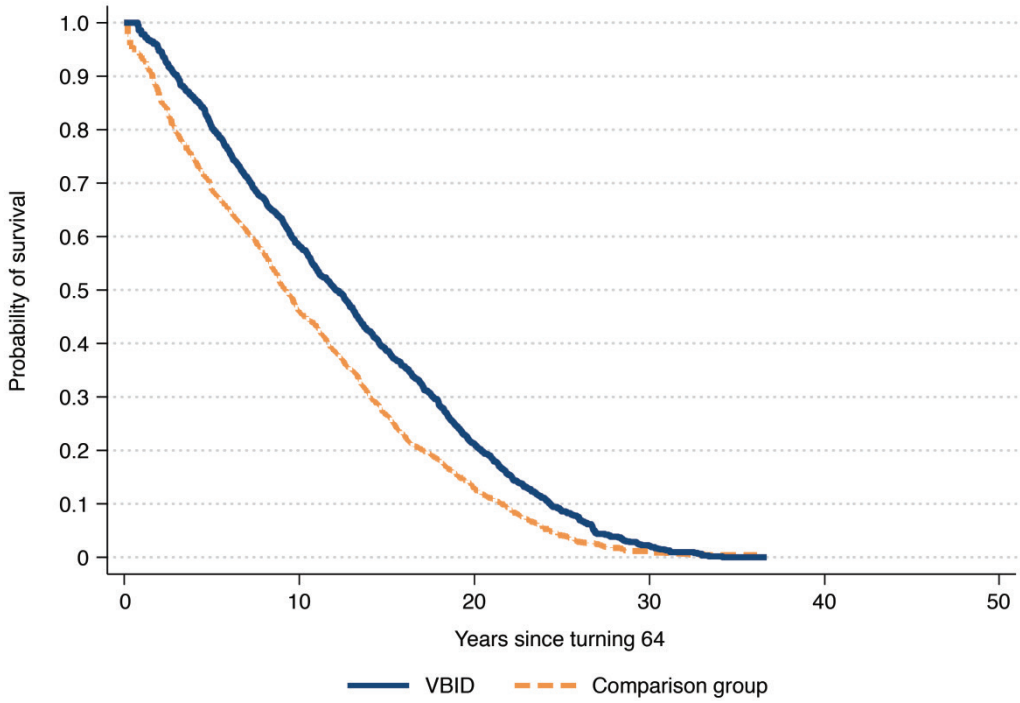
Figure H.20. Adjusted Probability of Survival, Male Beneficiaries



NOTE: Kaplan-Meier curve plotting average probability of survival across ages against number of days exposed to the VBID model test, where day 0 is January 1, 2017.

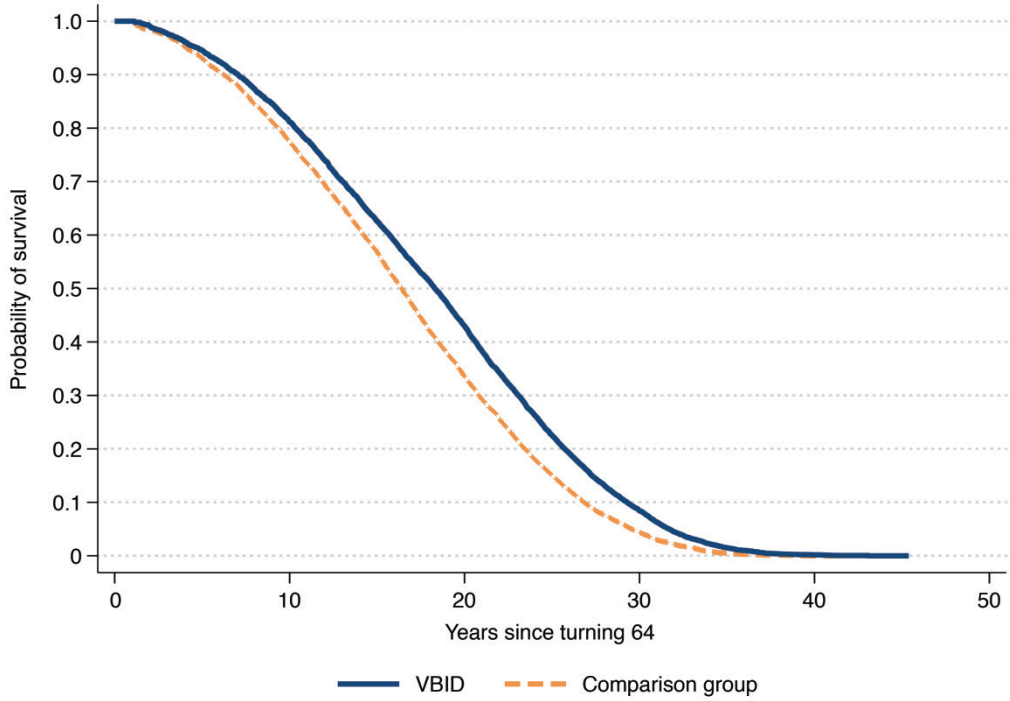
Both Kaplan-Meier curves show that in the early months of the VBID model test (i.e., from January 2017 to approximately March-April 2017), survival among VBID beneficiaries was higher than among comparison beneficiaries. However, the mortality rate among VBID beneficiaries subsequently increased, such that by May 2017 it was similar to the mortality rate among comparison beneficiaries. This finding is similar to the differences and trends in unadjusted mortality rates between VBID and comparison beneficiaries described in Table H.17. This difference in mortality rates was most pronounced among disabled beneficiaries just older than 65. Figures H.21 and H.22 show Kaplan-Meier curves plotting the probability of survival by age and VBID status, separately for disabled and nondisabled beneficiaries. Figure H.21 shows that among disabled beneficiaries, there is an immediate separation in the survival curves at age 65, with disabled VBID beneficiaries having a higher probability of survival than disabled comparison beneficiaries. In contrast, Figure H.22 shows that among nondisabled beneficiaries just older than 65, survival is similar in both the VBID and comparison groups.

Figure H.21. Adjusted Probability of Survival, Disabled Beneficiaries



NOTE: Kaplan-Meier curve plotting average probability of survival by age.

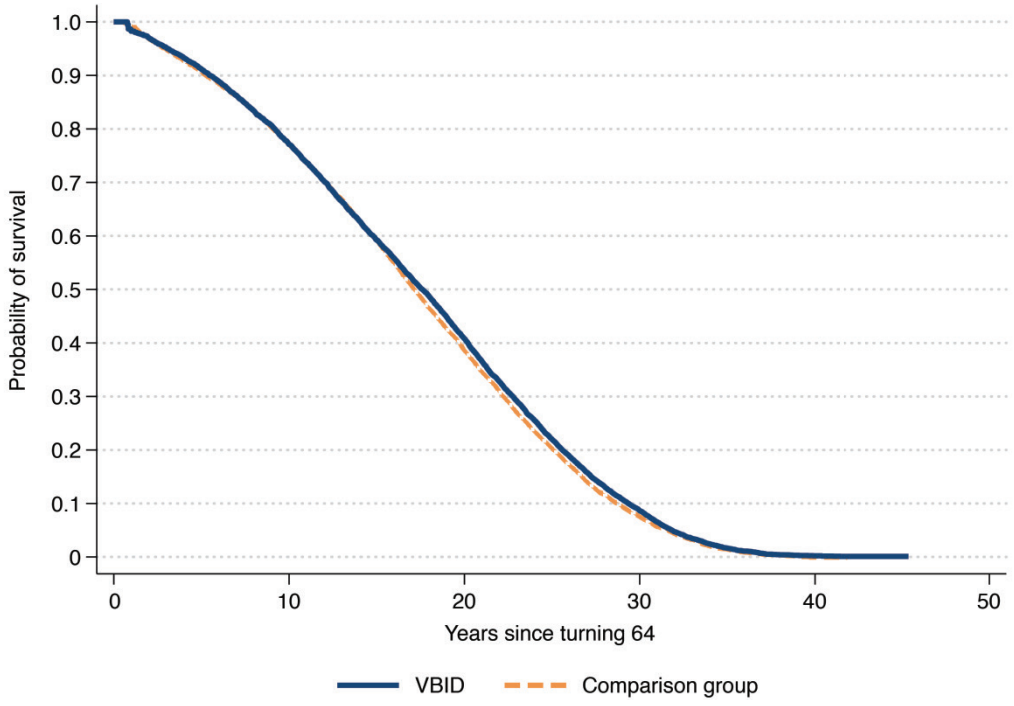
Figure H.22. Adjusted Probability of Survival, Nondisabled Beneficiaries



NOTE: Kaplan-Meier curve plotting average probability of survival by age.

Across *all* beneficiaries, both disabled and nondisabled, the difference in mortality by VBID status is largely concentrated in the first four months of 2017. Figure H.23 again shows the Kaplan-Meier survival curves for all beneficiaries, but conditional on surviving until at least May 1, 2017. The survival probability of VBID and comparison beneficiaries who survived at least until May 1, 2017 is more similar.

Figure H.23. Adjusted Probability of Survival, Beneficiaries Surviving At Least Four Months After VBID Model Test Start



NOTE: Kaplan-Meier curve plotting average probability of survival by age.

To summarize, there was a higher mortality rate in the early months of the VBID model test among comparison beneficiaries than among VBID beneficiaries, and this difference was most pronounced among disabled beneficiaries. It is unclear why this temporal variation in mortality rates occurred.

Limitations

Aside from the failure of the proportional hazard assumption, a limitation of our mortality analysis was the relatively short follow-up period of two years. It is possible that VBID might have delayed effects on mortality that can only be detected with longer follow-up.

Appendix I. Health Care Cost Analytic Results

This appendix describes the methods used for the health care costs analyses presented in Chapter 8. The general approaches for the study design are described in Chapter 1, and the PBP and beneficiary matching processes are described in Appendix D. Table I.1 presents the different measures used in these analyses, their definitions, data source(s), and the VBID model years included in the analyses.

Table I.1. Health Care Cost Measures

Cost Measure	Definition	Data Source(s)	Model Years Included in Analyses
PBP spending			
Medical services	Spending by the PO on medical services on behalf of PBP enrollees	PBP bids	2017
Prescription drugs	Spending by the PO on prescription drugs on behalf of PBP enrollees	PDE data	2017, 2018
Beneficiary costs			
Premiums	Monthly amount paid by beneficiaries enrolled in PBPs for MA and Part D coverage	Public MA and Part D landscape files, IDR premium data	2017–2019
Part D OOP	Annual beneficiary copayments and coinsurance payments for covered Part D drugs	PDE data	2017, 2018
PBP bids			
Medical services	Projected costs of health care coverage for coming year; submitted six months before start of plan year	PBP MA bid data	2017–2019
Prescription drugs	Projected costs of prescription drug coverage for coming year; submitted six months before start of plan year	PBP Part D bid data	2017–2019
Costs to Medicare			
Medical services	Costs paid by Medicare for health care coverage. These include the portion of the MA PBP bids paid by CMS, adjusted by the risk score for each enrollee in the PBP, and MA PBP rebates paid.	PBP MA bids, risk scores	2017, 2018
Prescription drugs	Costs paid by Medicare for prescription drug coverage. These include the portion of the Part D PBP bids paid by CMS, adjusted by the risk score for each enrollee in the PBP, and additional Medicare costs associated with Part D reinsurance and low-income subsidies.	PBP Part D bids, risk scores, publicly available Part D payment data, PDE low-income cost-sharing data	2017, 2018

Data Sources

The outcomes examined in our analyses of health care costs include PBP spending on medical services and prescription drugs, and beneficiary costs, costs to Medicare, and standardized PBP bids. Data on these outcomes were drawn from many different sources, several of which were used in constructing more than one outcome. In this section, we provide a brief overview of these data sources; full details on the definition of each cost measure are presented later in this appendix.

PBP spending on health care services was calculated using data submitted to OACT in support of PBP bids. PBP spending on prescription drugs was derived from the PDE data. Beneficiary premiums and Part D OOP costs were calculated using publicly available and beneficiary-specific MA and Part D premium data from the IDR, as well as PDE data. The OACT extracted data on VBID-participating and matched comparison PBPs from BPT spreadsheets submitted for 2014–2019 MA and Part D bids. In addition to the standardized PBP bids for MA and MA-PD coverage, data from the BPT included numerous variables reflecting costs to Medicare and realized MA spending. The construction of MA costs to Medicare also required data from the CMS IDR on PBP-average final risk scores. Finally, publicly available information about average bids, premium subsidies for low-income-subsidy enrollees, and other variables were downloaded from the CMS website.

Our analysis, which focused on changes within PBPs over time, required a data set with consistently defined PBPs over time. This was complicated by the fact that many PBPs were merged together or relabeled over time; mergers were especially common between 2014 and 2016. To consistently define plans over time, we cross-walked all PBPs in our analysis sample so that every plan in each year was mapped to a PBP in 2017 (the first year of VBID implementation). We accordingly treat the 2017 PBP as the unit of analysis for studying health care costs other than Part D OOP costs; outcomes for all PBPs or segments that cross-walk to each 2017 PBP are aggregated by taking an enrollment-weighted average of PBP bids, PBP premiums, PBP spending, or costs to Medicare observed at the segment or PBP level.

Study Population

The bids, premiums, costs, and spending analyses were conducted at the PBP level, using VBID-participating and matched comparison PBPs. Analyses reported in Chapter 8 were limited to PBPs offering both MA and Part D benefits; supplementary analyses for the full sample of PBPs participating in VBID (including MA-only PBPs) are presented in this appendix. The beneficiary Part D OOP cost analyses were conducted at the beneficiary level for all VBID-eligible and matched comparison beneficiaries, identified as described in Appendix D. We restrict the analysis of beneficiary Part D OOP costs to those enrolled in the same PBP for the entire year, as switching between PBPs during the year can affect a beneficiary's OOP costs.

Outcome Measures

This section describes how we constructed each of the outcome measures used for the health care costs analyses.

Plan Benefit Package Spending

Total PBP spending on MA was calculated using base-period net per-beneficiary per-month data submitted as part of PBP bids. Total PBP spending on Part D (prescription drugs) was calculated using beneficiary-level data accessed via the IDR. Unlike the bids themselves, which are submitted in the year preceding the contract year (year *t*), the base-period net per-beneficiary per-month amounts are reported in the year after the contract year (year $t + 1$) as part of the bids for coverage in the following year (year $t + 2$). For example, 2019 bids, which are submitted in CY 2018, are supported by net per-beneficiary per-month amounts calculated from experience in the 2017 contract year. In contrast, the prescription drug spending data accessible via the IDR are available on a shorter lag (usually the summer after the close of the coverage year). Accordingly, we were able to analyze MA PBP spending for the 2014–2017 contract years and were able to analyze Part D PBP spending for the 2014 through 2018 contract years.

The equation for constructing health care spending is as follows:

$$\text{Per-beneficiary per-month PBP health care services spending} = \sum_s^S \text{net spend}_s, \qquad \text{(I.1)}$$

where Net Spend_s is the net per-beneficiary per-month spending amount for service category *s*, and *S* represents the set of all service categories reported in the BPT.

The equation for constructing prescription drug spending is as follows:

$$\text{Per-beneficiary per-month PBP prescription drug spending} = \frac{CPP+NPP+0.15*GDCA}{N}, \qquad \text{(I.2)}$$

where *CPP* corresponds to the total amount paid by the PBP for the drug, *NPP* is the total amount paid by the PBP for drugs based on enhanced coverage, and gross drug cost above out-of-pocket threshold (*GDCA*) is the total gross drug cost above the catastrophic threshold. We summed all of these values in the PDE data for all PBP enrollees in a given year and then divided by the number of beneficiary months in the plan (*N*) to obtain a per-beneficiary per-month amount.

Beneficiary Costs

We analyzed four different beneficiary cost variables: MA PBP premiums, Part D PBP premiums, total MA-PD premiums, and prescription drug (Part D) OOP costs. We obtained monthly PBP-level premium data for MA and Part D coverage for the years 2014–2019 and constructed annual beneficiary prescription drug costs for the years 2014–2018. We obtained

premium data from the publicly available PBP landscape files and from beneficiary-level IDR MA and Part D premium data for PBPs with missing values in the public files.

We constructed Part D OOP spending from the PDE data housed on the IDR. The sum of the patient payment (PTNT_PMT) variable across all PDEs associated with the given beneficiary's ID is the total amount the beneficiary was responsible for paying for all Part D outpatient prescriptions for the given year.

Costs to Medicare

Costs to Medicare reflect payments made by Medicare to POs for PBPs offering coverage for MA and (for MA-PD PBPs) Part D benefits. MA costs to Medicare are calculated by multiplying the standardized MA bid with the final average risk score across all enrollees, then adding per-beneficiary MA rebates. The equation for deriving the per-beneficiary per-month MA costs to Medicare is as follows:

$$\text{Per-beneficiary per-month MA costs to Medicare} = \text{Standardized MA bid} \times \text{average MA risk score} + \text{MA rebate} \quad \text{(I.3)}$$

Part D costs to Medicare are calculated as the sum of four types of payment: the direct subsidy, payments for the low-income premium, payments for the low-income cost-sharing subsidy (LICS), and payments for reinsurance. The direct subsidy can be calculated at the per-beneficiary per-month level using data on standardized bids, the average Part D risk score, and several other figures that we describe below. For the three other components of Part D costs to Medicare, we measured total costs at the PBP level and then divided by the number of beneficiary months in the PBP to derive per-beneficiary per-month costs. We are able to analyze MA and Part D costs to Medicare for the 2014–2018 contract years. The equation for and Part D costs to Medicare is as follows:

$$\text{Per-beneficiary per-month Part D costs to Medicare} = \text{Direct subsidy} + \frac{LIS\ PS + LICS + 0.8{*}GDCA}{N}, \quad \text{(I.4)}$$

The first term in Equation (I.4) is the direct subsidy. The direct subsidy is equal to the difference between the standardized Part D Bid multiplied by the average risk score and the Part D basic premium:

$$\text{Direct subsidy} = \text{standardized Part D bid} \times \text{average Part D risk score} - \text{Part D basic premium} \quad \text{(I.5)}$$

The Part D basic premium for a PBP is equal to

$$\text{Part D basic premium} = \text{standardized Part D Bid} - \textit{NAMBA} + \text{national average BBP}, \quad \text{(I.6)}$$

where *NAMBA* is the national average monthly bid amount, and national average BBP is the base beneficiary premium. These two quantities are publicly available from the CMS website.

The second term in Equation (I.4) reflects per-beneficiary per-month payments for the low-income premium, LICS, and reinsurance. *LIS PS* is the total amount of low-income subsidy premium subsidies paid to each PBP. The per-beneficiary per-month low-income subsidy payment is based on the difference between the Part D PBP's basic premium and the regional low-income subsidy benchmark premium. We calculated this per-beneficiary per-month amount using data on the Part D premium obligations with full premium assistance from the 2014 to 2018 Medicare Part D plan reports (2019), which are available as part of the landscape and plan and premium information files available from the CMS website. The difference between the Part D basic premium and the Part D premium obligations with full premium assistance reflects LIS premium subsidies after accounting for POs' decisions to apply some or all of the MA rebate to buying down the Part D premium. We then used data on member months of LIS enrollment from IDR to calculate the total value of LIS premium subsidy payments.

LICS is the total amount of low-income cost-sharing subsidies paid to the PBP, which were calculated by summing all cost-sharing subsidies that appear in the PDE data in the IDR. *GDCA* is the gross drug cost above the catastrophic threshold. We summed all GDCA amounts in the PDE for each PBP and then multiplied this PBP level total by 0.8 to obtain Medicare's costs for reinsurance payments. These three PBP-level totals (*LIS PS*, *LICS*, and $0.8 \times GDCA$) were then divided by the total number of beneficiary months in the PBP to obtain a per-beneficiary per-month amount that could be added to the per-beneficiary per-month direct subsidy.

As noted in Chapter 8, we were unable to include risk corridor payments in our Part D cost measure. In addition to the per-beneficiary per-month payments included in our measure of Part D costs, Part D costs to Medicare for a PBP may also include payments (transfers from CMS to PBPs) or recovery of overpayments (transfers from PBPs to Medicare) via the risk corridors, a mechanism intended to limit the gains or losses of each PBP. From the inception of Part D through 2018, aggregate risk corridor payments have overwhelmingly been payments to Medicare. In 2013, for instance, the Medicare Payment Advisory Commission reported that 78 percent of Part D POs made risk corridor payments *to* Medicare in 2013, whereas only 14 percent of POs received risk corridor payments (Medicare Payment Advisory Commission, 2015). PBP-level data on risk corridor payments were not readily available for this study, however, and so we were not able to incorporate risk corridor payments into our cost measure. The predominance of payments from POs to Medicare in earlier years suggests that we may overestimate the final, reconciled cost of Part D coverage to Medicare. However, we have no reason to believe that the omission of risk corridor payments would lead us to differentially underestimate or overestimate costs to Medicare for VBID PBPs during the model test, and so we think it is unlikely that the risk corridor payment data would change the findings of our difference-in-difference models.

Plan Benefit Package Bids

Data on standardized MA (reflecting projected costs for Parts A and B coverage) and Part D bids were provided to RAND by OACT. MA and Part D bids must be submitted by the first Monday in June of the year preceding the contract year. For example, 2019 contract year PBP bids were submitted on or before June 4, 2018. We are thus able to analyze bids for the 2014–2019 contract years, giving us three years of VBID model outcomes to examine. Because of the timing of bid submissions, changes in PBP bids associated with VBID for a given contract year should be interpreted as a reflection of POs' expectations based on the first half of the year preceding the contract year in question. In particular, we note that bids for the first model year (2017) were submitted before the POs had any experience with the VBID model test.

We consider both the MA bid and the Part D bid, separately and summed together, for those MA PBPs also offering Part D. In Chapter 8, we presented analyses focusing on the 33 MA-PD PBPs offering VBID in 2017 and 2018, including one PBP that did not implement VBID until 2018. We also present analyses in this appendix looking at the effect of VBID on MA bids alone for all 46 MA VBID-participating PBPs.

Analysis

In our analysis of beneficiary costs, the unit of analysis was the PBP year for both MA and Part D premiums and the beneficiary-year level for the Part D OOP costs. In our analysis of bids, costs to Medicare, and realized medical spending, the unit of observation was the PBP year. All analyses used the same basic research design involving difference-in-differences regression models with inverse propensity weights, but we used a different statistical model for beneficiary costs to accommodate beneficiary-level data and the statistical properties of data on health care costs. We discuss the regression model used for beneficiary costs before turning to the model used for the other, PBP-level outcomes analyzed in Chapter 8.

Beneficiary Costs

We constructed descriptive statistics showing differences in beneficiary Part D OOP costs. We also ran difference-in-differences models to evaluate the effect of VBID on Part D OOP costs. We modeled the relationship between VBID and OOP costs using a generalized linear model (GLM) with a log link function to account for the skewed distribution of OOP costs. We estimated this model by quasi-maximum likelihood under the assumption that the conditional variance of the error term is proportional to the conditional mean. As discussed by Buntin and Zaslavsky (2004), the model remains consistent under misspecifications of the variance function as long as the exponential conditional mean assumption holds. To express this model formally, let y_{pit} be one of the outcomes for beneficiary i in PBP p in year t, and let $VBID_{pt}$ be

an indicator that the *p*th PBP was implementing VBID in year *t*. Our GLM can be written as follows:

$$y_{pit} = \exp\left(\alpha + \alpha_t + \theta_p + \gamma_t * VBID_{pt} + \beta^T X_{pit}\right) + \varepsilon_{pt}, \quad \text{(I.7)}$$

where

- α = overall intercept
- α_t = year fixed effect (with $\alpha_{2014} = 0$) that captures the trend over time
- θ_p = PBP fixed effect capturing time-invariant differences between PBPs
- γ_t = interaction effect between time and VBID-participating PBPs (with $\gamma_t = 0$ for $t \leq 2016$) that captures the difference-in-differences estimates between participating and comparison PBPs
- β = effect of the additional characteristics included in the model
- ε_{pt} = error term.

Both α_t and θ_p account for unmeasured factors; for example, if participating PBPs in general tend to contract with doctors who are more focused on CM than providers in nonparticipating PBPs, this difference will be captured in θ_p. Time varying factors that differ between beneficiaries in participating PBPs and matched beneficiaries in comparison PBPs can be controlled in vector X_{pit}, as long as these factors are observed. Because the term γ_t is time varying, we will be able to assess whether the effects of VBID become stronger, weaker, or show no change over time.

For the vector X_{pit}, we used the following beneficiary-level controls:

- Age
- Gender
- Race/ethnicity
- Dual eligibility and LIS status (monthly indicator, rolled up to year for annual models – beneficiary is considered dual/LIS if dual/LIS for >6 months of year)
- Disability status
- ESRD status
- HCC/RxHCC (the latter for the prescription drug-focused models).

Bids, Premiums, Costs to Medicare, and Spending

We calculated each measure for all VBID-participating PBPs and for a set of matched comparison PBPs. We then used a difference-in-differences regression to estimate how participation in the VBID model test affected each outcome. In particular, let y_{pt} be one of the outcomes for PBP p in year t, and let $VBID_{pt}$ be an indicator that the pth PBP was implementing VBID in year t to reflect the fact that one PBP began VBID participation in 2018, making these

VBID participation indicators PBP specific.[1] We estimated the following difference-in-differences regression:

$$y_{pt} = \alpha + \alpha_t + \theta_p + \gamma_t * VBID_{pt} + \varepsilon_{pt}, \quad \text{(I.8)}$$

where

- α = overall intercept
- α_t = year fixed effect (with $\alpha_{2014} = 0$) that captures the trend over time
- θ_p = PBP fixed effect capturing time-invariant differences between PBPs
- γ_t = interaction effect between time and VBID-participating PBPs (with $\gamma_t = 0$ for $t \leq 2016$) that captures the difference-in-differences estimates between participating and comparison PBPs.

Models for all outcomes were estimated using weighted least squares (WLS) with inverse propensity score weights constructed as described in Appendix D. We estimated SEs clustered on PBP to allow for arbitrary correlation of the error term within PBPs over time.

Results

Below we present results of our empirical analysis for each of the four groups of outcomes examined in this appendix: PBP spending, beneficiary costs, costs to Medicare, and PBP bids. We begin by evaluating the assumption that trends in outcomes were parallel between VBID and comparison PBPs prior to VBID implementation. We also present descriptive figures illustrating how outcomes for VBID PBPs changed after VBID implementation. We then present regression tables showing the results discussed in the report. Finally, we discuss limitations of our analysis and report sensitivity analyses intended to address those limitations.

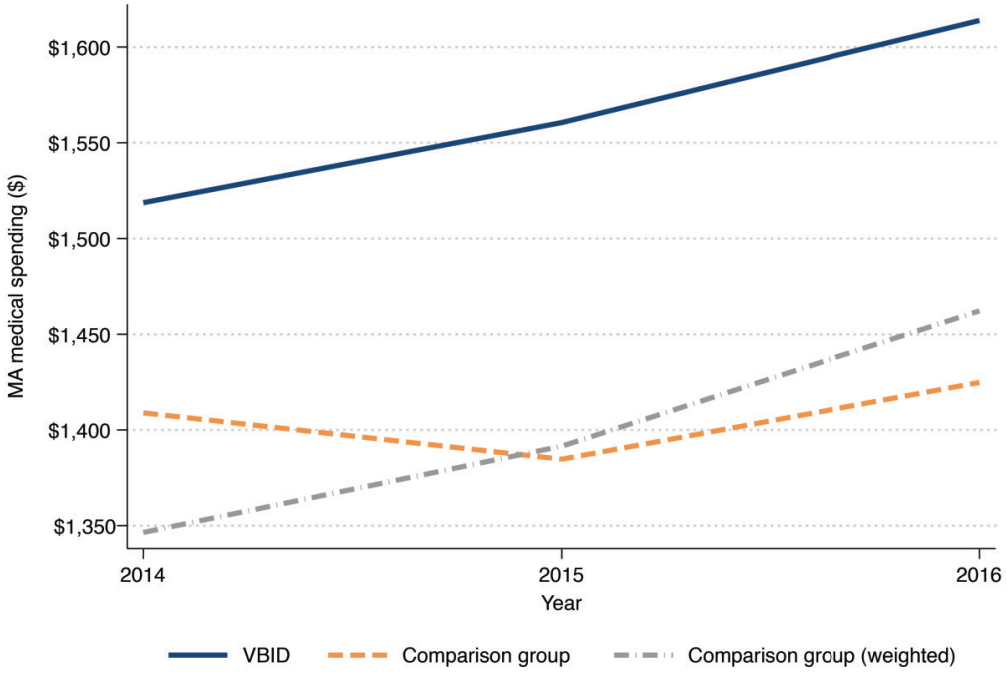
Plan Benefit Package Spending

Parallel Trends

Figures I.1 and I.2 present unadjusted trends in MA health care spending and Part D prescription drug spending by PBPs, for VBID and matched comparison PBPs. The sample used in all figures is restricted to PBPs offering Part D benefits.

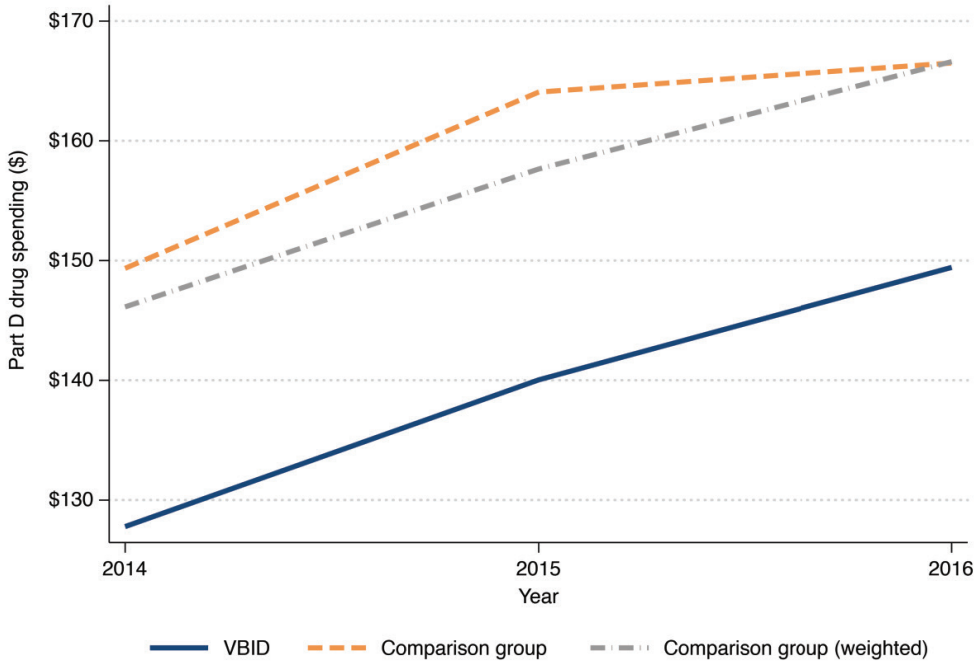
[1] Regression models for PBP spending, bids, and costs to Medicare also included PBP-specific indicators for the VBID model year γ_t separately from the interaction term $\gamma_t * VBID_{pt}$. The presence in our sample of a PO that did not implement VBID until 2018 prevents these model year fixed effects from being perfectly collinear with the time fixed effects. However, inclusion of the separate VBID model year γ_t term has no meaningful impact on the estimated VBID effect. These coefficients are labeled VBID Year 1, VBID Year 2, and VBID Year 3 in regressions where they were included.

Figure I.1. MA Medical Spending by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.01 before weighting and 0.87 after weighting.

Figure I.2. Part D Drug Spending by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.42 before weighting and 0.93 after weighting.

Table I.2. Test Statistics and P-Values for Parallel Trends Assumptions Regarding PBP Spending, Unweighted and with Inverse Propensity Weights

Outcome	Unweighted F-Statistic	p-Value	Weighted F-Statistic	p-Value
Per-beneficiary per-month realized MA spending	4.93**	0.01	0.14	0.87
Per-beneficiary per-month realized Part D spending	0.89	0.42	0.07	0.93

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. F-statistic for test of hypothesis that pre-VBID trends are parallel between VBID and comparison PBPs between 2014 and 2016. Significance assessed using $F_{2,64}$ distribution.

Table I.2 reports test statistics and p-values for the hypothesis that trends were parallel between VBID and comparison PBPs prior to VBID. We rejected (at the 5-percent significance level) the hypothesis that VBID and matched comparison PBPs had parallel trends in MA spending between 2014 and 2016. As discussed in Chapter 1, we constructed inverse propensity weights for each outcome. The inverse propensity weights were estimated using the change in outcomes between 2014–2015 and 2015–2016 to predict VBID status in a logistic regression model. These weights serve to improve the balance between VBID and matched comparison plans in terms of changes in outcomes leading up to VBID implementation, as suggested by the greater similarity in trends between the weighted comparison group and the VBID PBPs, and by the small and insignificant F-statistics reported in Table I.2. All difference-in-differences estimates for PBP spending are estimated using WLS regression with inverse propensity weights.

Descriptive Results

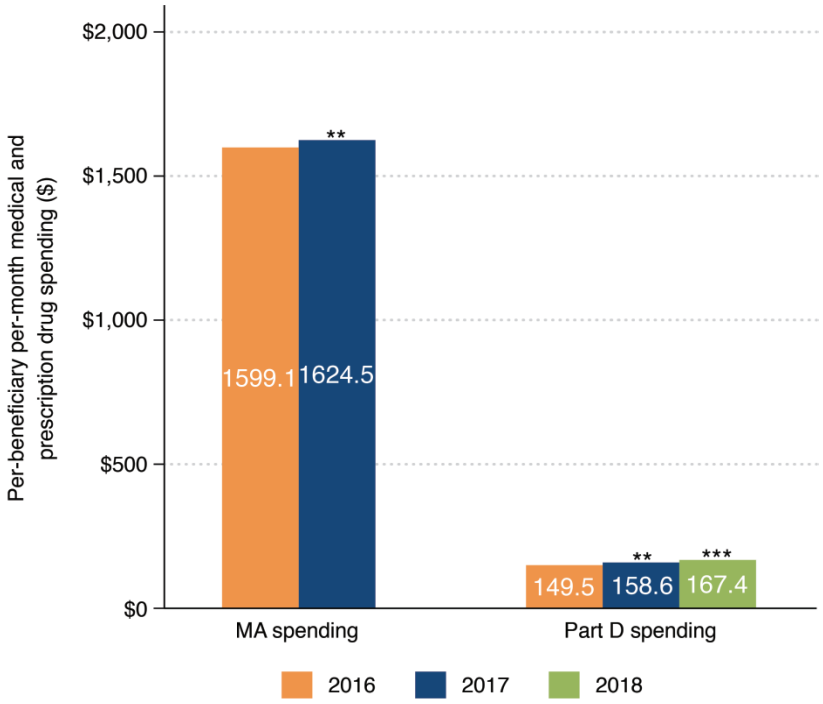
Figure I.3 shows average MA spending from 2016 through 2017 and average Part D spending from 2016 through 2018 for VBID-participating PBPs. Relative to 2016, spending on both MA and Part D rose in the VBID plans after the model test began.

Table I.3 presents summary statistics for the outcome variables analyzed in this chapter. Differences in the number of observations available for each outcome reflect differences in the range of years for which data were available.

Regression Results

Table I.4 shows the difference-in-differences model results for PBP spending. These results are discussed in Chapter 8.

Figure I.3. Descriptive Statistics for Per-Beneficiary Per-Month Spending



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing VBID PBP spending in each model year with spending in 2016. Sample size (number of unique PBPs in each year) ranges between 32 and 33.

Table I.3. Summary Statistics for Health Care Services and Prescription Drug Spending by PBPs, MA-PD Plans

Variable	VBID Average	VBID SD	VBID Min.	VBID Median	VBID Max.	VBID N	Comp. Average	Comp. SD	Comp. Min.	Comp. Median	Comp. Max.	Comp N.
MA medical spending by PBPs												
2014	1,501	390	696	1,490	2,300	30	1,346	359	737	1,293	2,249	32
2015	1,540	381	968	1,506	2,338	31	1,392	349	838	1,389	2,288	32
2016	1,599	415	871	1,584	2,446	32	1,462	349	952	1,436	2,381	32
2017	1,624	405	871	1,604	2,478	32	1,525	400	826	1,530	2,478	32
Part D drug spending by PBPs												
2014	129	39	62	125	258	31	146	37	51	145	277	32
2015	142	45	69	141	312	32	158	40	88	155	376	32
2016	149	51	75	140	313	33	167	45	103	164	307	32
2017	159	63	74	157	377	33	185	58	95	181	346	32
2018	167	60	79	160	319	32	198	59	88	194	380	32

Table I.4. Difference-in-Differences Model Results for PBP Spending Outcomes

	MA Spending	SE	Part D Medical Spending	SE
Number of observations	257		322	
Intercept	1,538.19***	9.63	N/A	
VBID indicator	(omitted)		(omitted)	
VBID Year 1	33.28	29.15	−26.29***	5.79
VBID Year 2	N/A		−114.48***	8.06
Year (2016 reference)				
2014	−112.98***	18.46	−21.56***	3.89
2015	−63.89***	14.98	−9.78***	2.25
2017	33.04***	10.35	43.64***	1.94
2018	N/A		144.75***	5.45
VBID indicator*VBID Year 1	−41.66	34.48	−10.61	7.23
VBID indicator*VBID Year 2	N/A		−7.87	9.11

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. Models include PBP fixed effects. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Sensitivity Analysis: Medicare Advantage Spending Regression Results for All Value-Based Insurance Design Plan Benefit Packages

As a sensitivity analysis, we used the difference-in-differences regression model presented above to estimate the change in PBP spending on health care services for the full sample of all VBID PBPs, including both MA-PD PBPs and MA-only PBPs. As in our analysis of PBP spending for MA-PD PBPs, we find that changes in PBP spending on health care services associated with VBID are generally small and statistically insignificant.

Sensitivity Analysis: Medicare Advantage Spending Results for Plan Benefit Packages with Per-Beneficiary Per-Month Spending Directly Reported

As discussed in Chapter 8, not all PBPs had per-beneficiary per-month MA spending amounts reported in the BPT spreadsheets. Because per-beneficiary per-month spending is reported retrospectively, data on spending amounts for PBPs that merged may reflect an average for multiple base period PBPs. For these PBPs, we examined information in the BPT spreadsheet identifying the base period PBPs that contributed experience to each PBP's bids and assigned each PBP the average per-beneficiary per-month spending amount for all PBPs that referenced each base period PBP as a source. For instance, if a PBP offered in 2016 merged with other PBPs between 2016 and 2018, we would identify all 2018 PBPs that referenced that PBP as a source of experience in bids for the 2018 contract year and assign the 2016 PBP the average of base period per-beneficiary per-month spending for all such 2018 PBPs.

We were concerned that this imputation approach might introduce measurement error into our realized MA spending outcome variable, so we estimated our difference-in-difference

Table I.5. Difference-in-Differences Model Results Using Pooled MA-PD and MA-Only PBPs for MA Spending

	MA Spending	SE
Number of observations	361	
Intercept	1,556***	15.93
VBID indicator	(omitted)	
VBID Year 1	27.32	48.33
VBID Year 2		
VBID Year 3		
Year (2016 reference)		
2014	−150.4***	39.23
2015	−70.67***	25.27
2017	18.31	20.39
2018		
2019		
VBID Indicator*VBID Year 1	−34.32	51.08
VBID Indicator*VBID Year 2		
VBID Indicator*VBID Year 3		

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. Models include PBP fixed effects. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

regression model on a sample restricted to PBPs for which per-beneficiary per-month spending amounts could be assigned without any ambiguity. Fortunately, 84 percent of PBP-year observations for MA-PD plans could be assigned realized MA spending amounts without ambiguity. (Among MA-only PBPs, 97 percent of PBP-year observations could be assigned realized per-beneficiary per-month amounts without ambiguity.)

Table I.6 shows regression results for this restricted sample of MA-PD plans.

Table I.6. Difference-in-Differences Model Results for PBPs with Directly Reported Per-Beneficiary Per-Month Spending

	MA Spending	SE
Number of observations	216	
Intercept	1,557.02***	10.81
VBID indicator	(omitted)	
VBID Year 1	27.85	33.39
Year (2016 reference)		
2014	−110.23***	19.53
2015	–57.43***	19.53
2017	36.11***	10.81
VBID indicator*VBID Year 1	−44.66	39.04

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

As in the full sample, VBID-associated changes in 2017 per-beneficiary per-month MA spending are small and insignificant.

Sensitivity Analysis: Medical and Prescription Drug Spending Estimates Adjusted for Risk Scores

Our discussion of the medical and prescription drug spending estimates in Chapter 8 referenced a set of estimates that included controls for PBP-average risk scores to disentangle the effects of VBID on spending from potential compositional changes that could affect spending. Although our main estimates omit controls for risk scores to avoid overcontrolling for changes in risk scores that might reflect improved health due to VBID, evidence from the beneficiary-level health outcomes analysis (in Chapter 7) indicated that VBID PBPs' risk scores were trending downward prior to VBID implementation. The following table reports difference-in-differences regression estimates with and without controls for PBP-average risk scores.

Table I.7. Model Results for Medical Spending with and Without Risk Scores Included

	MA Spending	SE	MA Spending	SE	Part D Spending	SE	Part D Spending	SE
Number of observations	257		257		322		322	
Intercept	1,538.19***	9.63	39.96	376.6	N/A		−232.2*	125.7
VBID indicator	(omitted)		(omitted)		(omitted)		(omitted)	
VBID Year 1	33.28	29.15	94.04**	44.58	−26.29***	5.79	−19.50**	9.500
VBID Year 2	N/A				−114.48***	8.06	−86.05***	16.94
Year (2016 reference)								
2014	−112.98***	18.46	−60.57***	16.34	−21.56***	3.89	−18.21***	4.143
2015	−63.89***	14.98	−51.67***	13.89	−9.78***	2.25	−16.66***	6.260
2017	33.04***	10.35	−59.33**	29.01	43.64***	1.94	39.82***	3.825
2018	N/A				144.75***	5.45	129.7***	8.509
VBID indicator*VBID Year 1	−41.66	34.48	−20.70	31.32	−10.61	7.23	−1.745	6.782
VBID indicator*VBID Year 2	N/A				−7.87	9.11	2.202	8.306
Part C risk score			1,356***	342.9				
Part D risk score							378.7***	132.8

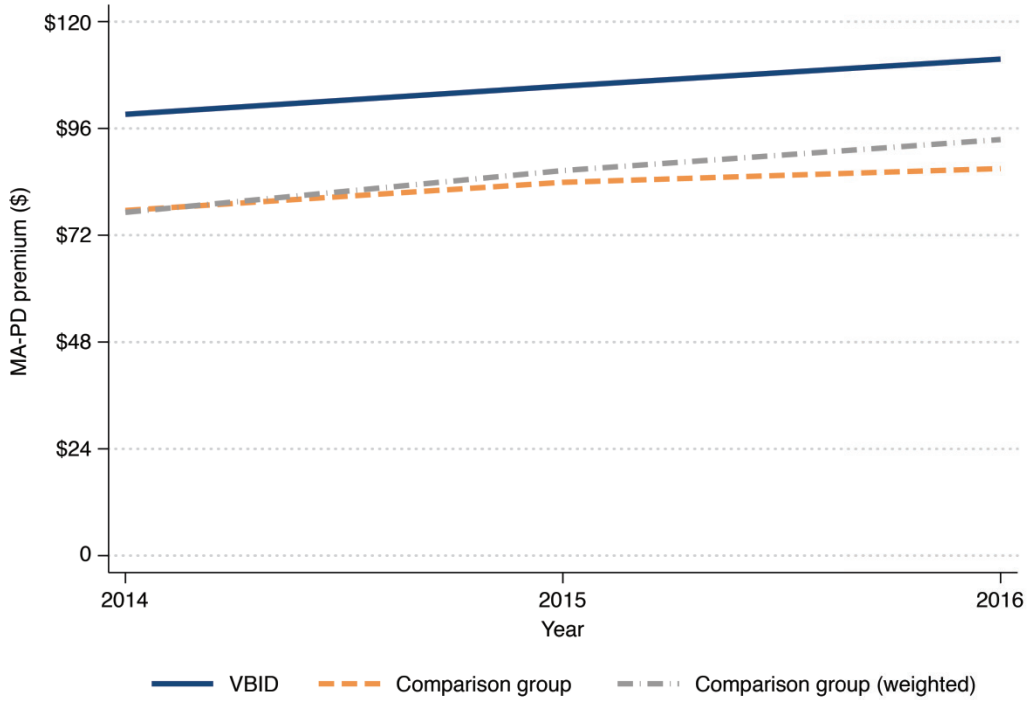
NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Beneficiary Costs

Parallel Trends

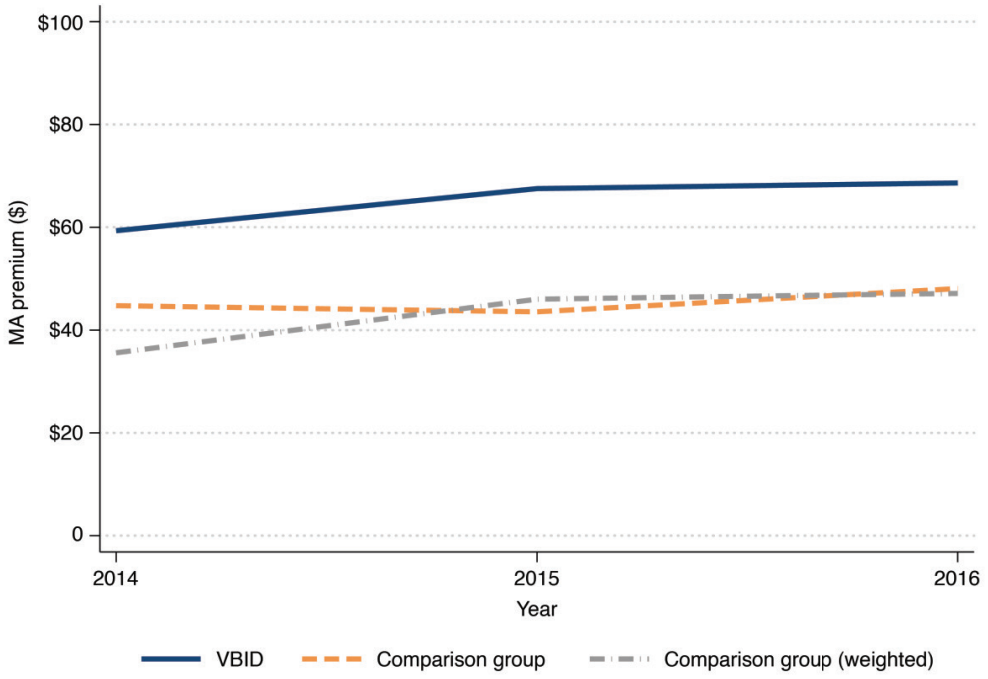
Figures I.4–I.6 present unadjusted trends in combined MA and Part D premiums, MA PBP premiums, and Part D PBP premiums for PBPs participating in VBID and the matched comparison PBPs. Figure I.7 presents unadjusted trends in Part D OOP costs for beneficiaries in VBID and matched comparison PBPs. The sample used in all figures is restricted to PBPs offering Part D benefits. Part D OOP costs were compared only for VBID beneficiaries (Figure I.7) enrolled in PBPs offering Part D benefits and their matched comparisons. In addition to the three beneficiary cost measures presented in Chapter 8, these figures provide additional detail on beneficiary premiums and the Part D OOP costs.

Figure I.4. Total MA and Part D Premiums by VBID Status, 2014–2016



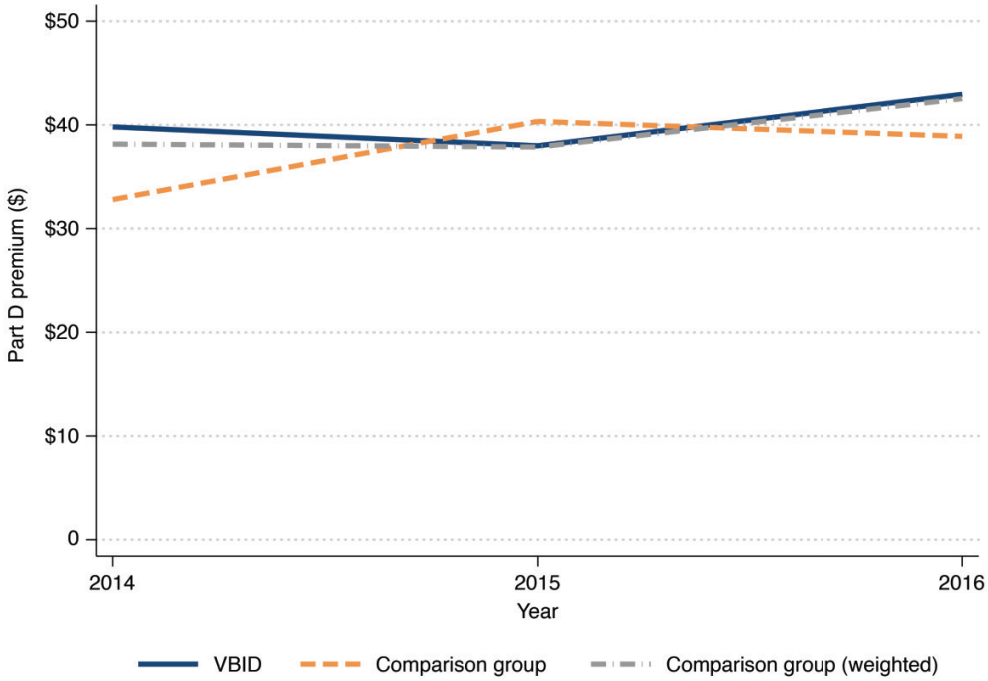
NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.44 before weighting and 0.996 after weighting.

Figure I.5. MA Premiums by VBID Status, 2014–2016



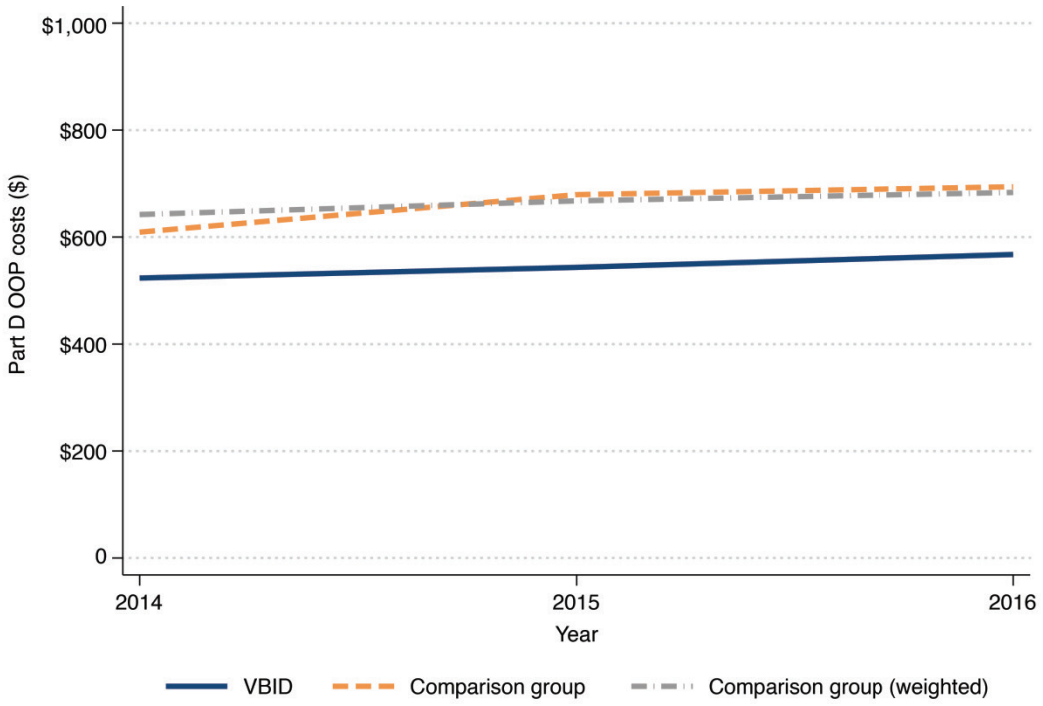
NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.09 before weighting and 0.97 after weighting.

Figure I.6. Part D Premiums by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.04 before weighting and 0.98 after weighting.

Figure I.7. Part D OOP Costs by VBID Status, 2014–2016



NOTE: Sample restricted to beneficiaries enrolled in MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was < 0.001 before weighting and 0.003 after weighting.

Table I.8 reports test statistics and p-values for the hypothesis that trends were parallel between VBID and comparison PBPs prior to VBID. For each outcome variable, we first tested for parallel trends in the unweighted data using the approach described in Appendix D, but with models consistent with Equation (I.5). We then generated weights using inverse probability weights (for PBP-level premium analyses) and entropy balancing (for the beneficiary-level Part D OOP costs analysis) and repeated the test on the weighted data to verify that our weights

Table I.8. Test Statistics and p-Values for Parallel Trends Assumptions, Unweighted and with Weights

Outcome	Unweighted F Statistic/Unweighted Chi-Squared	p-Value	Weighted F Statistic/Weighted Chi-Squared	p-Value
MA PBP premiums*	2.55	0.086	0.03	0.972
Part D PBP premiums*	3.38	0.040	0.02	0.985
MA-PD PBP premiums (total)*	0.83	0.441	0.00	0.996
Part D OOP costs**	32.30	<0.001	11.56	0.003

NOTE: *F for test of hypothesis that pre-VBID trends are parallel between beneficiaries enrolled in VBID-participating PBPs and matched comparison beneficiaries between 2014 and 2016. Significance assessed using F distribution. **Chi-squared for test of hypothesis that pre-VBID trends are parallel between beneficiaries enrolled in VBID-participating PBPs and matched comparison beneficiaries between 2014 and 2016. Significance assessed using a chi-squared distribution with two degrees of freedom.

improved the balance for pre-VBID trends in outcomes. Table I.8 reports the F and chi-squared statistics. Our null hypothesis is that the trends are parallel for both groups. Null results ($p > 0.05$) for an outcome mean that we found no evidence that the parallel trends assumption fails for that outcome; significant results for an outcome ($p < 0.05$) indicate that the trends are not parallel for that outcome

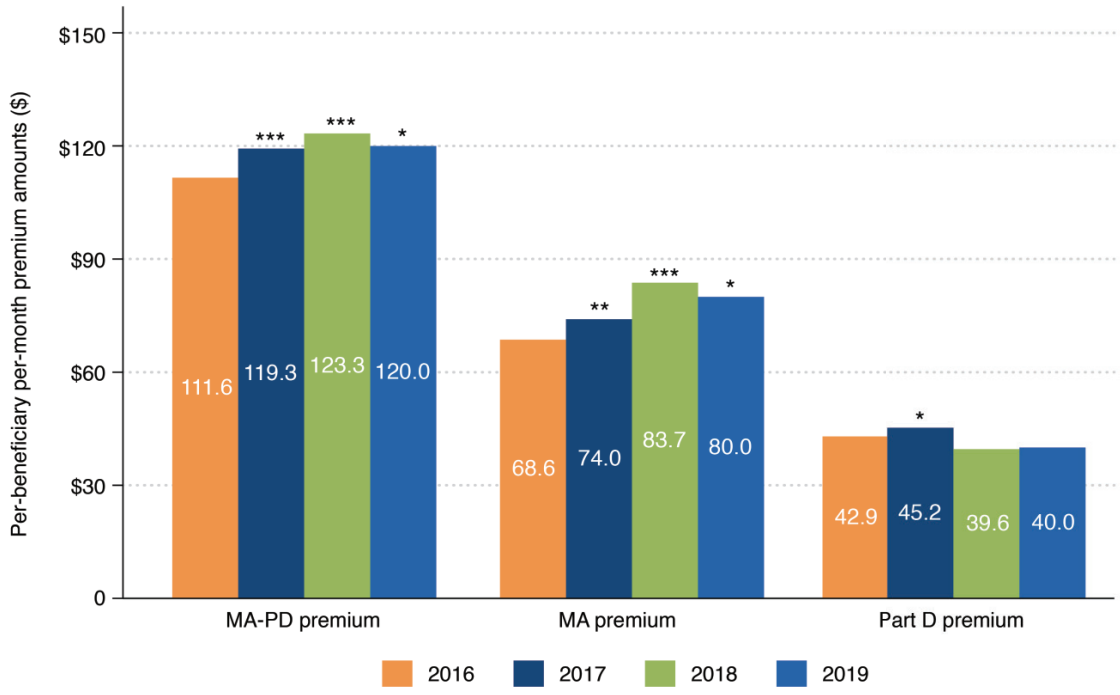
As suggested by the figures and table above, two of the three PBP premium outcomes evolved in parallel for the VBID-participating and comparison PBPs between 2014 and 2016. Trends in Part D premiums were not parallel prior to weighting, however. Part D OOP costs did not evolve in parallel, either. The results in Table I.7 suggest that, for these two outcomes, reweighting data from the matched comparison group might help improve comparability between the VBID and comparison groups.

Table I.8 shows that, although the weights serve to somewhat improve the balance of outcomes in the preperiod, we still find statistically significant evidence ($p=0.003$) of nonparallel trends in Part D OOP costs between beneficiaries in VBID and comparison plans. The trends for the reweighted data do appear to be more parallel, however, as shown in Figure I.7. All difference-in-differences estimates reported below are therefore estimated using weights.

Descriptive Results

Figure I.8 presents descriptive statistics for the three premium outcomes reported in Chapter 8 for VBID-participating PBPs offering Part D (MA-PDs). Both MA premiums and the combined

Figure I.8. Descriptive Statistics for PBP Premiums



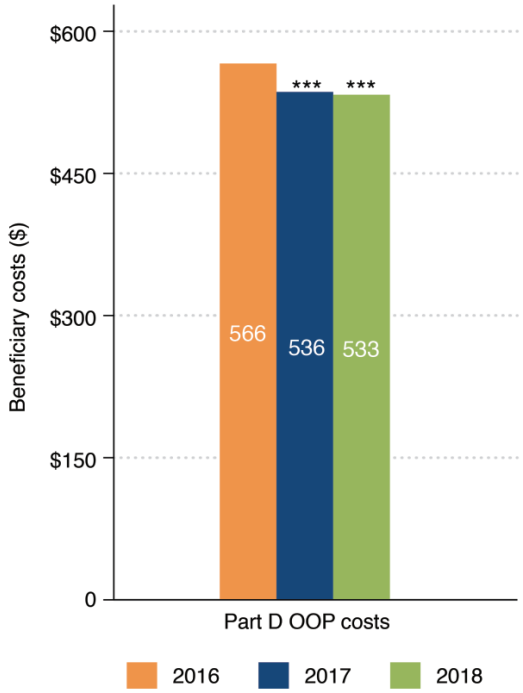
NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing 2016 and 2017 outcomes. Sample size (number of PBPs in each year) ranges between 31 and 33.

MA-PD premiums (the sum of MA and Part D premiums) increased significantly in each year of the model, relative to 2016. Part D premiums increased significantly in 2017 relative to 2016, but by a small magnitude ($2.30).

Figure I.9 presents descriptive statistics for the Part D OOP costs outcome, for beneficiaries enrolled in VBID-participating PBPs offering Part D (MA-PDs) and eligible for VBID. Average Part D OOP costs declined from 2016 to 2017 and 2018; these results were statistically significant, though relatively small in magnitude ($21 and $15 decline).

Table I.9 reports summary statistics for beneficiary costs.

Figure I.9. Descriptive Statistics for Part D OOP Costs, 2017–2018



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing 2016 and 2017, and 2016 and 2018 outcomes. Sample size (number of beneficiaries in each year) ranges between 42,220 and 66,384.

Table I.9. Summary Statistics for Beneficiary Costs (in Dollars), MA-PD Plans, 2014–2019

Variable	VBID Avg.	VBID SD	VBID Min.	VBID Median	VBID Max.	VBID N	Comp. Avg.	Comp. SD	Comp. Min.	Comp. Median	Comp. Max.	Comp. N
MA premiums												
2014	59	55	0	61	159	31	45	56	0	28	266	32
2015	68	60	0	50	183	32	44	48	0	27	195	32
2016	69	63	0	48	184	33	48	47	0	31	183	32
2017	75	72	0	44	222	33	44	49	0	34	189	32
2018	83	75	0	66	250	33	43	52	0	20	213	32
2019	76	75	0	55	237	32	42	52	0	21	219	32
Part D premiums												
2014	40	27	0	36	117	31	33	24	0	36	72	32
2015	38	21	0	37	80	32	40	25	0	39	89	32
2016	43	24	0	44	88	33	39	27	0	36	87	32
2017	45	24	0	47	79	33	42	27	0	40	82	32
2018	39	22	0	35	73	33	44	27	0	39	88	32
2019	39	22	0	40	78	32	38	31	0	34	103	32
MA-PD premiums												
2014	99	74	0	96	252	31	78	68	0	65	328	32
2015	105	76	0	91	263	32	84	60	0	80	253	32
2016	112	79	0	92	272	33	87	64	0	83	249	32
2017	119	86	0	94	295	33	86	69	0	80	270	32
2018	123	87	0	98	292	33	86	72	0	78	296	32
2019	116	89	0	98	292	32	81	72	0	75	295	32
Part D OOP costs												
2014	521	691	0	295	13,061	37,078	605	783	0	343	17,862	39,580
2015	541	781	0	282	22,581	51,178	675	859	0	394	16,802	52,574
2016	566	841	0	281	22,843	66,384	692	910	0	386	14,972	66,587
2017	536	831	0	248	23,945	58,463	672	910	0	373	25,559	59,777
2018	533	850	0	249	38,164	42,220	659	901	0	352	16,362	44,542

NOTE: Dollar amounts rounded to the nearest dollar. Premiums are reported in monthly amounts, whereas Part D OOP costs are in annual amounts.

Regression Results

Tables I.10 and I.11 show the difference-in-differences model results for beneficiary costs. As shown in Chapter 8, we find that VBID was associated with increases in MA premiums in all three years of the model test, and was associated with declines in Part D premiums in 2018, as well as Part D OOP costs in 2017 and 2018, relative to the matched comparison groups. These results are discussed in Chapter 8.

Table I.10. Difference-in-Differences Model Results for Part D OOP Costs (in Dollars)

	Part D OOP costs	SE
Number of observations	127,877	N/A
Intercept	6.69***	0.04
Year (2014 reference)		
2015	0.05***	0.0
2016	0.11***	0.01
2017	0.12***	0.01
2018	0.14***	0.01
Female	0.06***	0.01
Age	−0.01***	0.00
LIS	−0.94***	0.02
Dual	−0.38***	0.02
Disabled	0.14***	0.01
Race (white, reference)		
Black	−0.11***	0.03
Hispanic	−0.24***	0.05
Asian/Pacific Islander	−0.15***	0.05
American Indian/Alaska Native	1.12**	0.49
Multiple	−2.46***	0.67
ESRD	0.11***	0.02
Risk score	0.11***	0.00
VBID indicator*Year 1	−0.04***	0.01
VBID indicator*Year 2	−0.03***	0.01

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results are presented. Models include PO fixed effects. Coefficient estimates are from a GLM Poisson model and are shown with robust SEs in the next column. The intercept is defined as the value that estimates the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Table I.11. Difference-in-Differences Model Results for PBP Premiums (in Dollars), MA-PD PBPs

	MA Premiums	SE	Part D Premiums	SE	MA-PD Premiums	SE
Number of observations	65	N/A	65	N/A	65	N/A
Intercept	45.97***	2.18	38.68***	2.05	86.46***	3.33
Year (2014 reference)						
2015	10.52***	2.02	−0.64	2.21	9.49***	2.12
2016	12.01***	2.47	4.09	2.56	16.31***	3.11
2017	6.75***	2.51	6.97**	2.83	17.95***	2.91
2018	5.41*	2.76	9.39***	2.53	18.74***	7.00
2019	5.26	3.26	5.49	4.51	12.60	8.05
VBID indicator*Year 1	11.36**	4.51	−1.34	3.19	5.78	6.65
VBID indicator*Year 2	21.43***	5.94	−8.90**	3.82	8.65	8.00
VBID indicator*Year 3	18.34**	7.38	−4.20	5.19	12.26	9.70

NOTE: ***, **, and * represent statistical significance at 1-, 5-, and 10-percent, respectively. Difference-in-differences results are presented. Models include PBP fixed effects. Coefficient estimates are shown with robust SEs in the next column. The intercept is defined as the value that estimates the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Sensitivity Analysis: Medicare Advantage Premiums for All Value-Based Insurance Design Plan Benefit Packages

As a sensitivity analysis, we used the difference-in-differences regression model presented above to estimate the change in MA PBP premiums for the full sample of all VBID PBPs, including both MA-PD PBPs and MA-only PBPs. We continue to find an increase in MA premiums associated with VBID in all three years. However, the increases are smaller than those estimated in the MA-PD sample and not statistically significant in 2017.

Table I.12. Difference-in-Differences Model Results for PBP Premiums, Across All MA PBPs (in Dollars)

	MA Premiums (All PBPs)	SE
Number of observations	91	N/A
Intercept	52.25***	2.70
Year (2014 reference)		
2015	11.76***	2.67
2016	14.53***	3.01
2017	13.02***	4.16
2018	10.71**	4.68
2019	5.06	6.64
VBID indicator*Year 1	7.09	4.45
VBID indicator*Year 2	17.87***	6.09
VBID indicator*Year 3	20.98**	8.49

NOTE: ***, **, and * represent statistical significance at 1-, 5-, and 10-percent, respectively. Difference-in-differences results are presented. Models include PBP fixed effects. Coefficient estimates are shown with robust SEs in the next column. The intercept is defined as the value that estimates the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Sensitivity Analysis: MA Premium Results Excluding Outlier PBPs

Linear regression estimates can be sensitive to outliers, so we evaluated the sensitivity to outliers of our difference-in-differences estimates for the effect of VBID on MA premiums by dropping data from PBPs with unusually large year-on-year increases in MA premiums and re-running our difference-in-differences regression models. While PBP fixed effects control for long-run differences across PBPs in the level of premiums, large year-on-year changes in certain PBPs could drive our results even if premium changes among most PBPs did not differ between VBID and comparison PBPs. To determine criteria for identifying outliers, we calculated year-on-year changes (i.e., first differences) in MA premiums for all PBPs in each year from 2015 through 2019 and examined the distribution of the year-on-year change in MA premiums. We then set outlier cutoffs at $50, $40, and $30: in the sensitivity analyses, we dropped data from all years for any PBP that had a one-year increase in premiums above each of these thresholds. Regression estimates and information about the number of VBID and comparison PBPs included in the sample are presented in Table I.13.

Table I.13. Difference-in-Differences Model Results for PBP Premiums Excluding Outliers

	MA Premiums (main results)		Drop PBPs with Increase in MA Premium Larger Than $50		Drop PBPs with Increase in MA Premium Larger Than $40		Drop PBPs with Increase in MA Premium Larger Than $30	
Number of Observations	387		369		351		309	
N VBID PBPs	33		30		27		25	
N Comparison PBPs	32		32		32		27	
Total N PBPs	65		62		59		52	
Intercept	45.97***	(2.18)	44.21***	(1.92)	42.59***	(1.53)	43.51***	(1.49)
Year (2014 reference)								
2015	10.52***	(2.03)	11.04***	(2.09)	8.30***	(1.86)	7.10***	(2.25)
2016	12.01***	(2.48)	11.33***	(2.42)	9.27***	(2.25)	5.16***	(1.77)
2017	6.75***	(2.51)	6.13**	(2.52)	4.57*	(2.45)	1.04	(2.52)
2018	5.41*	(2.76)	4.82*	(2.65)	3.49	(2.57)	0.65	(2.24)
2019	5.26	(3.26)	4.86	(3.20)	3.50	(3.17)	–1.98	(2.19)
VBID Indicator*Year 1	11.36**	(4.51)	9.17**	(3.61)	6.39**	(3.17)	6.54**	(3.17)
VBID Indicator*Year 2	21.43***	(5.94)	15.70***	(4.14)	11.89***	(3.66)	14.03***	(3.75)
VBID Indicator*Year 3	18.34**	(7.38)	12.17*	(6.43)	8.11	(6.41)	17.00***	(4.76)

NOTES: SE = standard error. ***, ** and * represent statistical significance at 1 percent, 5 percent, and 10 percent, respectively. Difference-in-differences results are presented. Models include PBP fixed effects, year fixed effects, and an intercept (not reported). Coefficient estimates are shown with cluster-robust standard errors in the next column. N VBID (Comparison) PBPs is the number of VBID (Comparison) PBPs retained after dropping outliers according to the criterion in the column header. Total N PBPs = N VBID PBPs + N Comparison PBPs.

Trimming outliers reduces the magnitude of the estimated VBID effects, but the significance and sign of post-VBID changes in MA premiums are robust to the exclusion of PBPs with large

increases in premiums. Further, using $30 as the threshold drops 20 percent of PBPs in the sample, suggesting that this threshold begins to cut into the heart of the distribution, leading to exclusion of PBPs that did not have unusually large changes in MA premiums. We conclude that the MA premium results are not driven by just a few outlier PBPs.

Sensitivity Analysis: MA Premium Results Reweighted by PBP Enrollment

Another potential concern with our MA premium findings is that our estimates weight all VBID PBPs equally regardless of their enrollment. This opens the door to the possibility that large premium changes among very small PBPs could drive our results.

Table I.14. Difference-in-Differences Model Results for PBP Premiums, Weighted by PBP Enrollment

	MA Premiums, Original Results		MA Premiums, Enrollment-Weighted	
	coef.	SE	coef.	SE
Number of Observations	387		387	
Intercept	45.97***	2.18	28.28***	2.82
Year (2014 – reference)				
2015	10.52***	2.03	12.96***	3.93
2016	12.01***	2.48	12.52***	3.62
2017	6.75***	2.51	7.99***	2.70
2018	5.41*	2.76	6.67**	2.50
2019	5.26	3.26	3.10	6.08
VBID Indicator*2017	11.36**	4.51	7.54*	4.14
VBID Indicator*2018	21.43***	5.94	14.21**	6.20
VBID Indicator*2019	18.34**	7.38	17.25**	7.82

NOTES: SE = standard error. ***, ** and * represent statistical significance at 1 percent, 5 percent, and 10 percent respectively. Difference-in-differences results are presented. Models include PBP fixed effects, year fixed effects, and an intercept (not reported). Coefficient estimates are shown with cluster-robust standard errors in the next column. All results are weighted to ensure parallel trends in the pre-VBID period. The "enrollment weighted" results are further weighted to reflect PBP-level enrollment.

Table I.14 compares regression estimates for MA premiums between our main analysis and an analysis that incorporates PBP-level enrollment into the weighting scheme. The inverse propensity weights are re-estimated using plan enrollment weights. The difference-in-difference models apply a weight equal to the inverse propensity weights multiplied by plan enrollment. While the coefficients are slightly smaller, we continue to find a statistically significant relationship between VBID implementation and MA premiums in all years.

Sensitivity Analysis: Exploration of PBP Costs and Administrative Expenses Contributing to MA Premiums

To understand mechanisms that may have contributed to the increases in MA premiums that we observed among VBID plans, we analyzed data from OACT's bid pricing tool (BPT) to assess whether medical spending on additional services, including mandatory supplemental

benefits, increased for VBID-participating PBPs relative to matched comparators in post-implementation years (2017, 2018, and 2019). We also examined non-benefit expenses (which include administrative costs and some disease management activities not involving health care providers or DME) to test whether VBID plans exhibited faster growth in administrative expenses. As with our data on PBP bids, these measures are submitted before the year for which the bid applies. These data are not retrospective measures of realized costs, but rather prospective measures of the supplemental benefit costs and expenses that PBPs anticipate for the following year. Estimates are presented in Table I.15.

We found that VBID was associated with statistically significant increases in each of these measures. We found a marginally significant $7.35 increase in projected spending on supplemental benefits in 2018 ($p < 0.10$), but not in 2017 or 2019. Relative to the comparison group, VBID PBPs experienced statistically significant increases in administrative costs for MA-covered services in all three years, with increases ranging from nearly $10 dollars in 2017 to nearly $19 in 2019. Administrative costs for mandatory supplement benefits also increased among VBID PBPs relative to comparators in 2017 and 2018, although these effects were relatively small in magnitude.

Table I.15. Difference-in-Differences Model Results for PBPM Cost of Additional Services and Non-Benefit Expenses

	Per-Beneficiary Per-Month Cost of Additional Services		Non-Benefit Expenses for MA-Covered Services		Non-Benefit Expenses for Mandatory Supplemental Services	
	coef.	(se)	coef.	(se)	coef.	(se)
Number of Observations	387		387		387	
Intercept	13.95***	(0.77)	86.37***	(2.19)	8.15***	(0.19)
VBID Year 1	0.55	(1.48)	–7.29***	(1.41)	0.091	(0.36)
VBID Year 2	–4.55**	(2.15)	–4.91	(4.03)	–0.16	(0.52)
VBID Year 3	5.77	(4.82)	–7.62	(6.73)	2.38**	(1.11)
Year (2016– reference)						
2014	0.96	(1.13)	–5.45*	(2.77)	0.091	(0.36)
2015	0.86*	(0.49)	–4.73*	(2.64)	-0.16	(0.53)
2017	-0.29	(0.45)	–8.57***	(1.72)	2.38**	(1.11)
2018	5.46***	(1.38)	5.16**	(2.50)	–1.30***	(0.30)
2019	3.22	(4.15)	–0.13	(4.93)	–0.72***	(0.26)
VBID Indicator*Year 1	2.71	(1.89)	9.62***	(2.12)	0.77*	(0.44)
VBID Indicator*Year 2	7.35*	(4.16)	12.24**	(4.96)	2.06**	(0.97)
VBID Indicator*Year 3	1.98	(4.12)	18.56***	(5.93)	1.20	(0.81)

NOTES: SE = standard error. ***, ** and * represent statistical significance at 1 percent, 5 percent, and 10 percent respectively. Difference-in-differences results are presented. Models include PBP fixed effects, year fixed effects, and an intercept (not reported). Coefficient estimates are shown with cluster-robust standard errors in the next column. All results are weighted to ensure parallel trends in the pre-VBID period. The "enrollment weighted" results are further weighted to reflect PBP-level enrollment. Per-Beneficiary Per-Month Cost of Additional Services taken from BPT Worksheet 4, Subsection C., column (p), line u. Non-Benefit Expenses for MA-Covered Services taken from BPT Worksheet 4, Subsection C., column (o), line v6. Non-Benefit Expenses for Mandatory Supplemental Services taken from BPT Worksheet 4, Subsection C., column (r), line v6.

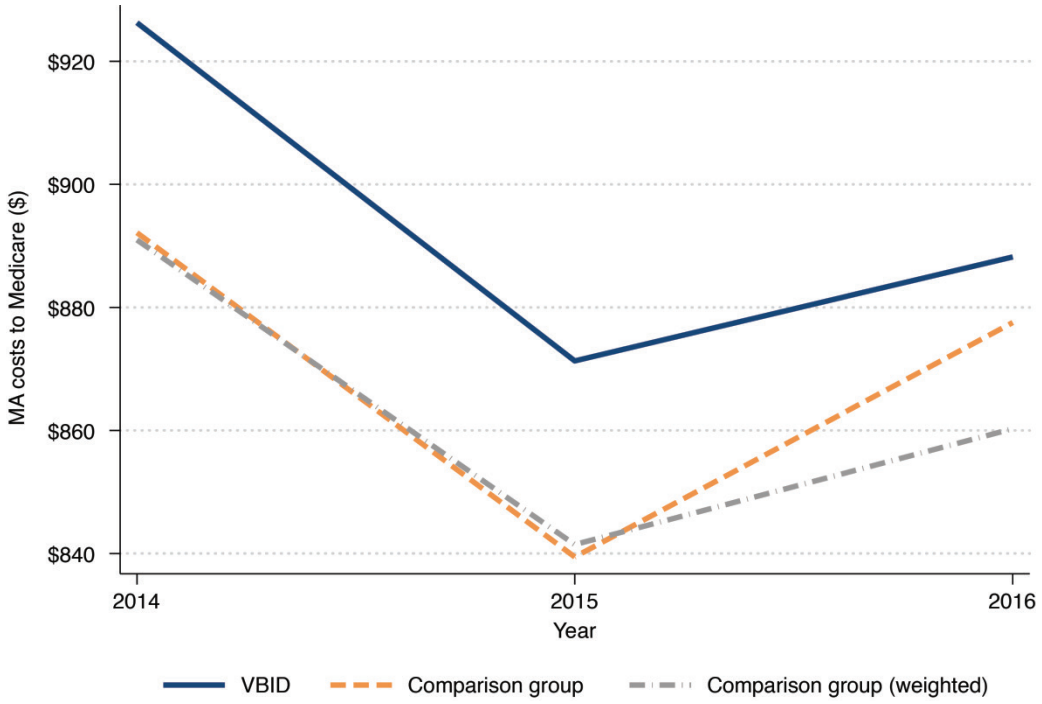
Costs to Medicare

Parallel Trends

Figures I.10 and I.11 present unadjusted trends in MA-PD PBP bids, costs to Medicare, and PBP spending for VBID and matched comparison PBPs. The sample used in all figures is restricted to PBPs offering Part D benefits.

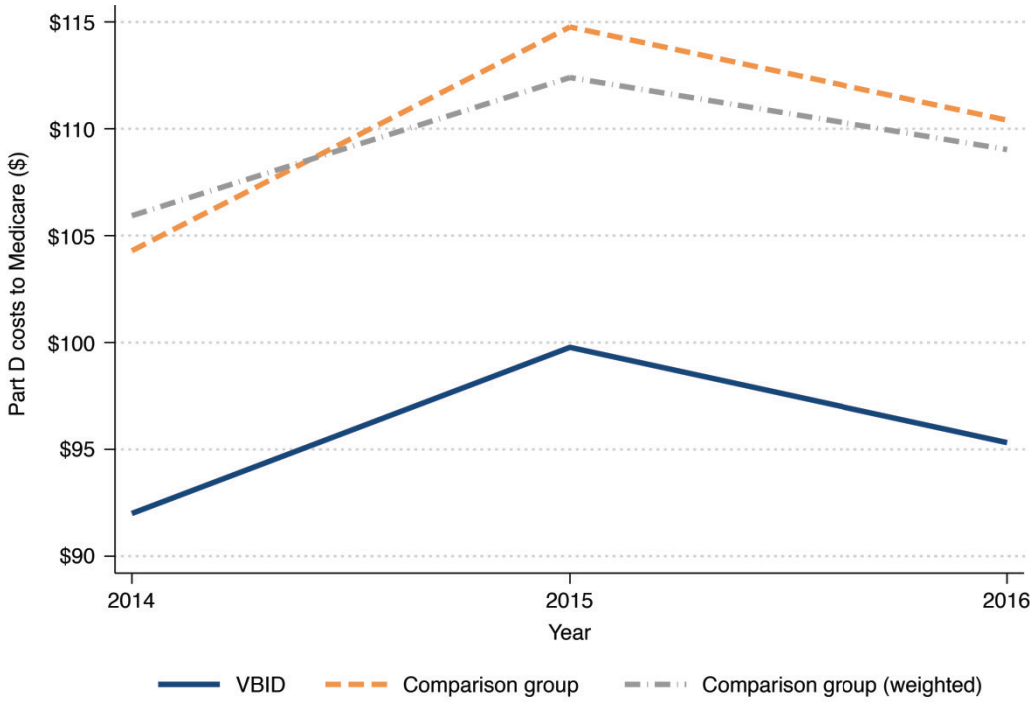
Table I.16 reports test statistics and p-values for the hypothesis that trends were parallel between VBID and comparison PBPs prior to VBID. We failed to reject the hypothesis of parallel trends at the 10-percent level for either measure of costs to Medicare. However, the F-statistics in Table I.16 and the graphical evidence in Figures I.10 and I.11 suggest that inverse propensity weights resulted in even greater similarity of the pre-VBID trends. All difference-in-differences estimates for costs to Medicare are estimated using WLS regression with inverse propensity weights.

Figure I.10. MA Costs to Medicare by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.21 before weighting and 0.99 after weighting.

Figure I.11. Part D Costs to Medicare by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.48 before weighting and 0.99 after weighting.

Table I.16. Test Statistics and p-Values for Parallel Trends Assumptions, Unweighted and with Inverse Propensity Weights

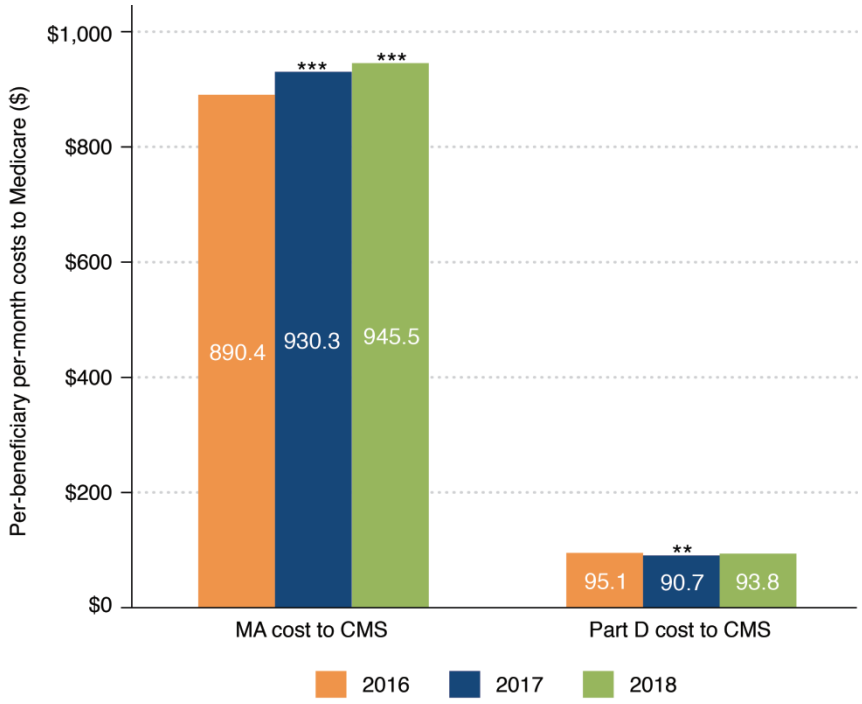
Outcome	Unweighted F-Statistic	p-Value	Weighted F-Statistic	p-Value
Per-beneficiary per-month cost to Medicare for MA	1.59	0.21	0.01	0.99
Per-beneficiary per-month cost to Medicare for Part D	0.73	0.48	0.01	0.99

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively.

Descriptive Results

Figure I.12 shows average costs to Medicare for VBID-participating PBPs from 2016 through 2018. MA costs rose throughout this period, while Part D costs were essentially flat.

Figure I.12. Descriptive Statistics for Costs to Medicare



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing VBID PBP costs to Medicare in each model year with costs in 2016. Sample size (number of unique PBPs in each year) ranges between 32 and 33.

Table I.17 reports summary statistics for costs to Medicare.

Table I.17. Summary Statistics for Costs to Medicare (in Dollars), MA-PD Plans

Variable	VBID Average	VBID SD	VBID Min.	VBID Median	VBID Max.	VBID N	Comp. Average	Comp. SD	Comp. Min.	Comp. Median	Comp. Max.	Comp N.
MA costs to Medicare												
2014	926	210	571	882	1,314	31	891	179	500	870	1,290	32
2015	873	201	542	836	1,240	32	841	155	506	822	1,264	32
2016	890	200	553	887	1,251	33	860	152	531	833	1,351	32
2017	930	207	559	914	1,350	33	913	160	496	882	1,341	32
2018	945	205	562	924	1,318	32	934	163	518	926	1,426	32
Part D costs to Medicare												
2014	93	23	53	88	150	31	106	20	33	108	154	32
2015	101	27	58	95	177	32	112	20	69	113	178	32
2016	95	32	51	88	183	33	109	21	66	108	191	32
2017	91	37	41	80	196	33	113	29	48	103	206	32
2018	94	36	48	84	193	32	117	34	45	115	207	32

Regression Results

Table I.18 shows the difference-in-differences model results for PBP spending. These results are discussed in Chapter 8.

Table I.18. Difference-in-Differences Model Results for Costs to Medicare and Spending Outcomes (in Dollars)

	MA Costs to Medicare	SE	Part D Costs to Medicare	SE
Number of observations	322		322	
Intercept	875.04***	3.88	102.43***	1.13
VBID indicator	(omitted)		(omitted)	
VBID Year 1	−63.01***	12.24	−16.65***	3.87
VBID Year 2	−140.72***	14.45	−64.27***	6.33
Year (2016 reference)				
2014	30.55***	8.41	−3.37	2.15
2015	−19.29***	5.22	3.28**	1.55
2017	115.77***	4.20	20.23***	1.15
2018	214.52***	11.25	71.82***	3.09
VBID indicator*VBID Year 1	−13.49	15.14	−9.63**	4.53
VBID indicator*VBID Year 2	−8.63	16.01	−7.21	6.73

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. Models include PBP fixed effects. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Sensitivity Analysis: MA Costs to Medicare Regression Results for All VBID PBPs

As a sensitivity analysis, we used the difference-in-differences regression model presented above to estimate change in MA outcomes for the full sample of all VBID PBPs, including both MA-PD PBPs and MA-only PBPs. Results are reported in Table I.19. As in our analysis of MA Costs to Medicare for MA-PD PBPs, we find that changes in health care costs associated with VBID are generally small and statistically insignificant.

Sensitivity Analysis: Components of Part D Costs to Medicare

In Chapter 8, we noted some tension between our results for Part D bids (which were insignificantly different from comparison PBPs in 2017 but significantly lower for 2018 and 2019) and Part D costs to Medicare (which were significantly lower in 2017 but insignificantly different from comparison PBPs in 2018). Estimating our difference-in-difference models on the different components that affect Part D costs to Medicare pointed toward two factors that explained why 2017 Part D costs were lower while standardized Part D bids remained unchanged. First, per-beneficiary per-month reinsurance payments to PBPs appear to have grown more slowly in VBID plans than in comparison PBPs between 2016 and 2017: Per-beneficiary

Table I.19. Difference-in-Differences Model Results Using Pooled MA-PD and MA-Only PBPs for MA Costs to Medicare (in Dollars)

	MA Costs to Medicare	SE
Number of observations	452	
Intercept	872.7***	4.877
VBID indicator	(omitted)	
VBID Year 1	−56.04***	16.34
VBID Year 2	−128.9***	20.94
VBID Year 3		
Year (2016 reference)		
2014	37.41***	7.619
2015	−14.16***	4.564
2017	119.8***	3.743
2018	219.5***	9.209
2019		
VBID indicator*VBID Year 1	−21.48	17.88
VBID indicator*VBID Year 2	−26.42	22.06
VBID indicator*VBID Year 3		

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. Models include PBP fixed effects. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

per-month reinsurance payments to VBID plans in 2017 were $7.05 lower than would have been expected in the absence of VBID. Second, point estimates suggest that, in 2017, VBID was associated with an insignificant decrease of $1.95 per beneficiary per month in the direct subsidy, which in turn was driven primarily by an insignificant decrease of $2.19 per-beneficiary per month in the risk-adjusted Part D bid. Table I.20 presents estimates supporting this discussion.

Table I.20. Model Results for the Components of Part D Costs to Medicare (in Dollars)

	Per-Beneficiary Per-Month Reinsurance Payments	SE	Per-Beneficiary Per-Month Direct Subsidy for Part D	SE	Risk-Adjusted Part D Bid	SE
Number of observations	322		322		322	
Intercept	52.40***	0.99	67.12***	1.51	67.12***	1.51
VBID indicator	(omitted)		(omitted)		(omitted)	
VBID Year 1	9.42***	3.32	−9.75	6.05	−9.75	6.04
VBID Year 2	−0.09	5.84	9.61	5.12	9.61*	5.12

	Per-Beneficiary Per-Month Reinsurance Payments	SE	Per-Beneficiary Per-Month Direct Subsidy for Part D	SE	Risk-Adjusted Part D Bid	SE
Year (2016 reference)						
2014	−14.60***	2.05	12.01***	1.93	12.01***	1.94
2015	−4.88***	1.49	5.70***	1.20	5.70***	1.20
2017	−1.53	1.06	10.79***	0.86	10.79***	0.86
2018	12.12***	2.76	−8.39***	1.95	−8.39***	1.95
VBID indicator*VBID Year 1	−7.05*	3.91	−2.19	6.20	−2.19	6.20
VBID indicator*VBID Year 2	−5.24	6.57	−12.14**	5.23	−12.14**	5.23

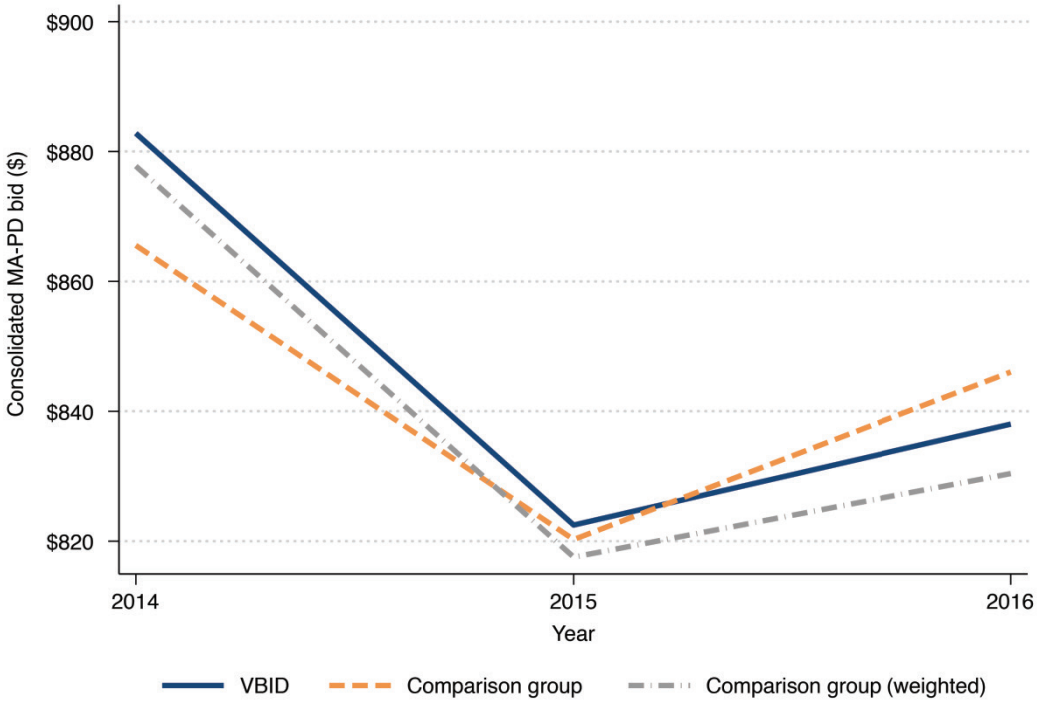
NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Plan Benefit Package Bids

Parallel Trends

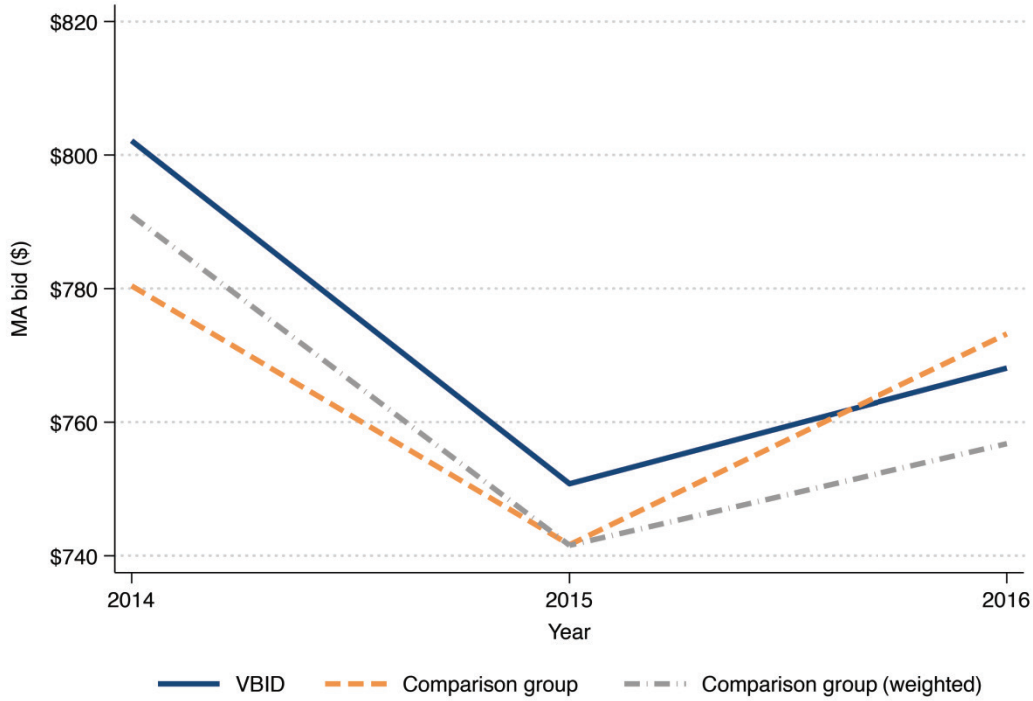
Figures I.13–I.15 present unadjusted trends in MA-PD PBP bids for VBID and matched comparison PBPs. The sample used in all figures is restricted to PBPs offering Part D benefits.

Figure I.13. Consolidated MA-PD PBP Bids by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.13 before weighting and 0.81 after weighting.

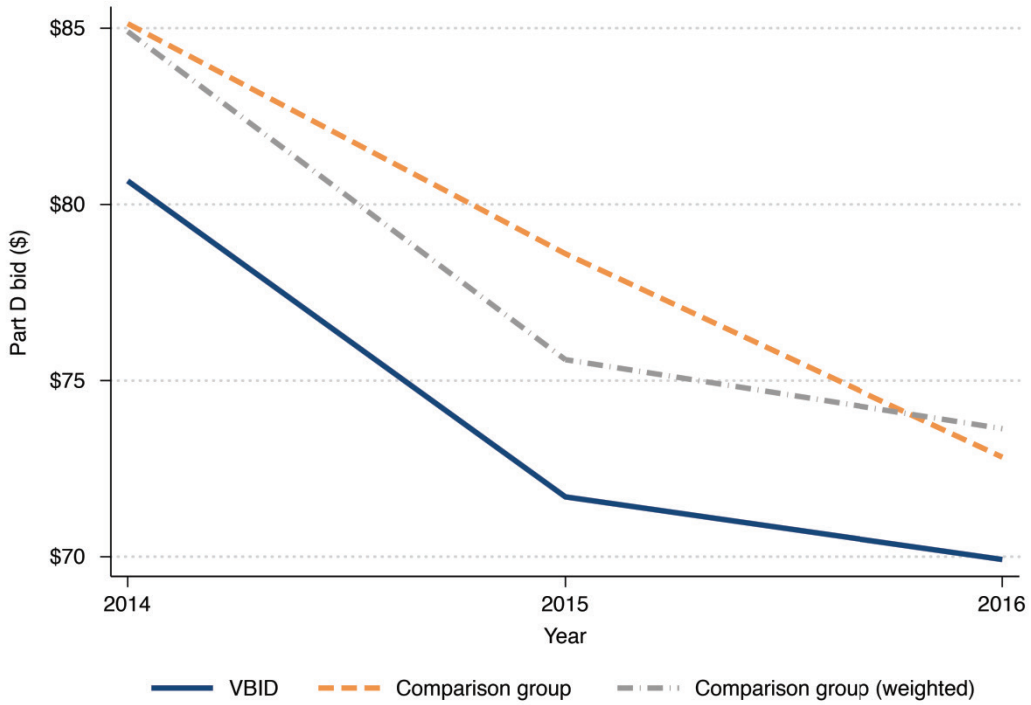
Figure I.14. MA Bids by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.08 before weighting and 0.98 after weighting.

For each outcome variable, we tested for parallel trends using the approach described in Appendix D, using the unweighted data and then repeating the test on the weighted data to verify that our weights improved the balance between VBID and comparison PBPs, in terms of pre-VBID trends in outcomes. Table I.21 reports the F-statistics and p-values from these tests. Using unweighted data, we reject parallel trends in MA bids at the 10-percent level. Although we did not reject parallel trends in MA-PD and Part D bids at the 10-percent level, Figures I.13 and I.15 show some suggestive evidence of nonparallel trends. Table I.21 shows that the use of inverse propensity weights results in parallel trends between VBID and comparison plans prior to VBID implementation. All difference-in-differences estimates for PBP bids are estimated using WLS regression with inverse propensity weights.

Figure I.15. Part D Bids by VBID Status, 2014–2016



NOTE: Sample restricted to MA-PD PBPs and matched comparison PBPs. Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.14 before weighting and 0.95 after weighting.

Table I.21. Test Statistics and p-Values for Parallel Trends Assumptions, Unweighted and with Inverse Propensity Weights

Outcome	Unweighted F-Statistic	p-Value	Weighted F-Statistic	p-Value
MA-PD bid	2.07	0.13	0.22	0.81
MA bid	2.59	0.08	0.08	0.98
Part D bid	2.05	0.14	0.05	0.95

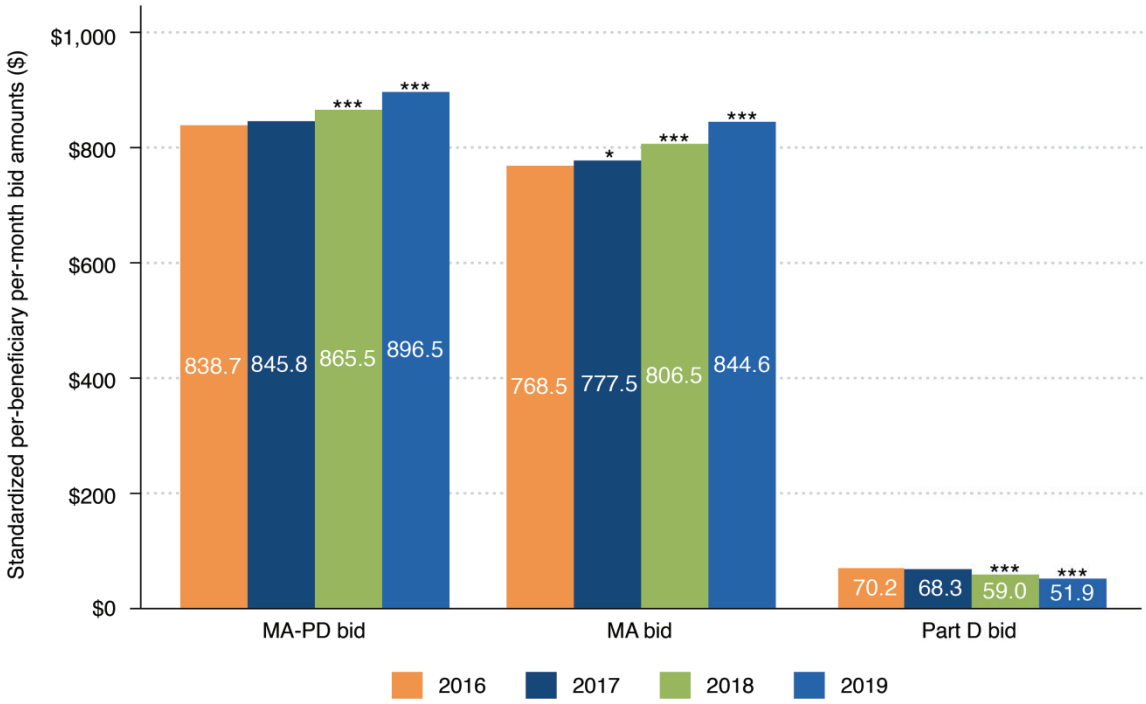
NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. F-statistic for test of hypothesis that pre-VBID trends are parallel between VBID and comparison PBPs between 2014 and 2016. Significance assessed using $F_{2,64}$ distribution.

Descriptive Results

Figure I.16 shows average PBP bids from 2016 through 2019 for VBID-participating PBPs. MA-PD bids increased for VBID-participating PBPs, driven primarily by increasing bids for MA.

Table I.22 shows summary statistics for PBP bids.

Figure I.16. Descriptive Statistics for PBP Bids



NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively, from a paired t-test comparing VBID PBP bids in each model year with bids in 2016. Sample size (number of unique PBPs in each year) ranges between 32 and 33. See Appendix I for details.

Table I.22. Summary Statistics for PBP Bids (in Dollars), MA-PD Plans

Variable	VBID Average	VBID SD	VBID Min.	VBID Median	VBID Max.	VBID N	Comp. Average	Comp. SD	Comp. Min.	Comp. Median	Comp. Max.	Comp N.
MA bids												
2014	801	53	724	789	898	31	791	79	659	763	935	32
2015	752	64	634	737	892	32	742	67	631	728	884	32
2016	769	67	610	761	906	33	757	75	644	754	906	32
2017	777	72	671	765	945	33	773	86	590	769	937	32
2018	807	78	633	798	983	33	789	74	630	787	932	32
2019	845	89	670	843	1,033	32	809	88	638	796	979	32
Part D bids												
2014	80	20	49	73	124	31	85	13	62	86	121	32
2015	72	11	49	69	94	32	76	21	40	81	115	32
2016	70	15	39	69	97	33	74	22	35	79	105	32
2017	68	16	41	66	99	33	72	26	34	78	190	32
2018	59	15	30	56	90	33	74	22	44	74	166	32
2019	52	13	27	50	76	32	68	27	36	65	191	32
MA-PD bids												
2014	882	59	797	865	1,017	31	878	81	755	850	1,036	32
2015	823	69	715	815	986	32	818	73	704	791	981	32

Variable	VBID Average	VBID SD	VBID Min.	VBID Median	VBID Max.	VBID N	Comp. Average	Comp. SD	Comp. Min.	Comp. Median	Comp. Max.	Comp N.
2016	839	69	686	829	979	33	830	78	714	804	1,004	32
2017	846	76	737	827	1,025	33	846	88	701	827	1,018	32
2018	866	77	723	857	1,023	33	861	82	745	849	1,018	32
2019	896	88	746	890	1,074	32	878	93	703	857	1,060	32

Regression Results

Table I.23 shows the difference-in-differences model results for MA-PD bids, MA bids, and Part D bids. These results are discussed in Chapter 8.

Table I.23. Difference-in-Differences Model Results for MA-PD Bids (in Dollars)

	MA-PD Bid	SE	MA Bid	SE	Part D Bid	SE
Number of observations	387		387		387	
Intercept	834.52***	3.11	762.67***	3.44	71.85***	1.21
VBID indicator	(omitted)		(omitted)		(omitted)	
VBID Year 1	−24.12**	9.45	−14.61	8.94	−9.97***	2.95
VBID Year 2	−42.97***	15.78	−51.13***	16.24	11.03***	3.25
VBID Year 3	−72.50***	18.53	−98.94***	21.35	26.81***	4.74
Year (2016 reference)						
2014	43.90***	5.87	31.99***	5.78	10.87***	2.08
2015	−14.56***	4.39	−16.41***	4.69	1.62	1.07
2017	37.83***	3.17	29.43***	3.08	7.99***	0.91
2018	72.04***	6.43	82.71***	6.85	−11.11***	2.07
2019	118.21***	11.82	150.21***	12.70	−32.45***	2.96
VBID indicator*VBID Year 1	−5.18	10.74	−5.90	10.50	1.20	3.29
VBID indicator*VBID Year 2	−1.15	16.33	5.93	17.00	−9.93***	3.47
VBID indicator*VBID Year 3	17.24	17.22	28.60	20.24	−11.69***	4.59

NOTE: ***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column.

Sensitivity Analysis: Medicare Advantage Bid Regression Results for All Value-Based Insurance Design Plan Benefit Packages

As a sensitivity analysis, we used the difference-in-differences regression model presented above to estimate change in MA outcomes for the full sample of all VBID PBPs, including both MA-PD PBPs and MA-only PBPs (Table I.24). As in our analysis of MA bids for MA-PD PBPs, we find that changes in health care costs associated with VBID are generally small and statistically insignificant.

Table I.24. Difference-in-Differences Model Results Using Pooled MA-PD and MA-Only PBPs for MA Bids (in Dollars)

	MA Bids	SE
Number of observations	543	
Intercept	765.6***	2.799
VBID indicator	(omitted)	
VBID Year 1	−12.30	7.576
VBID Year 2	−48.26***	15.54
VBID Year 3	−94.05***	17.51
Year (2016 reference)		
2014	33.49***	5.186
2015	−17.28***	3.512
2017	29.64***	2.482
2018	81.41***	5.362
2019	151.4***	9.668
VBID indicator*VBID Year 1	−6.697	8.727
VBID indicator*VBID Year 2	2.060	15.92
VBID indicator*VBID Year 3	25.17	16.71

NOTE:***, **, and * represent statistical significance at the 1-, 5-, and 10-percent levels, respectively. Difference-in-differences results with dynamic effects are presented. Coefficient estimates are shown with robust SEs in the next column. Models include PBP fixed effects. The intercept is defined as the value that makes the predicted value when all explanatory variables are set to their sample averages equal to the sample average of the outcome variable.

Appendix J. Detailed Analysis of the Effects of Value-Based Insurance Design Part D Intervention on Utilization and Adherence

In Chapters 6 and 7, we present findings on the effect of VBID on utilization and health outcomes, treating VBID as one intervention. However, there is variation in the type of VBID interventions that POs have implemented, as discussed in Chapter 2. In particular, three POs implemented "Part D" interventions, that is, interventions that reduce cost-sharing for high-value drugs. To the extent that reduced cost-sharing leads to higher utilization of high-value drugs, one may expect to see an effect of Part D intervention on utilization of health services and possibly adherence. In this appendix, we consider whether beneficiaries in VBID plans that offer a Part D intervention ("D plans") experience different outcomes from their counterparts in VBID with no Part D intervention and from matched comparators who are not in VBID plans. We consider five outcome variables (number of 30-day refills, any inpatient stay, any ED admission, adherence to cholesterol-lowering drugs, and adherence to hypertension medication) and perform two types of analyses, corresponding to the following two choices of comparison groups:

- **Analysis 1:** We compare individuals in D plans with matched comparators in VBID plans without Part D intervention.
- **Analysis 2:** We compare individuals in D plans with matched comparators who are not in VBID plans.

For both analyses, we used the same regression methodology and control variables described in Appendix G. The only difference is that the key intervention variable in this appendix is not whether someone is in a VBID PBP, but rather whether someone is in a VBID PBP with a Part D intervention. All the data were reweighted using the entropy balance method (Hainmueller, 2012) discussed in the "Difference-in-Differences Model and Parallel Trends Assumption" section of Appendix D. In the following, we present the results of the regression Analyses 1 and 2. These results only include data from two of the three POs with Part D intervention (PO D and PO G), since PO J joined the model late and its data are only suitable for descriptives at the moment. Therefore, to provide some visibility to the data of PO J, we conclude this Appendix with descriptives of PO D, G, and J relating to drug adherence.

Analysis 1: Individuals in VBID Plans with Part D Interventions and Matched Comparator in VBID Plans Without Part D Intervention

The composition of the VBID plans with and without Part D intervention is quite different, making it challenging to perform fair comparisons between the two groups. As shown in Table J.1, individuals in POs offering a Part D intervention are younger, less likely to be disabled or receive low-income support, and generally in better health than those in VBID POs with no Part D component.

Table J.1. Descriptive Statistics for VBID plans with and Without Part D Intervention, 2016

Covariate	VBID No Part D	VBID with Part D
Age	76.8	74.2
Low-income support (percent)	16	8
HCC risk score	1.91	1.06
RxHCC risk score	1.37	0.95
Disabled (percent)	19	12
Died before 2018 (percent)	7	3

To account for the differences in population composition, prior to running the regressions we performed 1-to-1 coarsened exact matching (CEM), obtaining intervention and comparison groups with similar distributions of confounders prior to the intervention. Three different matched data sets were produced, because different missing patterns of the dependent variables lead to three different subpopulations: The three utilization measures have virtually no missing values, the cholesterol medication adherence measure is only available for a small group of individuals in PO G, and the hypertension medication adherence is available for PO D and PO G. Univariate and multivariate imbalance measures for the three matched data sets are shown in Table J.2. The figures corresponding to variable names are univariate imbalance measures. Matching was performed using the R package CEM. The variables along which the data set displays most imbalance are OOP maximum, premium, and HCC risk score. Although in some cases the matching process significantly reduces the size of the data set, it greatly improves the comparability of the two groups, as demonstrated by the large reduction in the multivariate imbalance measure displayed in all three cases. The choice of the matching variables and the level of coarseness applied to each have been dictated purely by the criterion of reducing the multivariate imbalance measure.

Table J.2. CEM Quality Statistics for Matching Beneficiaries in VBID with and Without Part D Intervention, 2016

	Prior to Matching	After Matching
Matching for utilization analysis		
Female (percent)	0.003	0
Age	0.182	0
Low-income support (percent)	0.081	0
Dual eligible (percent)	0.047	0
Disabled (percent)	0.08	0
OOP maximum ($)	0.288	0
Premium ($)	0.338	0
HCC risk score	0.427	0
Multivariate imbalance measure	0.799	0
Size of treatment group	27,385	7,169
Size of control group	42,526	7,169

	Prior to Matching	After Matching
Matching for cholesterol medication adherence analysis		
Age	0.108	0
Disabled (percent)	0.014	0
HCC risk score	0.132	0.075
Low-income support (percent)	0.008	0
Premium ($)	0.196	0
OOP maximum ($)	0.6	0
Multivariate imbalance measure	0.861	0.378
Size of treatment group	455	442
Size of control group	9,284	442
Matching for hypertension medication adherence analysis		
Age	0.245	0
HCC risk score	0.603	0.131
Premium ($)	0.149	0
OOP maximum ($)	0.217	0
Multivariate imbalance measure	0.859	0.482
Size of treatment group	7,459	1,418
Size of control group	5,644	1,418

In Table J.3, we report the regression results for Analysis 1. Statistical tests for the parallel trend hypothesis failed in most cases, although after reweighting the hypothesis holds. Therefore, we show in Table J.3 the results corresponding to the specification that assumes parallel trends. Importantly, results of the regressions with and without assumption of parallel trends were qualitatively similar and led to the same conclusions. In the second column of the table, we show the association between the outcome and the presence of a Part D intervention. Technically, this is the effect corresponding to the interaction of Part D with a dummy for year 2017, measured on the original scale of the variable (computed using the predictive margins option in Stata).

The table shows that the presence of a Part D intervention is associated with a statistically significant increase in the number of 30-days prescription fills. Because the average number of 30-days prescription fills for the comparison group was 47.6 in 2016, this corresponds to a small 1.6-percent increase in utilization. The table also shows that there was a statistically significant reduction in the probability of any inpatient stay and any ED admission. Because the corresponding baseline probabilities in year 2016 for the comparison group were 0.25 and 0.41, the table implies that Part D is associated with a reduction in probability of any hospital or ED admission of 8.8 and 5 percent, respectively.

Regarding adherence, Table J.3 shows an association between Part D intervention and increased adherence. However, the effect is only statistically significant for hypertension, possibly because cholesterol adherence was measured only for a small group of approximately 450 individuals in PO G. For hypertension medication, the average adherence in the comparison group was 0.82 in 2016, and therefore the table suggests a 2.9-percent increase in adherence

associated with Part D intervention. Notably, PO D's intervention was focused specifically on eliminating copayments for hypertension drugs.

Table J.3. Regression Result, Change in Utilization and Adherence Measures, Actual Versus Expected if VBID with Part D Intervention Had Not Been Available

Outcome	Effect	SE	z	p-Value	Lower Bound	Upper Bound
Number of 30-day fills	0.785	0.230	3.418	0.001	0.335	1.235
Any inpatient stay	−0.022	0.007	−2.960	0.003	−0.036	−0.007
Any ED admission	−0.020	0.008	−2.351	0.019	−0.036	−0.003
Adherence to cholesterol medication	0.007	0.026	0.272	0.785	−0.043	0.057
Adherence to hypertension medication	0.024	0.006	3.939	0.000	0.012	0.036

NOTE: Estimates based on difference-in-differences regressions comparing beneficiaries in POs with Part D interventions to matched comparisons in VBID POs without Part D interventions. Sample size was 50,004 beneficiary years for utilization regressions, 2,627 for cholesterol adherence, and 37,724 for hypertension adherence.

Analysis 2: Individuals in Value-Based Insurance Design Plans with Part D Intervention and Matched Comparators Not in Value-Based Insurance Design Plans

In this section, we compare beneficiaries in D plans with a matched set of comparators who are not in VBID plans. We used the same regression methodology applied in the previous section and report the results in Table J.4.

Table J.4. Regression Result, Change in Utilization and Adherence Measures, Actual Versus Expected in VBID if Part D Intervention Had Not Been Available

Outcome	Effect	SE	z	P-Value	Lower Bound	Upper Bound
Number of 30-day fills	0.430	0.117	3.669	0.000	0.200	0.660
Any inpatient stay	0.002	0.003	0.746	0.456	−0.004	0.009
Any ED admission	−0.004	0.004	−0.854	0.393	−0.012	0.005
Adherence to cholesterol medication	0.023	0.025	0.930	0.352	−0.026	0.073
Adherence to hypertension medication	0.006	0.005	1.133	0.257	−0.004	0.015

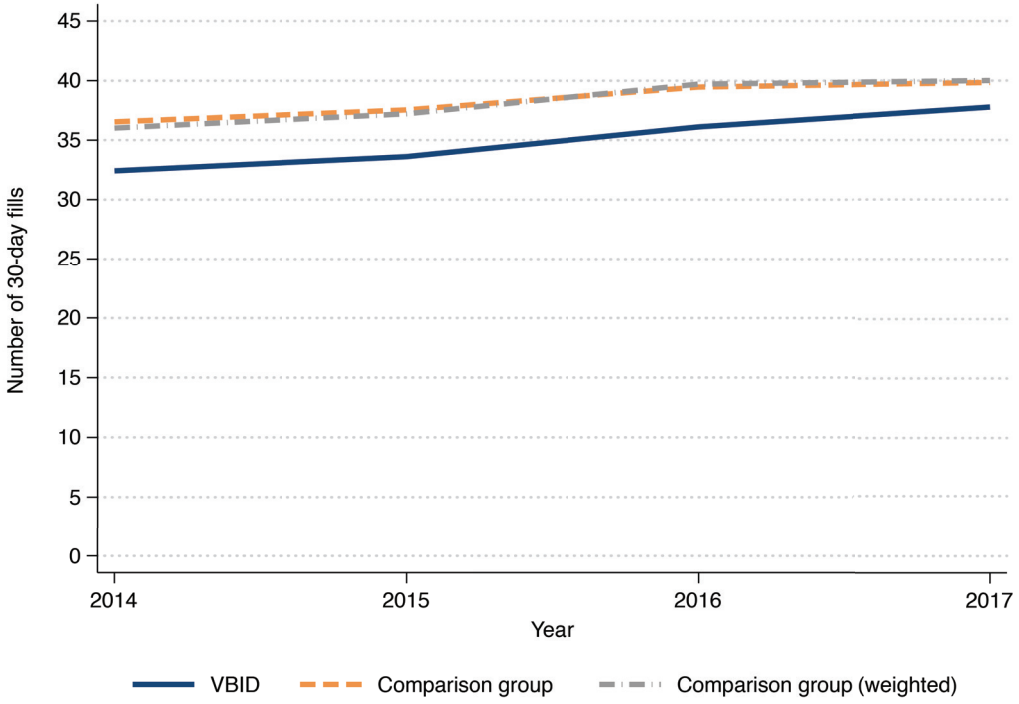
NOTE: Estimates based on difference-in-differences regressions comparing beneficiaries in POs with Part D interventions to matched comparators in non-VBID POs. Sample size was 187,423 beneficiary years, for utilization regressions, 3,159 for cholesterol adherence, and 48,272 for hypertension adherence.

The table shows the VBID Part D intervention is associated with a statistically significant higher number of 30-day prescription fills, although the size of effect is small and corresponds to an increase of 1.1 percent with respect to matched comparators. Unlike with Analysis 1, we did not find evidence of a statistically significant association between Part D intervention and probability of hospital or ED admissions. The sign for ED admission is negative and the one

for hospital admission is positive, but they are both very small and have large p-values. The association between Part D and adherence to cholesterol and hypertension medication is positive, as with Analysis 1, but not statistically significant at conventional levels. Therefore, overall the analysis only shows a small but significant positive association between Part D intervention and the number of 30-day fills.

To provide a more concrete view of how each of the five outcomes varies over time (from 2014 to 2017) and across different groups, we report in Figures J.1–J.5 the time trends for the two groups of individuals of Analysis 2. In addition to the trends over time, we also report the "reweighted" trend, obtained using the entropy balancing technique.

Figure J.1. Average Number of 30-Day Fills for Beneficiaries in VBID Part D Plans and Matched Comparators, 2014–2017

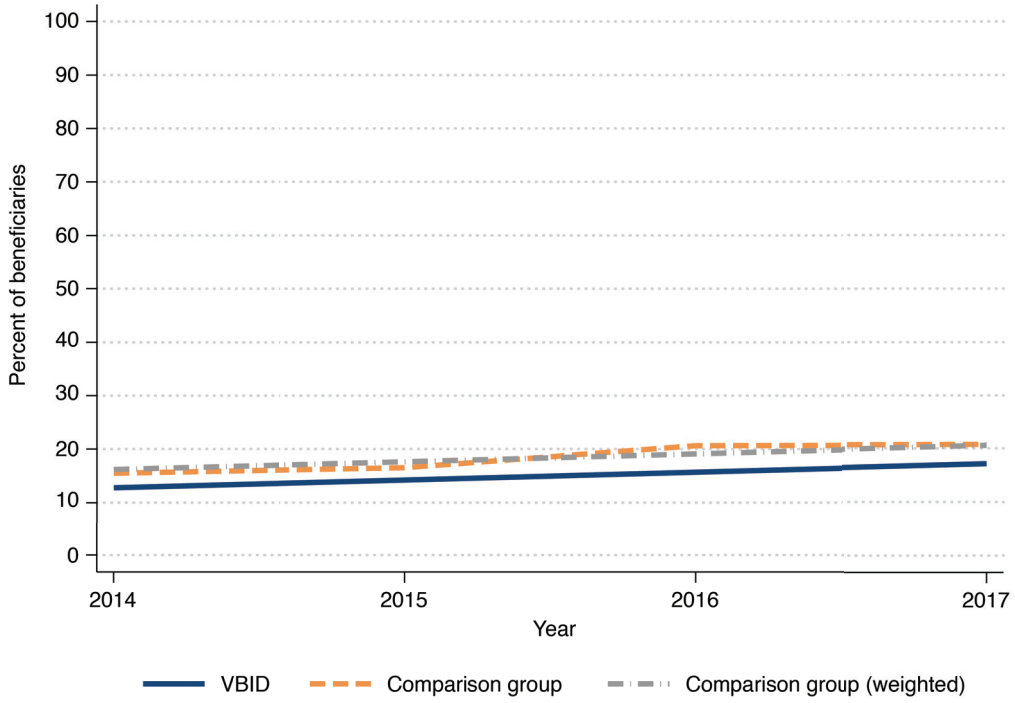


NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.00 before weighting and 0.00 after weighting.

Figure J.1 is an example of how the parallel trends hypothesis may be rejected by statistical tests, as it was the case in our regression analysis, but hold well visually. In this case, the small difference in trends is well compensated by the reweighting of the comparison group performed using entropy balancing.

Figure J.2 is an example of how well the entropy balancing works: In the unweighted data the trend for the comparison group in the period 2015–2016 is much steeper than in the Part D group, but reweighting the data perfectly matches the trends.

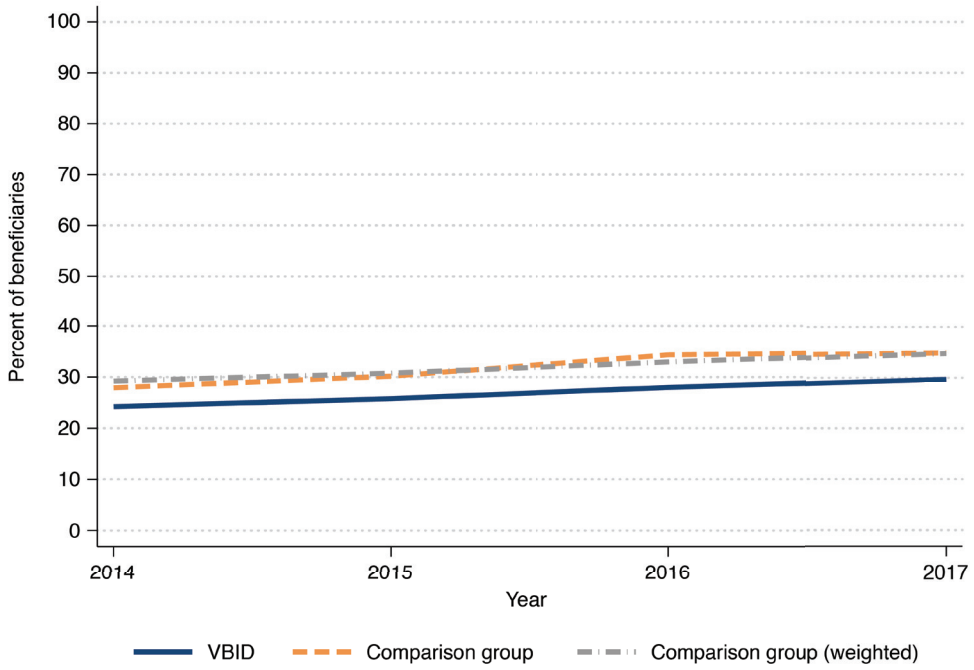
Figure J.2. Average Probability of Any Inpatient Stay for Beneficiaries in VBID Part D Plans and Matched Comparators, 2014–2017



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.17 before weighting and 0.00 after weighting.

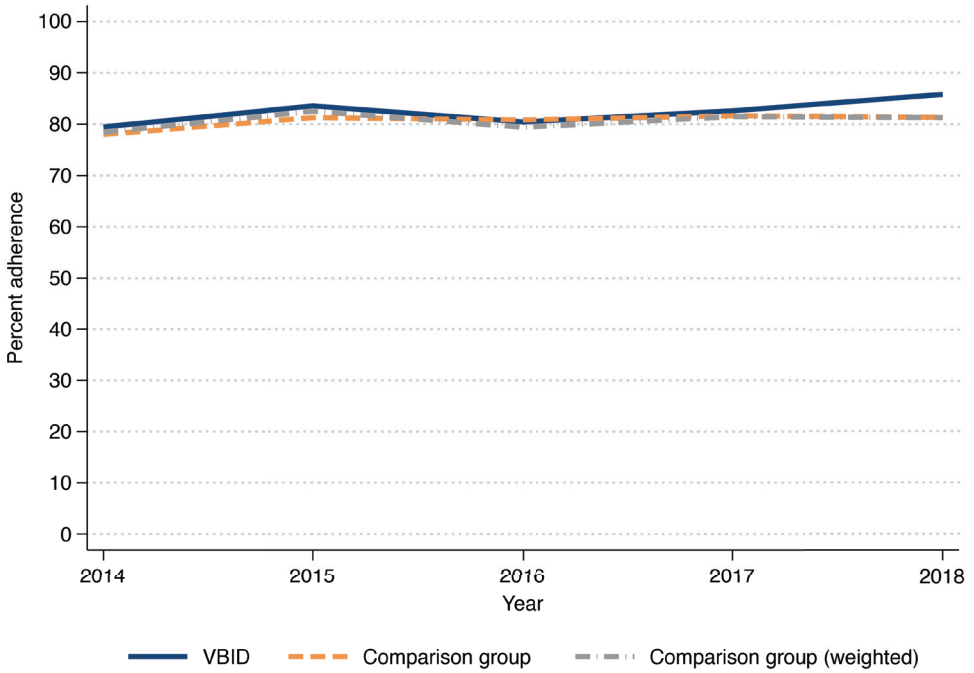
Figure J.3 shows that the temporal pattern of the comparison group seen in Figure J.2 is not an isolated case. We find in this group that, for inpatient stays, ED admission, and other utilization variables not shown here such as primary care visits or laboratory claims, utilization sharply increases in year 2016 and then flattens in year 2017. This is in contrast with the Part D group, which tends to exhibit steadily increasing trends. A breakdown of trends in the comparison group by POs shows that there is variation among POs: Some show steady trends, and some show the angled patterns of Figures J.2 and J.3. However, it is currently not known what characteristic of the POs may be associated with such patterns. Notice that Figures J.1–J.3 confirm the notion that people in the Part D group are somewhat healthier than their counterparts, because they exhibit lower utilization rates. Overall, however, there is no discernible difference in utilization after year 2016, as suggested by the regression analysis and Table J.4. Figures J.4 and J.5 make it apparent why the regression analysis does not show any significant effect of VBID with Part D intervention on adherence. The trends in the treatment and comparator groups are very similar to each other, and even if a closeup of the picture would show the treatment group performing slightly better than the comparator, the differences in outcomes are simply too small to be significant.

Figure J.3. Average Probability of Any ED Admission for Beneficiaries in VBID Part D Plans and Matched Comparators, 2014–2017



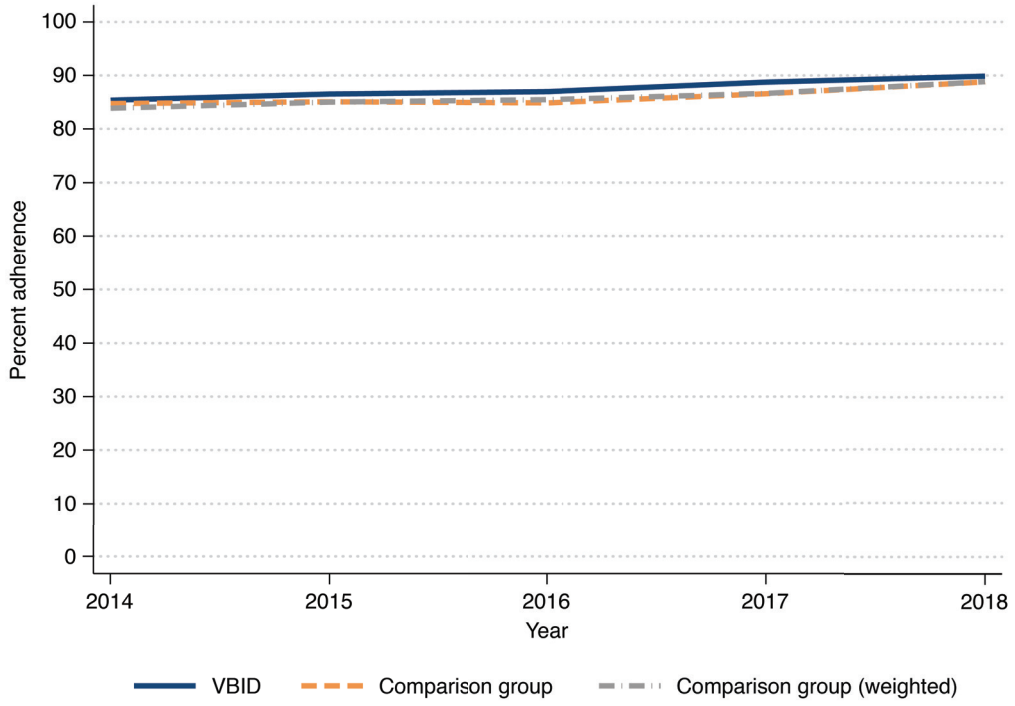
NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.00 before weighting and 0.00 after weighting.

Figure J.4. Average Adherence for Cholesterol Medication for Beneficiaries in VBID Part D Plans and Matched Comparators, 2014–2018



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.18 before weighting and 0.55 after weighting.

Figure J.5. Average Adherence for Hypertension Medication for Beneficiaries in VBID Part D Plans and Matched Comparator, 2014–2017



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.43 before weighting and 0.93 after weighting.

Descriptive Analysis of Plans Offering a Part D Intervention

The regression analyses presented in the previous section include data from two of the three POs that offer a Part D intervention, PO D and PO G. Some data for PO J, which also offers a Part D intervention, are available but not sufficient to support a regression analysis. Therefore, in this section, we present a simpler, descriptive analysis that includes all three POs and focus on adherence measures for medications related to cholesterol, hypertension, and diabetes.

The analysis is complicated by the fact that different POs have implemented different Part D interventions. PO D eliminated cost-sharing for select hypertension drugs; PO G eliminated copayments for select generic drugs for CHF, conditional on participation in CM; and PO J eliminated copays for select drugs for CAD.

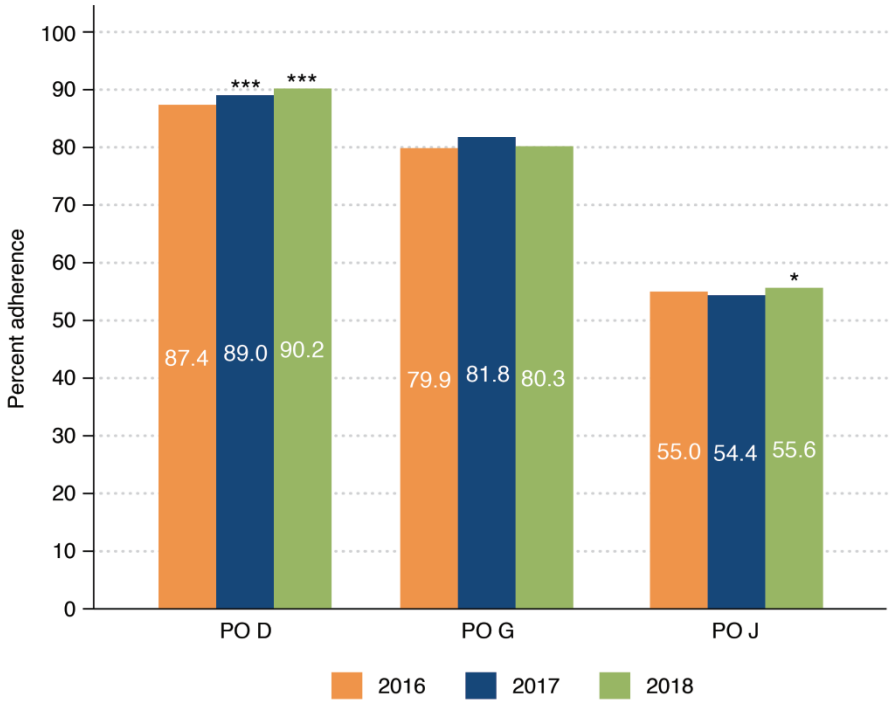
In addition, adherence measures are not uniformly available across POs. In Table J.5, we report for each of the POs and for each the adherence measures, the number of observations available in year 2016.

Table J.5. Number of Observations of Adherence Measures by POs, 2016

PO	Cholesterol	Hypertension	Diabetes
PO D	0	7,052	0
PO G	455	407	0
PO J	32,377	32,377	32,377

In Figure J.6, we report the average adherence to hypertension medication for all three plans offering a Part D intervention for years 2016–2018, covering the period before and after the implementation of VBID. We use a two-sided Wilcox test to assess the significance of changes between year 2017 and 2016 and between year 2018 and 2016.

Figure J.6. Average Adherence to Hypertension Medication for Plans Offering Part D Intervention, 2016–2018



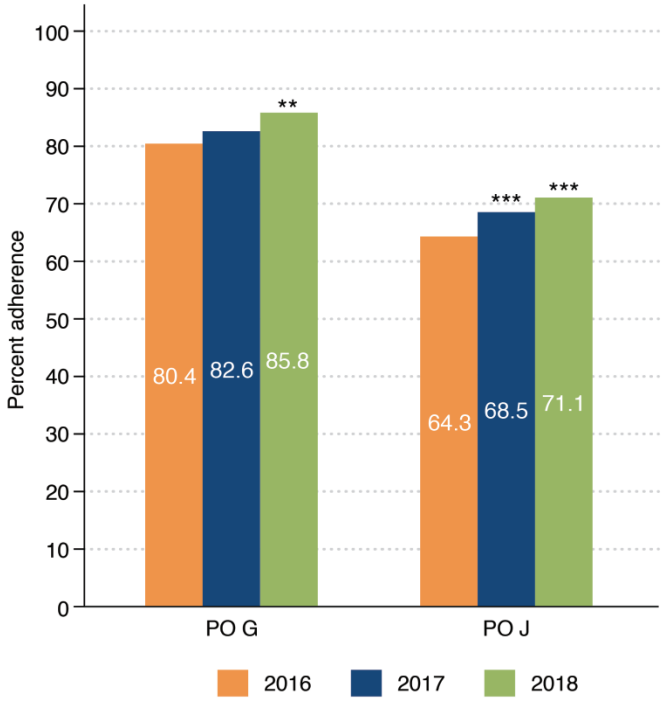
NOTE: ***, **, and * represent statistically significant differences at the 1-, 5-, and 10-percent levels, respectively, from the 2016 estimate.

A striking feature of this figure is the disparity in adherence between PO J and the other two POs, which is currently unexplained. As part of its Part D intervention, PO D targeted hypertension drugs, and as expected we observe a statistically significant increase in adherence postintervention, albeit relatively small.

In Figure J.7, we report the adherence to cholesterol lowering drugs, for which we have data from only PO G and PO J. There is still a sizable difference between adherence levels in PO G

and PO J, although not as large as the one observed for the hypertension drugs. Adherence seems to change significantly over time, much more than what we observe in Figure J.7, especially for PO J, with a similar trend for both plans.

Figure J.7. Average Adherence to Cholesterol Lowering Medication for Plans Offering Part D Intervention, 2016–2018



NOTE: ***, **, and * represent statistically significant differences at the 1-, 5-, and 10-percent levels, respectively, from the 2016 estimate.

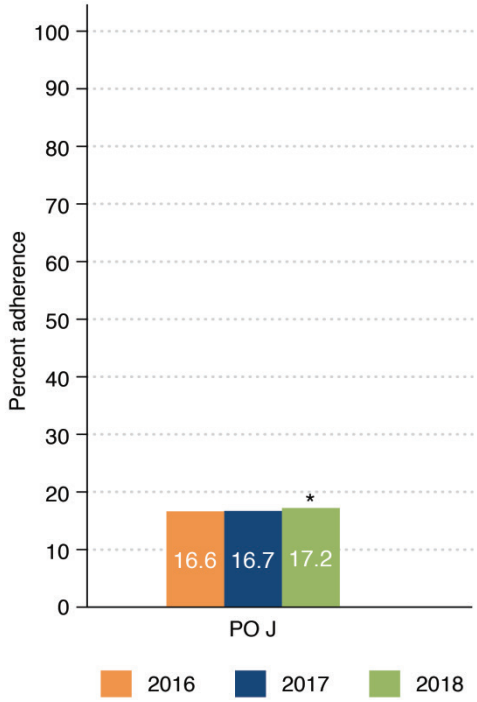
Finally, in Figure J.8, we report the measured average adherence for diabetes medication, for the only plan for which data are currently available (PO J). We report this figure mostly for completeness, because it is hard to interpret. It shows an unusually low level of adherence that changes minimally over time, which points to the need to investigate the data source more closely.

Several conclusions can be drawn from the discussion above. The most apparent one is that there is a fairly large amount of unexplained variation in adherence levels across the POs offering Part D interventions. For hypertension medication adherence ranges from 0.9, for PO D, to 0.5, for PO J. Although this falls in the range of observed adherence, it is not clear whether measuring error contributes in any way to this result and which beneficiaries' characteristics contribute to its explanation.

Figures J.6–J.8 have also shown that overall there has been some statistically significant increases in adherence for hypertension and cholesterol in the period from 2016 to 2018. In

particular, we have observed significant changes in adherence to hypertension medication in PO D, which has actually targeted hypertension, and significant changes in adherence to cholesterol medication in PO G, which has targeted cardiovascular conditions such as CHF.

Figure J.8. Average Adherence to Diabetes Medication for Plans Offering Part D Intervention, 2016–2018



NOTE: ***, **, and * represent statistically significant differences at the 1-, 5-, and 10-percent levels, respectively, from the 2016 estimate.

Appendix K. Effect of Value-Based Participation Requirements on Utilization

In Appendix J, we considered the issue of whether beneficiaries in VBID plans with a Part D intervention experienced different outcomes from their counterparts in VBID with no Part D intervention and from matched comparators who are not in VBID plans. The analysis was motivated by the fact that there is variation in the type of VBID interventions that POs have implemented. Part D is not the only dimension along which POs implementations differ, though; seven POs imposed requirements to receive VBID benefits, typically requiring beneficiaries to participate in CM/DM. In this appendix, we perform a subgroup analysis and investigate the effects of participation requirements on the following five outcome variables: number of 30-day refills, any inpatient stay, any ED admission, number of primary visits, and number of specialty visits. We refer to the plans with participation requirements as "CM/DM plans" and, as in Appendix J, we perform the following two types of analyses:

- **Analysis 1:** We compare individuals in CM/DM plans with matched comparators in VBID plans without participation requirements.
- **Analysis 2:** We compare individuals in CM/DM plans with matched comparators who are not in VBID plans.

For both analyses, we use the same difference-in-differences methodology and control variables described in Appendixes G and J.

Analysis 1: Individuals in Case Management/Disease Management Plans and Comparators in VBID Plans Without Participation Requirements

The composition of the VBID plans with and without participation requirements is quite different, making it difficult to perform fair comparisons. As shown in Table K.1, individuals in CM/DM plans are older, more likely to be disabled, more likely to receive low-income support, and generally in worse health than those in VBID plans with no participation requirements.

To account for the differences in population composition and to make the results less sensitive to the functional form specification, prior to running the regressions we performed 1-to-1 CEM, using the same methodology described in the Analysis 1 section of Appendix J. The effect of matching on the distribution of key confounders is reported in Table K.2. Notice that multivariate imbalance measure after matching is zero, reflecting the fact that we have discretized the continuous variables and were able to obtain perfect matches within all the cells of the joint probability distribution. This is clearly obtained at the expense of a smaller sample, which is reduced to approximately 7,000 beneficiaries in each of the four years of the panel.

Table K.1. Descriptive Statistics for VBID Plans with and Without Participation Requirements, 2016

Covariate	VBID with No CM/DM	VBID with CM/DM
Age	74.8	76.9
Low-income support (percent)	9	16
HCC risk score	1.24	1.88
RxHCC risk score	1.03	1.35
Disabled (percent)	12	19
Died before 2018 (percent)	4	7

Table K.2. CEM Quality Statistics for Matching Beneficiaries in VBID Plans with and Without Participation Requirements, 2016

	Prior to Matching	After Matching
Female (percent)	0.015	0
Age	0.133	0
Low-income support (percent)	0.068	0
Dual eligible (percent)	0.042	0
Disabled (percent)	0.073	0
Offers Part D	0.032	0
OOP maximum ($)	0.467	0
Premium ($)	0.27	0
HCC Risk score	0.356	0
Multivariate imbalance measure	0.817	0
Size of treatment group	42,355	7,052
Size of control group	30,755	7,052

In Table K.3, we report the results of the difference-in-differences regression. In the second column of the table ("Effect"), we show the association between the outcome and the presence of participation requirements, measured on the original scale. To facilitate the analysis, we report in Table K.4 the mean values for the comparison groups, so we can translate the estimate in percentage changes. The most notable findings in Table K.3 are that beneficiaries in CM/DM POs experienced statistically significant increases in inpatient stays and primary care visits, and a statistically significant decrease in 30-day prescription fills, relative to comparators in non-CM/DM VBID POs. The effect on primary care visits is quite large and corresponds to a 7.8-percent increase over the 2016 value for the comparison group. The size of the effect on prescription drugs use is much smaller, corresponding to a 2.6-percent decrease.

Table K.3. Regression Result, Change in Utilization Measures, Actual Versus Expected in VBID If VBID Had No Participation Requirements

Outcome	Effect	SE	z	P-Value	Lower Bound	Upper Bound
Any inpatient stay	0.023	0.007	3.196	0.001	0.009	0.038
Any ED admission	0.013	0.008	1.512	0.131	−0.004	0.029
Number of specialty visits	−0.047	0.114	−0.415	0.678	−0.272	0.177
Number of primary care visits	0.385	0.062	6.166	0.000	0.263	0.508
Number of 30-day fills	−1.123	0.271	−4.141	0.000	−1.654	−0.591

NOTE: Estimates based on difference-in-differences regressions comparing beneficiaries in POs with CM/DM interventions to matched comparators in VBID POs without CM/DM interventions. Sample size equals 50,500 beneficiary years.

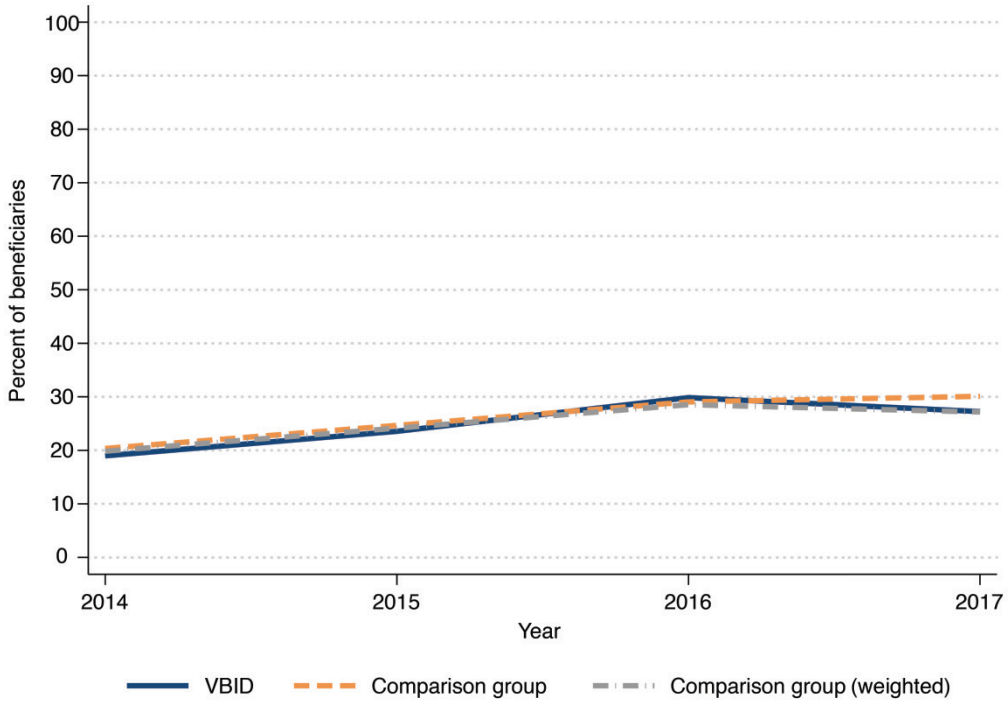
Table K.4. Mean Value of Utilization Measures for Beneficiaries in VBID Plans Without Participation Requirements, 2016

Outcome	Mean Value in 2016
Any inpatient stay	0.30
Any ED admission	0.44
Number of specialty visits	9.4
Number of primary care visits	4.9
Number of 30-day fills	43.6

The effects on PCP visits and prescription drugs are in the expected direction and consistent with the intent of the intervention to encourage beneficiaries to participate in CM/DM. Because POs with CM/DM interventions specifically mentioned medication management as a component of their interventions, it is possible that this is a partial explanation for the reduction in drug utilization. For example, better medication management may reduce polypharmacy and lead to a reduction in unnecessary or contraindicated scripts.

Table K.3 also shows an unexpected finding: CM/DM plans had a statistically significant and positive effect on probability of hospital admission, with an effect size that translates to a 7.7-percent increase over the comparison group of VBID beneficiaries not in CM/DM plans. The reason for this result can be seen in Figure K.1, where we plot the time trends of the probability of any hospital admission for CM/DM plans and comparators. Figure K.1 clearly shows that the probability of any hospital admission increased in the period postintervention in the CM/DM plans, while it decreased in the comparison group. What drives this behavior is not clear. Although one would expect CM/DM to be associated with a reduction in hospitalizations, in the long term, it is also possible that some of the additional hospitalizations are for unavoidable problems which are simply acted on sooner because of more frequent and better primary CM. The current data set does not extend beyond year 2017; therefore, we are unable to test a longer-term reduction in inpatient stays. We are also unable to test the type of additional hospitalizations, which would shed light on the origin of this effect.

Figure K.1. Probability of Any Hospital Admission in CM/DM Plans and Matched VBID Comparators, 2014–2017



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.00 before weighting and 0.50 after weighting.

Table K.3 also shows two results that are statistically insignificant, although they are not far from being significant. For the number of specialist visits, the sign of the association is negative and the size of the effect is small. The sign is consistent with a scenario in which beneficiaries are better managed through primary care and therefore need fewer specialty visits. The effect on the probability of any ED admission is larger and translates into a 3-percent increase in any ED admission. This sign is consistent with the increased frequency of hospitalizations, although we cannot test whether the additional ED visits were linked to an admission or not.

To summarize, this analysis has shown that participation requirements within VBID are associated with an increased number of primary care visits and reduced number of 30-day prescription fills. The analysis has also shown an unexplained increase in frequency of hospital admissions, although we do not know whether this effect is sustained over time. The results are robust with respect to changes to the specifics of the matching process. As a sensitivity analysis we ran the same type of regressions on the entire, unmatched data set. The estimates obtained had the same signs, were more statistically significant, most likely due to the increased size of the data set, and tended to be smaller in size, most likely because of noise introduced by unmatchable records.

Analysis 2: Individuals in Case Management/Disease Management and Matched Comparators Not in Value-Based Insurance Design

In this analysis, we compare beneficiaries in CM/DM plans with matched comparators who are not in VBID plans. We used the same difference-in-differences regression methodology used in Appendix J and described in detail in Appendix G. The sample size for this study is large (276,000 over four years), and the parallel trend assumptions failed. However, inspection of the trends after entropy balancing reweighting shows good alignments, and therefore we proceeded with the specification that assumes that the parallel trends hypothesis holds. The results of the analysis are shown in Table K.5, where the results are reported on the scale of the original variable. To facilitate the discussion, we show in Table K.6 the mean value of these variables for the comparison group in year 2016, so that we can easily convert the effects of Table K.5 into percentage changes.

Table K.5. Regression Result, Change in Utilization Measures, Actual Versus Expected If VBID with Participation Requirements Had Not Been Available

Outcome	Effect	SE	z	p-Value	Lower Bound	Upper Bound
Any inpatient stay	0.010	0.003	3.055	0.002	0.004	0.016
Any ED admission	0.008	0.004	2.095	0.036	0.000	0.015
Number of specialty visits	0.126	0.055	2.311	0.021	0.019	0.233
Number of primary care visits	0.271	0.026	10.492	0.000	0.221	0.322
Number of 30-day fills	0.138	0.124	1.114	0.265	−0.105	0.382

NOTE: Estimates based on difference-in-differences regressions comparing beneficiaries in POs with CM/DM interventions to matched comparators in non-VBID POs. Sample size equals 284,440 beneficiary years.

Except for the number of 30-day fills, all associations in Table K.5 are statistically significant at conventional levels. The most significant finding in the table is a strong positive association of participation requirements with primary care visits: Beneficiaries in CM/DM plans visit a primary care doctor 5.5 percent more often than matched comparators who are not in VBID plans. This increase in primary care visits is paralleled by a small increase in specialty visits of only 1.3 percent and a 2.9-percent increase in the probability of any hospital admission. The association with specialty visits has the opposite sign from what we found in the previous analysis, but this is not necessarily inconsistent with it because the comparison groups are quite different. The increase in hospital visits is accompanied by an increase of 1.6 percent in ED admission.

Unlike in Analysis 1, the association between CM/DM and number of 30-day fills is positive. However, it is not only not significant but also so small (corresponding to an increase of 0.3 percent) that it does not provide any evidence of change in drug utilization.

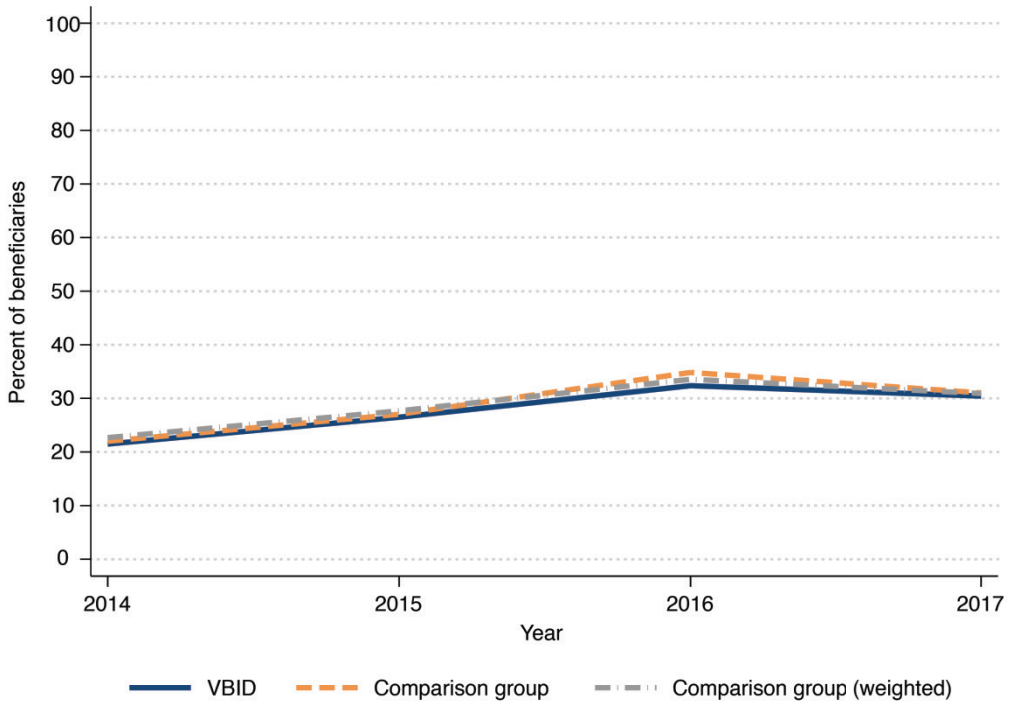
Table K.6. Mean Value of Utilization Measures for Beneficiaries Not in VBID plans, 2016

Outcome	Mean Value in 2016
Any inpatient stay	0.35
Any ED admission	0.51
Number of specialty visits	9.41
Number of primary care visits	4.91
Number of 30-day fills	43.67

Because the result for inpatient stays is not intuitive, it is useful to look at the original data. Therefore, we show in Figure K.2, the average probability of any inpatient stay for beneficiaries in CM/DM plans and matched comparators not in VBID.

Figure K.2 shows that the probability of any inpatient admission decreases after the intervention for the CM/DM beneficiaries, but it decreases more for the comparison group, leading to a relative increase. A similar pattern is observed for ED admissions (not shown). Notice that this is a very different qualitative pattern from the one observed in Figure K.1, where we performed comparisons within VBID plans. The data in Figure K.1 refer to the subset of the beneficiaries in CM/DM plans that could be matched to VBID beneficiaries in plans without

Figure K.2. Average Probability of Any Inpatient Stay for Beneficiaries in CM/DM Plans and Matched Comparators Not in VBID, 2014–2017



NOTE: Parallel trends were assessed using the approach described in Appendix D. The p-value for the test of parallel trends was 0.00 before weighting and 0.00 after weighting.

CM/DM. The contrasts between the two figures points to the fact that there is substantial heterogeneity in time trends within the VBID plans and, depending on which subset of the population one focuses on, different patterns may be observed, even if they lead to the same conclusion. In this specific case, a more detailed analysis has shown that the difference in the time trends of Figures K.1 and K.2 is well explained by the beneficiaries' risk score.

To summarize, we find evidence that participation in CM/DM plans is statistically significantly associated with an increase in utilization of both primary care and specialty services. We have also found evidence of an association with increased frequency of hospitalization, although additional data on the type of hospital stays are needed to gain a better understanding of this effect.

Appendix L. Beneficiary Responsiveness to PCP Copayments

Previous studies have examined whether consumers are responsive to the out-of-pocket price of medical care. For consumers with health insurance, the out-of-pocket price is the cost sharing paid at the time the patient receives the service (e.g., coinsurance, copay, and deductible payments). Price sensitivity is commonly measured using the elasticity, which is the percent change in service utilization stemming from a percent change in price. Elasticities that are greater than one (in absolute value) indicate relatively high price responsiveness (*elastic* demand), while elasticities below one (in absolute value) indicate lower price responsiveness (*inelastic* demand). A large body of work focused primarily on the population under age 65 has found that demand for medical care is relatively inelastic, with elasticities typically ranging from –0.10 to –0.44 (Manning et al., 1987; Scoggins and Weinberg, 2016; Ellis et al., 2017). These studies did not focus on older people who are chronically ill, a group that may be less sensitive to the price of medical care.

We conducted an exploratory analysis using the data generated for the Medicare Advantage Value-Based Insurance Design Model test (MA VBID) to analyze whether chronically ill Medicare beneficiaries are responsive to changes in the copayment level for primary care provider visits (PCP visits).

Methods

Briefly, we exploited variation in copayments introduced as part of the MA VBID model test, along with copayment changes in the comparison group, to estimate elasticities. Our sample consisted of VBID-eligible beneficiaries in the seven VBID-participating POs that did not offer cost-sharing rebates, and their matched comparators.[1] Among the seven VBID POs in our sample, three reduced cost sharing for primary care visits. We did not test the impact of VBID on the utilization of PCP visits in this analysis as those results have been reported in Chapter 6. We used the PCP outcome developed for the utilization analyses in Chapter 6, coded in two ways: 1) whether the beneficiary had at least one PCP visit in the year; and 2) the number of PCP visits the beneficiary had in a year. Table 1 describes the average PCP utilization across the sample and the average copay. We estimate elasticities using variation in copayments across plans and within plans over time.

Our sample consisted of beneficiaries in treatment PBPs who were eligible for their POs' VBID intervention as of January 1, 2017, and their matched comparators. We pooled VBID and

[1] We excluded two VBID-participating POs that offered rebates contingent on the use of specific services because it is unclear the extent to which these rebates—which are paid quarterly or annually—are perceived by beneficiaries as reductions in the price of the service.

Table L.1. Summary Statistics of PCP Visits and Copays for Sample

	2014	2015	2016	2017
N	55,162	80,734	104,756	95,610
PCP copay ($)	14.8 (0.023)	14.7 (0.016)	13 (0.016)	11.4 (0.021)
PCP visits	4.3 (0.018)	4.5 (0.015)	4.6 (0.014)	4.5 (0.015)
% with any PCP visit	83.1 (0.16)	83.7 (0.13)	83.9 (0.114)	82.2 (0.124)
Average within-beneficiary change in PCP copay from previous year ($)	N/A	–0.5 (0.011)	–1.5 (0.006)	–1.6 (0.008)

comparison beneficiaries together and ran one regression model for each outcome. Our matching process, detailed in Appendix D, ensured that treatment and comparison beneficiaries are highly comparable, with the key difference being that VBID beneficiaries faced changes to plan design starting in 2017, sometimes including reduced primary care copayments. We used data on utilization spanning from 2014 through 2017, although we observed utilization only for years when the beneficiary was enrolled in the treatment (or comparison) PBP. Because we used the VBID population, all beneficiaries included in our sample had at least one chronic condition (chronic obstructive pulmonary disorder, congestive heart failure, diabetes, or hypertension) in 2016.

The regression model assessed the changes in PCP use per-beneficiary per-year, with the key independent variable being the size of the copayment for PCP visits per plan per year (pcp_copay_{tp}), according to the following specification:

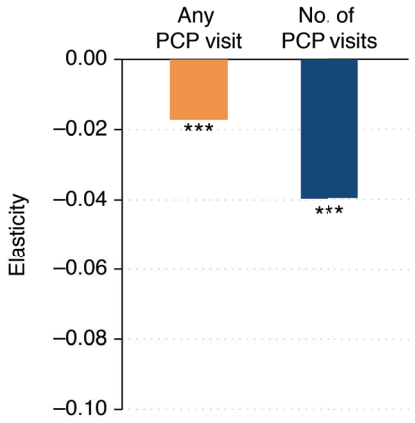
$$u_{itp} = \alpha + \beta spec\_copay_{tp} + \gamma\, pcp\_copay_{tp} + \delta\, VBID_i + \mu\, Year_t + \theta Z_{it} + \mu X_p + \varepsilon_{it}$$

We also controlled for the specialist visit copayment ($spec\_copay_{tp}$) since the decision to use PCP visits might also depend on the cost of specialist visits. We controlled for a set of beneficiary characteristics (Z_{it} − age, gender, race/ethnicity, Medicare-Medicaid dual-eligibility status, Part D low-income subsidy status, an indicator for whether the beneficiary was disabled, HCC risk score, and an indicator for whether the beneficiary was in a VBID plan – $VBID_i$) and plan characteristics (X_p −Part C + D premium, out-of-pocket maximum, parent organization and an indicator for whether the beneficiary was in a VBID plan that included care management requirements as part of their VBID intervention).

Results

As expected, we found that as patient cost sharing falls, demand for PCP visits increases slightly (Figure L.1). The elasticities for PCP visits (for both versions of the outcome) are statistically significant and in the expected direction (negative) but very small. For the binary indicator for at least one PCP visit, the elasticity is –0.02, which indicates that a 10 percent decrease in PCP copay is associated with a 0.2 percent increase (0.2 = –0.02 × –10) in the

Figure L.1. Elasticity Estimates for any PCP Visit and Number of PCP Visits Per-Beneficiary Per-Year, 2014–2017



probability of any PCP visit. Similarly, the elasticity for the number of PCP visits is 0.04. A 10 percent decrease in the PCP copay is associated with a 0.4 percent increase in the number of PCP visits per-beneficiary per-year. These estimates suggest that the chronically ill beneficiaries in our sample are relatively insensitive to the price of PCP visits.

Discussion

Our elasticities of demand for PCP visits are much lower than several previous estimates of the elasticity of demand for medical care. For example, using a randomized control trial, the RAND HIE found that the elasticity for medical care was approximately –0.2. However, the RAND HIE did not include adults over the age of 65, and so may not reflect elasticities for the Medicare population. Other studies that have considered general medical care or overall physician visits have found results somewhat closer to the RAND HIE. For example, Scoggins and Weinberg (2017) found an elasticity for medical services of between –0.12 and –0.31 for a population of public employees in Washington State, depending on the type of model they used. Chandra, Gruber, and McKnight (2010) examined responsiveness to changing copayments among enrollees in the California Public Employees' Retiree System (CalPERS), nearly all of whom were over the age of 65 and found the elasticity for physician office (both PCP and specialist) visits to be –0.10 (using a change in copayment from $0 to $10 for outpatient services. Another study of Medicare Advantage enrollees found that a near doubling of primary care copays (from an average of $7.38 to $14.38) led to a 20 percent reduction in all outpatient visits, suggesting an elasticity similar to the –0.2 estimated in the HIE (Trivedi et al., 2010). However, the treatment group in the Trivedi study was identified by selecting PBPs that explicitly raised copayments for ambulatory care, and the elasticity estimates focused on all outpatient use (not just primary care).

Our estimated elasticity may be lower than previously-reported elasticities of demand for medical care because we focused specifically on primary care visits. Ellis et al. (2017) used a

database of commercial claims from 2008–2014 to explore elasticities by type of service and found substantial variation, with lower elasticities for primary care services. Specifically, although the overall elasticity for medical services in Ellis et al.'s study was –0.44, the elasticity for nonspecialist visits was –0.25 and the elasticity for preventive care services was –0.02. Consumers may be less sensitive to the price of primary care because they are more likely to have established relationships with primary care providers than with other types of providers.

In addition, elasticity estimates may depend on the size of the changes in price observed in the data. The subset of beneficiaries in our sample exposed to changes in primary care copayments due to VBID faced relatively steep declines in cost sharing, with copays falling from around $10 to zero. However, few beneficiaries in our sample were exposed to such substantial changes. As shown in Table 1, the average changes in PCP copays from one year to the next are small (less than $2). In a more detailed analysis (not reported in the table), we found that the distribution of copay changes has a large mode at $0, with about 50 percent of the beneficiaries experiencing no change in copay. In the remaining sample the average change in copay was $3, and only 28 percent of those with any change in copay experienced a yearly change of larger than 50 percent. In contrast, other studies have focused on larger copayment changes. For example, Scoggins and Weinberg (2017) exploited variation in coinsurance rates ranges from 10 to 25 percent, and the RAND HIE include cost-sharing ranging from $0 to 95 percent coinsurance. It is possible that consumers are less sensitive to small price changes, which might partly explain our smaller elasticity estimates.

Further, our elasticity estimates may be lower than prior estimates because our population of chronically ill, older adults may be less sensitive to medical care prices than the general population of younger adults. Because of their age and health status, our population may view medical care as more of a necessity than other groups. Our sample may also have differed from the general Medicare population in ways beyond health status, because the PBPs that POs entered into the model test were systematically different from the overall population of PBPs. For example, our previous evaluation report found that participating PBPs were located in higher-income communities and had higher out-of-pocket maximums than nonparticipating PBPs (Eibner et al., 2018).

Finally, we found that the price elasticity of demand for any PCP visit (–0.02) was only about half as large as the price elasticity of demand for the total number of visits (–0.04). This finding suggests that chronically ill beneficiaries are more likely to increase their total number of visits in response to a price reduction than to move from not having any primary care visits to newly seeing their PCP in a given year. This pattern makes sense from a conceptual standpoint, as the vast majority of beneficiaries in our sample—more than 80 percent—have at least one PCP visit each year. A minimum of one visit per year may be necessary for this population, for example because most beneficiaries have at least one prescription, which would generally require a primary care visit for renewal.

References

Abadie, A., "Semiparametric Difference-in-Differences Estimators," *Review of Economic Studies*, Vol. 72, No. 1, 2005, pp. 1–19.

Agency for Healthcare Research and Quality (AHRQ), "Prevention Quality Indicators Overview," webpage, undated. As of November 11, 2019: https://www.qualityindicators.ahrq.gov/modules/pqi_resources.aspx

Austin, P. C., "Balance Diagnostics for Comparing the Distribution of Baseline Covariates Between Treatment Groups in Propensity-Score Matched Samples," *Statistics in Medicine*, Vol. 28, No. 25, 2009, pp. 3083–3107.

Austin, P. C., and E. A. Stuart, "Moving Towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies," *Statistics in Medicine*, Vol. 34, No. 28, 2015, pp. 3661–3679.

Buntin, M. B., and A. M. Zaslavsky, "Too Much Ado About Two-Part Models and Transformation? Comparing Methods of Modeling Medicare Expenditures." *Journal of Health Economics*, Vol. 23, No. 3, 2004, pp. 525–542.

Butler-Kisber, L., *Qualitative Inquiry: Thematic*, *Narrative and Arts-Informed Perspectives*, Los Angeles: Sage Publications, 2010.

Caliendo, M., and S. Kopeinig, "Some Practical Guidance for the Implementation of Propensity Score Matching," *Journal of Economic Surveys*, Vol. 22, No. 1, 2008, pp. 31–72.

Center for Medicare & Medicaid Innovation (CMMI), "Benefits Data," webpage, 2018. As of November 11, 2019: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/MCRAdvPartDEnrolData/Benefits-Data.html

———, "Medicare Advantage Value-Based Insurance Design Model CY 2018 Communications Guidelines," 2019b. As of November 20, 2019: https://innovation.cms.gov/Files/x/vbid-commsguidelines-yr2.pdf

———, "Medicare Advantage Value-Based Insurance Design Model CY 2019 Communications Guidelines," 2019c. As of November 11, 2019: https://innovation.cms.gov/Files/x/vbid-commsguidelines-yr3.pdf

Centers for Medicare & Medicaid Services (CMS), "Measure Information Form: 2014 Measure Information About the Acute and Chronic Ambulatory Care-Sensitive Condition Composite Measures, Calculated for the Value-Based Payment Modifier Program," 2015. As of April 8, 2020: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeedbackProgram/Downloads/2014-ACSC-MIF.pdf

———, "Medicare Advantage (MA) ValueBased Insurance Design (VBID) Model - Evaluation of Model Year 1 (2017)," 2017. As of February 4, 2020: https://innovation.cms.gov/Files/reports/vbid-yr1-evalrpt-fg.pdf

———, "Encounter Data Submission and Processing Guide: Medicare Advantage Program, Version 1.0," May 2018. As of April 8, 2020: https://www.csscoperations.com/internet/cssc4.nsf/files/ED_Submission_Processing_Guide.pdf/$FIle/ED_Submission_Processing_Guide.pdf

———, "Medicare Advantage Value-Based Insurance Design Model," webpage, 2019a. As of January 27, 2020: https://innovation.cms.gov/initiatives/VBID/

———, "Medicare Data for the Geographic Variation Public Use File: A Methodological Overview - February 2020 Update," 2019b. As of February 4, 2020: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/Downloads/Geo_Var_PUF_Methods_Paper.pdf

Chandra, A., J. Gruber, and R. McKnight, "Patient Cost-Sharing and Hospitalization Offsets in the Elderly," *American Economic Review*, Vol. 100, No. 1, 2010, pp. 193–213.

Chernew, M. E., M. R. Shah, A. Wegh, S. N. Rosenberg, I. A. Juster, A. B. Rosen, M.C. Sokol, K. Yu-Isenberg, and A. M. Fendrick, "Impact of Decreasing Copayments on Medication Adherence Within a Disease Management Environment," *Health Affairs (Project Hope)*, Vol. 27, 2008, pp. 103–112.

Choudhry, N. K., J. Avorn, R. J. Glynn, E. M. Antman, S. Schneeweiss, M. Toscano, L. Reisman, J. Fernandes, C. Spettell, J. L. Lee, R. Levin, T. Brennan, and W. H. Shrank, "Full Coverage for Preventive Medications after Myocardial Infarction," *New England Journal of Medicine*, Vol. 365, No. 22, 2011, pp. 2088–2097.

Choudhry, N. K., M. A. Fischer, J. Avorn, S. Schneeweiss, D. H. Solomon, C. Berman, S. Jan, J. Liu, J. Lii, M. A. Brookhart, J .J. Mahoney, and W. H. Shrank, "At Pitney Bowes, Value-Based Insurance Design Cut Copayments and Increased Drug Adherence," *Health Affairs (Project Hope)*, Vol. 29, 2010, pp. 1995–2001.

Choudhry, N. K., M. A. Fischer, B. F. Smith, G. Brill, C. Girdish, O. S. Matlin, T. A. Brennan, J. Avorn, and W. H. Shrank, "Five Features of Value-Based Insurance Design Plans Were Associated with Higher Rates of Medication Adherence," *Health Affairs*, Vol. 33, 2014, pp. 493–501.

CMS.gov, "Taxonomy," , February 27, 2018. As of November 14, 2019: https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/Medicare ProviderSupEnroll/Taxonomy.html

Congrees.Gov. "H.R. 1892 - Bipartisan Budget Act of 2018," 115th Cong. (2017–2018). As of November 14, 2019: https://www.congress.gov/bill/115th-congress/house-bill/1892/text

Creighton, S., R. Duddy-Tenbrunsel, and J. Michel, "The Promise and Pitfalls Of Medicare Advantage Encounter Data," Health Affairs Blog, February 25, 2019. As of May 28, 2020: https://www.healthaffairs.org/do/10.1377/hblog20190221.696651/full/

Eibner, C., D. Khodyakov, E. A. Taylor, C. Buttorff, C. Armstrong, M. Booth, . . . B. Eshete-Roesler, *First Annual Evaluation Report of the Medicare Advantage Value-Based Insurance Design Model Test*, Santa Monica, Calif.: RAND Corporation, 2018. As of November 20, 2019: https://innovation.cms.gov/Files/reports/vbid-yr1-evalrpt.pdf

Ellis, R. P., B. Martins, W. Zhu, "Health Care Demand Elasticities by Type of Service," *Journal of Health Economics*, Vol. 55, 2017, pp. 232–243.

Gibson, T. B., S. Wang, E. Kelly, C. Brown, C. Turner, F. Frech-Tamas, . . . E. Mauceri, "A Value-Based Insurance Design Program at a Large Company Boosted Medication Adherence for Employees with Chronic Illnesses," *Health Affairs*, Vol. 30, No. 1, 2011, pp. 109–117.

Haas, A., M. N. Elliott, J. W. Dembosky, J. L. Adams, S. M. Wilson-Frederick, J. S. Mallett, . . . A. M. Haviland, "Imputation of Race/Ethnicity to Enable Measurement of HEDIS Performance by Race/Ethnicity," *Health Services Research*, Vol. 54, No. 1, 2019, pp. 13–23.

Hainmueller, J. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies," *Political Analysis*, Vol. 20, No. 1, 2012, pp. 25–46.

Imai, K., and M. Ratkovic, "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Vol. 76, No. 1, 2014, pp. 243–263.

Kling, J. R., J. B. Liebman, and L. F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, Vol. 75, No. 1, 2007, pp. 83–119.

Lunceford, J. K., and M. Davidian, "Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects: A Comparative Study," *Statistics in Medicine*, Vol. 23, No. 19, 2004, pp. 2937–2960.

Maciejewski, M. L., D. Wansink, J. H. Lindquist, J. C. Parker, and J. F. Farley, "Value-Based Insurance Design Program in North Carolina Increased Medication Adherence but Was Not Cost Neutral," *Health Affairs*, Vol. 33, No. 2, 2014, pp. 300–308.

Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, A. Leibowitz, and M.S. Marquis, "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *American Economic Review*, Vol. 77, No. 3, 1987, pp. 251–277.

MAXQDA, software for qualitative data analysis (version 12), Berlin: VERBI GmbH, 2015.

McHugh, M. L., "Interrater Reliability: The Kappa Statistic," *Biochemia Medica*, Vol. 22, No. 3, 2012, pp. 276–282.

Medicare Payment Advisory Commission, *Report to the Congress: Medicare and the Health Care Delivery System*, Washington, D.C., 2015. As of November 11, 2019: http://www.medpac.gov/docs/default-source/reports/june-2015-report-to-the-congress-medicare-and-the-health-care-delivery-system.pdf?sfvrsn=0

———, "Chapter 7: Ensuring the Accuracy and Completeness of Medicare Advantage Encounter Data," in *Report to the Congress: Medicare and the Health Care Delivery System*, Washington, D.C., 2019. As of April 8, 2020: http://www.medpac.gov/docs/default-source/reports/jun19_ch7_medpac_reporttocongress_sec.pdf?sfvrsn=0

Mulcahy, A. W., M. E. Sorbero, A. Mahmud, A. Wilks, J. Gildner, A. Hornsby, and A. Pignotti, *Measuring Health Care Utilization in Medicare Advantage Encounter Data: Methods*, *Estimates*, *and Considerations for Research*, Santa Monica, Calif.: RAND Corporation, RR-2681-CMS, 2019. As of November 14, 2019: https://www.rand.org/pubs/research_reports/RR2681.html

Normand, S.-L. T., M. B. Landrum, E. Guadagnoli, J. Z. Ayanian, T. J. Ryan, P. D. Cleary, and B. J. McNeil, "Validating Recommendations for Coronary Angiography Following Acute Myocardial Infarction in the Elderly: A Matched Analysis Using Propensity Scores," *Journal of Clinical Epidemiology*, Vol. 54, No. 4, 2001, pp. 387–398.

Peaslee, A., M. Wickizer, J. Olson, and R. Topp, "Impact of a Combined Value-Based Insurance Design and Medication Therapy Management Program on Diabetes Medication Adherence," *Journal of Managed Care & Specialty Pharmacy*, Vol. 22, No. 11, 2016, pp. 1303–1309.

Reid, R., C. Damberg, and M. W. Friedberg, "Primary Care Spending in the Fee-for-Service Medicare Population," *JAMA Internal Medicine*, Vol. 179, No. 7, 2019, pp. 977–980.

Ridgeway, G., D. McCaffrey, A. Morral, L. Burgette, and B. A. Griffin, *Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial for the R TWANG Package*, Santa Monica, Calif.: RAND Corporation, 2017. As of November 20, 2019: https://www.rand.org/pubs/tools/TL136z1.html

Scoggins, J. F., and D. A. Weinberg, "Healthcare Coinsurance Elasticity Coefficient Estimation Using Monthly Cross-Sectional, Time-Series Claims Data," *Health Economics*, Vol. 26, No. 6, 2017, pp. 795–801.

Sekhon, J. S., "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R," *Journal of Statistical Software*, Vol. 42, No. 7, 2011, pp. 1–52.

Selim, A. J., "Updated U.S. Population Standard for the Veterans RAND 12-Item Health Survey (VR-12)," *Quality of Life Research*, Vol. 18, No. 1, 2009, pp. 43–52.

Stuart, E. A. "Matching Methods for Causal Inference: A Review and A Look Forward." *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, Vol. 25, No. 1, 2010, pp. 1–21.

Stuart, E. A., H. A. Huskamp, K. Duckworth, J. Simmons, Z. Song, M. E. Chernew, and C. L. Barry, "Using Propensity Scores in Difference-in-Differences Models to Estimate the Effects of a Policy Change," *Health Services and Outcomes Research Methodology*, Vol. 14, No. 4, 2014, pp. 166–182.

Trivedi, A.N., H. Moloo, and V. Mor, "Increased Ambulatory Care Copayments and Hospitalizations Among the Elderly," *New England Journal of Medicine*, Vol. 362, No. 4, 2010, pp. 320–328.

United States Department of Health and Human Services, Office of the Inspector General (HHS, OIG), *Medicare Advantage Encounter Data Show Promise for Program Oversight, but Improvements Are Needed*, Report (OEI-03-15-00060), 2018. As of April 8, 2020: https://oig.hhs.gov/oei/reports/oei-03-15-00060.asp

United States Government Accountability Office (GAO), *Medicare Advantage: Limited Progress Made to Validate Encounter Data Used to Ensure Proper Payments*, GAO-17-223, January 2017. As of April 8, 2020: https://www.gao.gov/assets/690/682145.pdf

Yeung, K., A. Basu, R. N. Hansen, J. B. Watkins, and S. D. Sullivan, "Impact of a Value-Based Formulary on Medication Utilization, Health Services Utilization, and Expenditures," *Medical Care*, Vol. 55, No. 2, 2017, pp. 191–198.

Washington Publishing Company, "Health Care Provider Taxonomy Code Set • ASC X12 External Code Source 682," *Washington Publishing Company*, July 1, 2019. As of November 14, 2019: http://www.wpc-edi.com/reference/codelists/healthcare/health-care-provider-taxonomy-code-set/

Zeger, S. L., K. Y. Liang, and P. S. Albert, "Models for Longitudinal Data: A Generalized Estimating Equation Approach," *Biometrics*, Vol. 44, No. 4, 1988, pp. 1049–1060.