Project Report

Appendices: Evaluation of the Medicare Advantage Value-Based Insurance Design Model: 2020–2024

Dmitry Khodyakov, Christine Eibner, Erin A. Taylor, Julia Bandini, Marika Booth, Christine Buttorff, Stephanie Dellva, Michael Dworsky, Alice Y. Kim, Julie Lai, Monique Martineau, Nabeel Qureshi, Afshin Rastegar, Max Rubinstein, Daniel Schwam, Shiyuan Zhang

RAND Health

PR-A4406-2
March 2026
Prepared for the Centers for Medicare & Medicaid Services
Center for Medicare and Medicaid Innovation
Under Research, Measurement, Assessment, Design, and Analysis Contract Number 75FCMC19D0093, Order Number 75FCMC20F0001

RAND

About These Appendices

These appendices supplement the evaluation of the Medicare Advantage Value-Based Insurance Design (VBID) Model test, initiated by the Center for Medicare and Medicaid Innovation (Innovation Center), for the years 2020 through 2024. Both the project report and these appendices focus on evaluating the impact of reduced cost sharing for Part C and Part D benefits, VBID-enabled supplemental benefits, and Rewards and Incentives programs and do not evaluate the Hospice Benefit component of the model. Because the analyses are similar year over year, these appendices contain some of the same information presented in the 2025 evaluation report (Eibner et al., 2025).

This research was funded by the Innovation Center under Research, Measurement, Assessment, Design, and Analysis for Contract Number 75FCMC19D0093 and Order Number 75FCMC20F0001, for which Julia Driessen is the contracting officer's representative. It was carried out within the Financing and Organization Program in RAND Health.

RAND Health, a division of RAND, strengthens health care systems and the societal conditions that support health. We provide decisionmakers across government, health care, and communities with objective, actionable evidence to address health challenges from clinical settings to neighborhoods. For more information, see www.rand.org/health, or contact Health@rand.org.

Acknowledgments

We thank Julia Driessen at the Centers for Medicare & Medicaid Services (CMS) for her guidance and support as we developed this project report. We also thank the CMS Office of the Actuary, the CMS VBID Model team, and Acumen for providing critical data. We are also indebted to our reviewers and expert advisers, including Betsy Q. Cliff, Beth Ann Griffin, Paul Koegel, Carrie Farmer, and Peter Hussey. We are grateful to Leslie Canterbury for excellent administrative assistance.

Abbreviations

ATT	average treatment effect on the treated
CI	confidence interval
CMS	Centers for Medicare & Medicaid Services
COVID-19	coronavirus disease 2019
C-SNP	chronic condition special needs plan
DD	difference in differences
DSNP	Dual Eligible Special Needs Plan
EB	entropy balancing
ED	emergency department
ESS	effective sample size
FFS	fee for service
GERBIL	Generalized Efficient Regression Based Imputation with Latent Processes
HCC	hierarchical condition category
ISNP	institutional special needs plan
LIS	low-income subsidy
MA	Medicare Advantage
MAPD	Medicare Advantage Prescription Drug
MSB	mandatory supplemental benefit
NA	not applicable
OOP	out of pocket
PDSS	Part D Senior Savings
PMPM	per member, per month
PO	parent organization
ppt	percentage point
RI	Rewards and Incentives
RxHCC	prescription drug hierarchical condition category
SES	socioeconomic status
SMD	standardized mean difference
SSBCI	Special Supplemental Benefits for the Chronically Ill
UF	uniformity flexibility
VBID	Value-Based Insurance Design

Contents

Figures and Tables

Figures

Tables

Appendix A. Methods for Statistical Analysis

The main project report includes quantitative analyses intended to estimate changes in outcomes associated with participation in one component of the Medicare Advantage (MA) Value-Based Insurance Design (VBID)—VBID General—at three levels of analysis: the plan level, the beneficiary level, and the contract level. As in our 2025 evaluation report (Eibner et al., 2025), all these analyses use an entropy-balanced difference-in-differences (DD) design to identify the effects of VBID General, hereafter referred to as *VBID*. Some minor modifications to these methods that were necessary to accommodate more-recent years of data are discussed in the next section.

Because our basic approach is similar to that in our 2025 evaluation report, this description of our methods draws heavily on the description provided there (Eibner et al., 2025).

Overview of Statistical Methods

As described in the main project report, VBID encompasses a variety of interventions, including Part C and Part D reduced cost sharing, VBID-enabled supplemental benefits, and Rewards and Incentives (RI) programs. Our analyses did not differentiate these VBID interventions. Instead, for each year of the model test, we sought to estimate the average effect of all VBID interventions in effect at the time.

There are a few challenges to successfully estimating the average impact of VBID participation, given the observational nature of the model test:

- First, plans' fidelity of implementation and beneficiaries' uptake of the proposed intervention can vary. For this reason, all analyses, unless otherwise noted, were based on the intention-to-treat principle—that is, plans were analyzed based on their participation in VBID, regardless of how their interventions were implemented or the extent to which beneficiaries used VBID benefits.
- Second, plans were allowed to join and leave the VBID Model on a year-to-year basis, which led to different participation patterns.
- Finally, plans that chose to participate in the VBID Model might differ in both observable and unobservable ways from those that did not.

To address the second and third concerns, this evaluation combines entropy balancing (EB) on observable characteristics with the DD framework established in Callaway and Sant'Anna (2021), which allows DD designs with differing patterns of participation. The Callaway and Sant'Anna approach involves running separate regressions for plans based on when they entered the VBID Model and addresses concerns that staggered adoption of the model could lead to a biased result if associations between VBID and key outcomes varied over time (Goodman-Bacon, 2021; de Chaisemartin and d'Haultfoeuille, 2024). EB serves to bolster the DD design, which accounts for both observable and unobservable characteristics if they are stable or trend similarly between VBID and comparison plans under certain assumptions. Table A.1 shows the data sources we used for this analysis.

Table A.1. Data Sources and Variable Domains Used in Project Report

Domain	Primary Data Sources	Years	Description	Analytic Use
Area-level demographic and socioeconomic characteristics	U.S. Census Bureau (including the American Community Survey), Area Health Resources Files, Integrated Data Repository	2017–2024	County- and plan-level measures of population composition, economic status, and area characteristics	EB
Geographic characteristics	CMS penetration files, Rural–Urban Continuum Codes, MA region files	2013–2024	Indicators of urbanicity, MA penetration	EB, descriptive
Plan benefits, costs, and characteristics	Health Plan Management System, Office of the Actuary Bid Pricing Tool, PBP	2017–2024	Standardized bids, estimated costs, premiums, MA rebates, supplemental benefit offerings, plan type	Outcomes
Contract performance	MA and Part D prescription drug Star Ratings	2017–2025	Contract-level performance scores	Outcomes, EB, descriptive
PO characteristics	CMS, Health Plan Management System	2017–2024	Blue Cross affiliate, for-profit status, enrollment, number of states in which plans are offered	EB, descriptive
Beneficiary health status, utilization, OOP costs, and demographic characteristics	Integrated Data Repository (risk scores, enrollment, claims, Part D event, beneficiary characteristics)	2017–2022	Measures of comorbidity and utilization, demographic characteristics for beneficiary analysis	Outcomes, EB, descriptive

NOTE: CMS = Centers for Medicare & Medicaid Services; OOP = out of pocket; PO = parent organization; PBP = plan benefit package; PDP = Part D prescription drug.

This analytic strategy can be summarized in four distinct stages, which are described in greater detail in subsequent sections:

1. definition of groups of participating plans and the effects of interest
2. identification of nonparticipating plans that are eligible for VBID
3. construction of outcome-specific comparison groups using EB for each of the groups in stage 1 using the comparisons identified in stage 2
4. estimation and summarization of DD models using the comparison groups derived in stage 3.

In the first stage, we grouped plans into participation patterns based on each plan's history of participation in VBID (or, for subgroup analyses, each plan's history of implementing specific types of VBID General interventions).

In the second stage, we identified eligible nonparticipant plans that could serve as members of a comparison group for the VBID plans. Because participation in the VBID Model test is voluntary (rather than randomly assigned), the comparison plans may differ from the VBID plans on both observable and unobservable characteristics that might predict differences in how outcomes of interest will evolve over time.

Therefore, in the third stage, we constructed a set of weights chosen to ensure that the observed covariate distribution of the comparison group (including both baseline characteristics and trends in outcomes prior to VBID implementation) approximately equaled the observed covariate distribution of the VBID-participating group.

Finally, in the fourth stage, we estimated a weighted two-way fixed-effect regression model for each outcome and participation pattern, using the weights constructed in the third stage to make the weighted comparison group as similar as possible to the VBID-participating group. Estimation of a separate model for each outcome and participation pattern is critical to ensure that two-way fixed-effect regression models identify average treatment effects on the treated (ATTs) (discussed below) rather than other quantities that are less relevant for policy evaluation. This approach is necessitated by the fact that VBID plans joined the model at different points in time; this *staggered-adoption* situation requires some care to obtain unbiased estimates if the changes in outcomes associated with VBID (which we call *treatment effects* for simplicity in this appendix) are heterogeneous over time or across plans adopting interventions in different years (Goodman-Bacon, 2021).

Because estimation of separate models by outcome and participation pattern yields a large number of estimates for each outcome, we combined the effects estimated for different participation patterns to estimate the average change in outcomes associated with VBID for each calendar year of the model test (that is, the average effect in 2020, 2021, 2022, 2023, or 2024). Reporting the average change in outcomes associated with VBID in each year of the model test offers a concise way to summarize findings for the large numbers of outcomes and years considered in our evaluation. The policy question answered by each year's estimate—"What was the average effect of all the VBID interventions implemented in that year?"—is likely of interest to the Centers for Medicare & Medicaid Services (CMS) because the answer reflects the overall effect of the model test.

This approach is not without limitations. The calendar-year effects for each year reflect a different group of plans and interventions both because of rapid growth in the number of participating plans and because of changes in the mix of interventions adopted over time. Even plans that participated for multiple years often added, removed, or modified parts of their interventions from year to year. Therefore, the average effects by calendar year that we report should not be interpreted as reflecting time-varying effects of a constant set of interventions. Alternative approaches to aggregating results across the many distinct participation patterns in the model (such as reporting average effects of interventions in year 1, 2, or 3 of implementation) would lead to greater complexity without adequately addressing the issues of heterogeneity in the interventions chosen by plans.

In rest of this section, we describe the steps in our estimation approach in greater detail. We provide a general description of our estimation strategy that applies to all levels of analysis; for simplicity, our methods described here focus on plan-level analyses. Additional detail on variables used in balancing and the extent to which balancing succeeded in making the comparison group observably similar to the VBID group is also presented here. We then present additional detail on differences between the plan-level, beneficiary-level, and contract-level estimation strategies, including details about the variables used in balancing at each level of analysis.

For a more complete listing of control variables used in the analysis, including data sources and definitions, please refer to Eibner et al. (2023b, Appendix D).

Defining Groups of Participating Plans

As in prior reports, we limited our analyses to Medicare Advantage Prescription Drug (MAPD) plans because very few MA-only plans participated and because we expected substantial differences in the design and structure of MAPD and MA-only plans. Although several parent organizations (POs) participated in a prior, Phase I (2017–2019) iteration of the VBID Model, we did not attempt in this analysis to model the effects of participation in Phase I.

Our analysis began with determining each plan's history of participation in VBID over the course of the present model, which began in 2020. Beginning with 2017 (the earliest year of pre-VBID data used in our analyses), a plan might have been observed annually for up to seven years (2017 through 2023), and VBID participation could have begun in 2020 or later. To define each plan's history of participation in VBID, we use the notation a_t to denote a binary indicator that is equal to 1 if a plan participates in VBID in year t and 0 otherwise; $a_t = 0$ for all years prior to the start of the model test (t = 2017, 2018, or 2019). Years in which a plan did not exist are coded as "NA," for "not applicable."

With this notation, the history of a plan's participation in VBID General in 2020 and later years can be encoded as vector (denoted by $\boldsymbol{a}$) containing these participation indicators. For example, the history of participation between 2020 and 2024 is a vector comprising five indicator variables $\boldsymbol{a} = (a_{2020}, a_{2021}, a_{2022}, a_{2023}, a_{2024})$.

As we discuss in this section, we estimated associations between VBID and outcomes by running separate statistical models for each year of the model, with plans grouped based on their participation history to date in each outcome year examined. We imposed two exclusion criteria:

- First, plans that entered VBID in their first year of existence were excluded because it is not possible to distinguish the effects of VBID from other factors affecting outcomes if a plan never existed without VBID. The exclusion of plans entering VBID in their first year of existence reduced the sample size of participating plans by seven plans in 2020, 26 plans in 2021, 112 plans in 2022, 295 plans in 2023, and 460 plans in 2024.
- After excluding plans with no history before VBID implementation, we also excluded participation patterns that had five or fewer plans. This led to the exclusion of an

additional eight plans in 2023 and ten plans in 2024. These patterns all reflect plans that entered the model test in 2020, exited the model test, and then subsequently reentered. No plans were excluded on this basis in 2020, 2021, or 2022.

Table A.2 lists the participation patterns used in our plan-level analyses for each outcome year between 2020 and 2024, as well as participation patterns that were excluded.

Table A.2. Number of Plans Participating in VBID, by Year, Participation History, and Inclusion in or Exclusion from the Plan-Level Analyses

Outcome Year and Inclusion Indicator	Participation History	Number of VBID General Plans	Number Entering VBID General in First Year of Existence	Number After Excluding Plans Entering VBID General in First Year of Existence
2020, included	1	141	7	134
	Total	**141**	**7**	**134**
2021, included	01	258	0	258
	11	96	6	90
	Total	**354**	**6**	**348**
2022, included	001	320	0	320
	011	254	0	254
	101	8	1	7
	111	87	4	83
	NA01	82	0	82
	Total	**751**	**5**	**746**
2023, included	0001	141	0	141
	0011	280	0	280
	0111	235	0	235
	1111	78	3	75
	NANA01	65	0	65
	NA001	38	0	38
	NA011	77	0	77
	Total	**914**	**3**	**911**
2024, included	00001	83	0	83
	00011	133	0	133
	00111	237	0	237
	01111	192	0	192
	11111	74	1	73
	NA0001	10	0	10
	NA0011	36	0	36
	NA0111	65	0	65
	NANA001	21	0	21

Outcome Year and Inclusion Indicator	Participation History	Number of VBID General Plans	Number Entering VBID General in First Year of Existence	Number After Excluding Plans Entering VBID General in First Year of Existence
	NANA011	54	0	54
	NANANA01	27	0	27
	Total	**932**	**1**	**931**
2021, excluded	NA1	20	20	0
	Total	**20**	**20**	**0**
2022, excluded	NA11	19	19	0
	NANA1	88	88	0
	Total	**107**	**107**	**0**
2023, excluded	0101	1	0	1
	1001	4	0	4
	1011	4	1	3
	NANANA1	212	212	0
	NANA11	62	62	0
	NA111	17	17	0
	Total	**300**	**292**	**8**
2024, excluded	01011	1	0	1
	01101	1	0	1
	10011	4	0	4
	10101	3	0	3
	10111	2	1	1
	NA1111	14	14	0
	NANA111	58	58	0
	NANANA11	167	167	0
	NANANANA1	219	219	0
	Total	**469**	**459**	**10**

NOTE: Participation history is a sequence indicating the plan's history of VBID participation (1) or nonparticipation (0) in each year from 2020 through the outcome year indicated in the row header. A 0 in a pattern indicates that the contract had no participating plans, a 1 indicates that the contract had at least one participating plan, and NA indicates that the contract did not exist in that year. For example, a pattern of 0111 indicates that the contract was observed in 2023 (four years into the model test) and had no participating plans in 2020 and at least one participating plan in 2021, 2022, and 2023. "Included" and "excluded" refer to whether this was included in the DD estimation. Patterns with no history of existence before entering VBID are excluded, as are two patterns (1001 and 101) that exited and reentered VBID but had too few participating plans to be appropriate for our estimation methods.

In total, our analysis of plan-level outcomes included 134 VBID participants in 2020, 348 VBID participants in 2021, 746 VBID participants in 2022, 911 VBID participants in 2023, and 931 VBID participants in 2024. We applied analogous restrictions in defining participation patterns for inclusion in the contract-level analysis, but these restrictions on plans' participation history were not necessary for the beneficiary-level analysis. These and other differences between these analyses and the plan-level approach are discussed under "Additional Estimation" later in this appendix.

Defining the Estimand of Interest

To evaluate changes in outcomes associated with VBID implementation, we sought to estimate the ATT. Following the potential-outcome framework proposed by Rubin (2005), among others, this estimand can be understood as the average difference between observed outcomes and potential outcomes (that is, outcomes that would have been observed in a counterfactual scenario in which the plans implementing VBID had not implemented VBID) among the plans that implemented VBID. The ATT can be defined at different points in time, because—as noted above—the group of participating plans changes from year to year and the effects of VBID may vary over time.

For an outcome of interest Y_{it} observed at time *t*, we define the potential outcomes for unit *i* as a function of unit *i*'s history of participation in VBID: $Y_{it}(\boldsymbol{a})$ denotes the potential outcome that would be observed if unit *i* had participation history $\boldsymbol{a}$, and $Y_{it}(0)$ denotes the potential outcome that would be observed if unit *i* had never participated in VBID.

We can then define the ATT for each VBID participation history $\boldsymbol{A} = \boldsymbol{a}$ as

$$ATT(\boldsymbol{a}, t) = E[y_{it}(a) - y_{it}(0) | \boldsymbol{A} = \boldsymbol{a}]. \qquad \text{(Equation A.1)}$$

Under the assumptions of our DD estimation strategy (discussed under "DD Analysis" later in this section), these $ATT(\boldsymbol{a}, t)$ can be estimated using comparisons between a group of participating plans defined by the VBID participation pattern $\boldsymbol{A} = \boldsymbol{a}$ and a group with $\boldsymbol{A} = \boldsymbol{0}$.

Once we have estimated $ATT(\boldsymbol{a}, t)$ for each participation pattern $\boldsymbol{a}$ and year t, we aggregated the estimates to obtain an average change in outcomes for each calendar year of VBID from 2020 through 2024. That is, for each year *t* = 2020, 2021, 2022, 2023, or 2024, we define ATT of VBID in year *t* as

$$ATT(t) = \sum_{a \in \mathcal{A}} w(\boldsymbol{a}, t) \cdot ATT(\boldsymbol{a}, t), \qquad \text{(Equation A.2)}$$

where $\mathcal{A}$ represents all possible participation histories and $w(\boldsymbol{a}, t)$ is a set of weights reflecting the proportion of year *t* VBID participants that belonged to participation pattern $\boldsymbol{a}$. For example, in 2023, 73 of the 931 VBID-participating plans had participated continuously since 2020, so the weight $w(\mathbf{11111}, 2024)$ was defined to be (0.078 = 73/931).

Identifying Nonparticipating Plans Eligible for VBID

Our prior reports documented in detail the plan-level eligibility criteria for the VBID Model, and we briefly summarize the process here. Key eligibility criteria include limiting to specific MA plan types (for example, employer plans are excluded) and—for some years—being of sufficiently high performance on quality ratings or other metrics (for example, having a three-star rating or higher and not under sanction). We also excluded several other plan types from the pool of eligible comparison plans, including end-stage renal disease chronic condition special needs plans (C-SNPs), because their patient populations differed greatly; plans that transitioned from 1876 Cost plans because they were ineligible for the model; and Part B–only plans because they were missing key outcome data. In each calendar year, newly added and discontinued VBID plans were included in descriptive analyses, but they were used in DD analyses only if they had at least one year of pre-VBID data and one year of post-VBID data for that model year.

EB for Outcome-Specific Comparison Groups

As noted above, plans volunteered to participate in VBID, and those that did so differed from eligible nonparticipating plans with respect to many observable characteristics. We sought to construct comparison groups to minimize these differences in an attempt to improve comparability between the groups and to justify the key assumptions of our DD regression models. To do this, we used EB weights. The weights increase comparability on observable characteristics between the VBID-participating and eligible nonparticipating plans by weighting the nonparticipating plans to be more similar to the VBID group.

To select the weights, we used EB to constrain the standardized mean differences (SMDs) of observable preintervention characteristics between VBID participants and the weighted comparison group. For a particular covariate Z, the SMD is defined as the mean in the treated group minus the weighted mean in the control group, divided by the standard deviation in the treated group. In other words, an SMD of 0 indicates that the mean of the covariate for the treated observations is equal to the weighted mean of the control observations, and an SMD of 0.1 indicates that the difference in means is equal to 0.1 standard deviations. A rule of thumb is that an SMD below 0.2 indicates acceptable balance between treatment and comparison groups, while an SMD above 0.2 indicates unacceptable balance (Cohen, 1977).

As in prior reports (e.g., Khodyakov et al., 2022), we modified the standard EB algorithm to produce weights that balance the covariates within a prespecified range (or *tolerance*) of SMDs. For example, we can estimate weights to consider any SMD with an absolute value below $\delta = 0.1$ to be balanced.

Choosing δ represents a trade-off between bias and variance (Wang and Zubizarreta, 2020). The amount of information in the weighted sample, and thus the potential statistical efficiency of the DD estimates, can be measured using Kish's effective sample size (ESS) (Kish, 1965) and is defined as $ESS(w) = \frac{(\sum w_i)^2}{\sum(w_i^2)}$ for a set of weights w_i. The ESS can range from 1 to the original sample size N. A low ESS implies that there might be insufficient information in the sample and that it is difficult to find comparable units between the two groups. Larger values of δ will lead to larger ESSs, but maintaining larger ESSs could reduce the balance between the groups. In practice, SMD values lower than δ = 0.1 are customarily used when the goal of balancing is to fully control for confounding from observable characteristics (Austin, 2009; Stuart, Lee, and

Leacy, 2013). Because our empirical strategy also uses DD (which does not require balance on baseline characteristics to deliver unbiased estimates), we set 0.2 as a target threshold value for the SMD between VBID and comparison plans after weighting.

As noted above, we derived EB weights separately for each participation pattern and outcome. For consistency in estimating the large number of outcomes and participation patterns in this evaluation, we used an automated approach to selecting the tolerance for the EB algorithm, in which we specified a maximum acceptable tolerance and then evaluated successively larger tolerances until the ESS of the weighted comparison group was no smaller than 90% of the number of VBID-participating plans. To calculate the weights, we used the smallest tolerance for which the ESS of the comparison group met this threshold.

Variables Included in EB

The practical value of weights derived from EB depends on the set of balancing variables included in the EB algorithm. We followed the approach used in our prior report (Eibner et al., 2023b) and balanced simultaneously on two groups of variables:

- characteristics of VBID participants and comparison plans observed prior to VBID plans' participation in the VBID Model (*baseline characteristics*)
- trends in the outcome variable observed prior to the first year of VBID implementation (*pre-VBID outcome trends*).

We used a comprehensive set of baseline characteristics, including beneficiary demographics, plan characteristics, and characteristics of the local health care market. The included set of baseline characteristics varied, depending on the level of analysis (plan, beneficiary, or contract), but we used the same set of baseline characteristics for all analyses at a given level of analysis. Details on the included variables and levels of balance achieved for each level of analysis (plan, beneficiary, and contract) are shown under "Additional Estimation Details."

The pre-VBID outcome trends that we included in balancing for plan- and contract-level analyses, in contrast, were specific to each outcome (for beneficiary-level outcomes, which have a larger sample, we included all outcome trends in balancing). For a given outcome and participation pattern, we constructed differences of the outcome variable over all combinations of years prior to VBID implementation. To illustrate, the EB weights for our analysis of an outcome variable Y_{it} (for example, MAPD bids) among plans with participation pattern $\boldsymbol{a} = \mathbf{11111}$, which denotes plans that implemented VBID in 2020 and continued through 2024, balanced on the change in the outcome between 2017 and 2018 ($Y_{i2018} - Y_{i2017}$), the change between 2018 and 2019 ($Y_{i2019} - Y_{i2018}$), and the change between 2017 and 2019 ($Y_{i2019} - Y_{i2017}$).

One challenge in implementing this approach is that some VBID plans existed for a relatively short length of time prior to joining VBID and did not have preperiod data extending back three years. For example, Table A.1 shows that 65 new plans that were established in 2022 joined

VBID in 2023 (pattern NANA01). We imputed preintervention outcome data for these plans jointly with other missing variables, using the approach described under “Imputation of Missing Data.” To avoid introducing artifacts from the imputation process, we restricted the comparison group for plans that did not exist in some preperiod years to include only eligible nonparticipating plans that were established in the same year. For example, the comparison group for pattern NANA01 was limited to plans with participation history NANA00.

Imputation of Missing Data

The raw data used in our analyses contain missing information about covariates and outcomes—both before and during the VBID Model—for at least some observations. However, EB and subsequent steps in our analysis required a dataset with no missing data.

Therefore, we imputed missing covariate and outcome information jointly using a Markov chain Monte Carlo method known as Generalized Efficient Regression Based Imputation with Latent Processes (GERBIL), a sophisticated approach for jointly imputing missing-at-random data (Robbins, 2024). GERBIL was developed to address theoretical and computational limitations of the better-known multiple imputation by chained equations (MICE) approach. Key advantages of GERBIL include the ability to accommodate arbitrary variable types, including binary, categorical, and other variables with restricted ranges, as well as formulation of the data-generating process that guarantees existence of a joint probability distribution over the included variables. MICE, in contrast, permits weaker parametric restrictions on the included variables but can yield results that are inconsistent with any possible joint distribution, which can lead to instability in the resulting imputation datasets. GERBIL specifies all variables as functions of jointly normal latent variables, using suitable transformations, such as copulas for continuous variables, where needed. Details on these transformations are given in Robbins, Ghosh, and Habiger (2013) and Robbins (2014).

After transformation, GERBIL assumes that all variables are latently joint multivariate normal, takes imputation draws from this joint distribution, and then back-transforms to return all imputed values to the observed scale. Robbins (2024) is the authoritative source on GERBIL and should be consulted for technical details.

This imputation strategy ensures that all plans contribute to all analyses. As the VBID Model has extended over time, the number of plans that had missing data in any postperiod year has grown. Thus, we needed an approach that kept such plans in the analysis.

DD Analysis

To identify $ATT(\boldsymbol{a}, t)$ using observed data, we used a DD design to estimate these models using the EB weights described above. For this, we used weighted least squares for all models at the contract, plan, and beneficiary levels of analysis. (We discuss further the choice of statistical models for beneficiary-level models under “Beneficiary-Level Analyses.”)

We specified DD models to account for any time-invariant unobserved differences between VBID and comparison plans and for any common factors that could simultaneously affect outcomes across all plans during the postintervention period. Specifically, we let Y_{it} denote the outcome for plan i at time t, let $VBID_i$ indicate that plan i is a VBID-participating plan, and let DD_{it} denote the DD indicator for plan i at time t (DD_{it} = 1 for VBID-participating plans in the postintervention period and 0 otherwise). For each participation pattern, we estimated a weighted DD model of the form

$$Y_{ti} = \alpha_i + \eta_t + \beta_t \cdot DD_{it} + \varepsilon_{it} \qquad \text{(Equation A.3)}$$

for each outcome year in which plans in that participation pattern participated in VBID. α_i is a plan-specific intercept; η_t is a time fixed effect; β_t is the effect of VBID participation in year t, and ε is an error term that is mean 0, conditional on the included explanatory variables. Plan outcomes in years of VBID discontinuation (nonparticipation after participation) are captured by the time-varying coefficients β_t and thus do not contribute to the estimated effects of VBID participation in other years.

As described above, we fit a separate DD model for each of the participation patterns ($\boldsymbol{a}$), so β_t is the estimate of $ATT(\boldsymbol{a}, t)$ for those groups of plans. We then aggregated β_t for each group of participating plans to obtain calendar-year effects as described above. We derived variance estimates using a smooth version of the bootstrap, such that plans were repeatedly reweighted using a beta distribution to approximate the sampling distribution.

Validity of the DD Design

DD designs rely on a parallel-trend assumption to identify causal effects. This assumption states that the postparticipation trend in the outcomes for the comparison group is equal to the trend for each VBID participation pattern had the comparison plans not participated in VBID. To bolster the plausibility of this assumption, we assumed that parallel trends hold within levels of observed variables $\boldsymbol{X}_{it}$ after applying the EB weights. Equation A.4 represents this assumption:

$$E[Y_{it*}(0) - Y_{it}(0) | \boldsymbol{A} = \boldsymbol{a}, \boldsymbol{X}_{it*} = \boldsymbol{x}] = E[Y_{it*}(0) - Y_{it}(0) | \boldsymbol{A} = \boldsymbol{0}, \boldsymbol{X}_{it*} = \boldsymbol{x}], \qquad \text{(Equation A.4)}$$

where $t*$ is some time period post–VBID implementation, t is some time period prior to VBID implementation, and the expectation on the right-hand side (for $\boldsymbol{A} = \boldsymbol{0}$) pertains to the distribution after EB weights have been applied to the comparison group. If this assumption holds, then the mean counterfactual outcome, absent treatment among members of the treatment group (which is inherently unobservable), can be expressed in terms of the pretreatment outcomes among members of the treated group plus the observed trend in the comparison group.

The DD methodology does not require that the balancing characteristics be perfectly balanced or that they be sufficient to control for confounding, as long as the parallel-trend assumption described in Equation A.4 holds. Rather, a DD model works under the assumption that the postparticipation trend in the outcome for the comparison plans is a proxy for the trend in the VBID-participating plans, had they not participated in VBID, and then comparing the participating and comparison plans' changes from preparticipation outcome to postparticipation outcome.

Under "Additional Estimation Details," we summarize balancing on preparticipation outcome trends by reporting the average SMD for each outcome pretrend variable used in balancing. The tables in that section show that weighting was generally able to achieve average SMDs in preparticipation outcome trends below 0.1 for each level of analysis. Readers interested in figures that illustrate balancing's impact on preparticipation trends for selected outcomes should consult Appendix C of our 2023 evaluation report (Eibner et al., 2023b).

Inference

We used a smooth version of the bootstrap that accounted for dependencies across time and within plans by generating 200 sets of plan-level weights, in which the weights were generated from the beta distribution. Holding the balancing weights fixed, we multiplied the balancing weights by the bootstrap weights and recomputed the two-way fixed-effect estimates for each new set of weights and conducted the relevant aggregation. We then calculated the empirical standard deviation of the bootstrap using a normal approximation to generate confidence across estimate intervals.

Additional Estimation Details

In this section, we provide details about methods that were specific to analyses at the three levels of analysis in this evaluation—plan, beneficiary, and contract—emphasizing

- details of balancing variables
- results of balancing
- departures from the overall methodology described above
- differences from last year's report specific to the level of analysis.

Plan-Level Analyses

Table A.3 lists selected baseline characteristics used for EB in the plan-level analyses and reports the SMDs between VBID and comparison plans in these outcomes, both before and after weighting. The table shows that EB succeeded in reducing the SMD between VBID and comparison groups below 0.2 for all balancing variables shown. Preparticipation outcome trends had an SMD of 0.03.

Table A.3. Selected Balancing Variables Included in Plan-Level Analyses

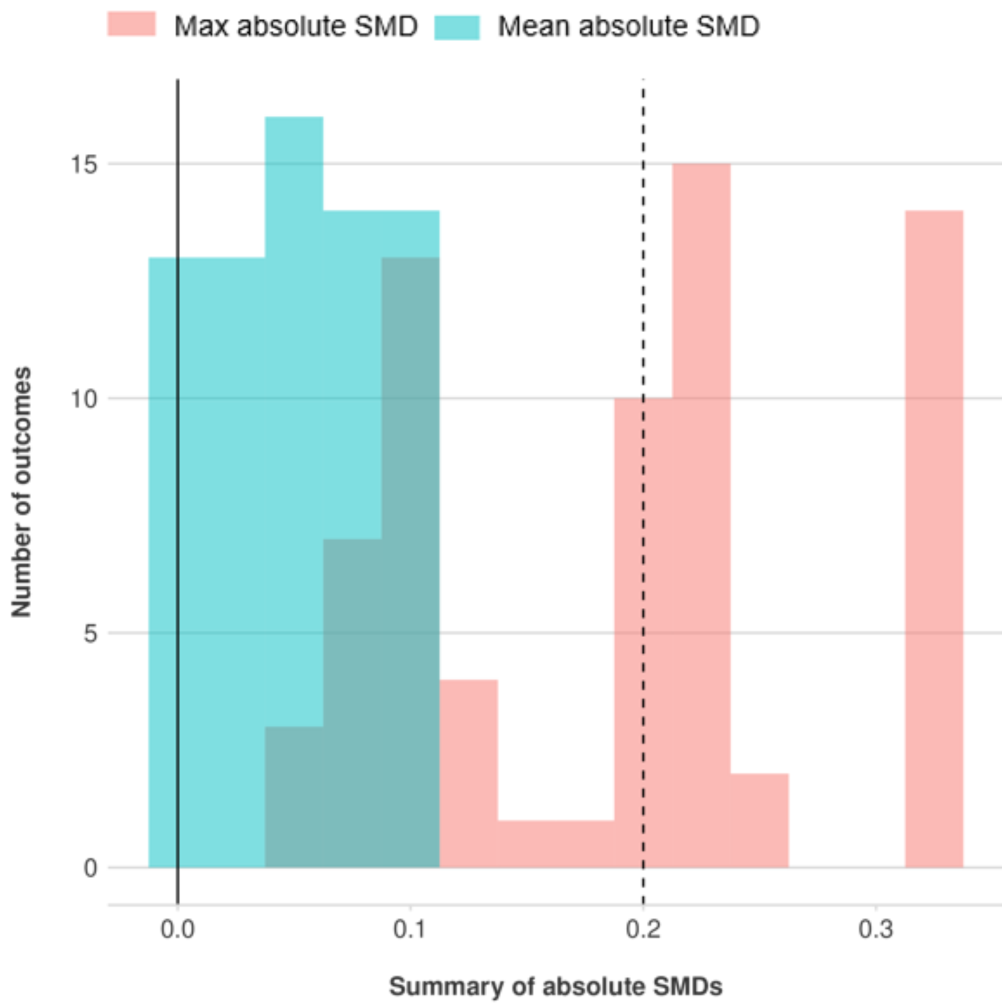
Variable	Unweighted Absolute SMD[a]	Weighted Absolute SMD[b]
Age	0.64	0.12
Star Rating (overall)	0.17	0.13
COVID-19 cases per 10,000	0.07	0.04
Percentage disabled	0.80	0.13
Percentage dual eligible	0.68	0.12
For profit (beneficiary months)	0.22	0.03
For profit (enrollment)	0.24	0.02
Percentage cancer	0.17	0.06
Percentage congestive heart failure	0.32	0.03
Percentage chronic obstructive pulmonary disease	0.63	0.14
Percentage diabetes	0.49	0.12
Hospice participant 2021	0.02	0.01
Hospice participant 2022	0.01	0.00
Hospice participant 2023	0.03	0.01
Hospice participant 2024	0.02	0.00
Health professional shortage area	0.01	0.00
Newly transitioned into bonus	0.02	0.01
Percentage LIS status	0.66	0.11
Sex	0.35	0.06
Area-level income	0.29	0.06
Missing outcomes	0.04	0.01
Newly transitioned out of bonus	0.03	0.01
Part D basic premium	0.24	0.05
PDSS participant 2021	0.04	0.02
PDSS participant 2022	0.05	0.06
PDSS participant 2023	0.05	0.07
Part D supplemental premium	0.50	0.02
Part D total premium	0.20	0.05
Offered SSBCI 2020	0.02	0.01
Offered SSBCI 2021	0.10	0.06
Offered SSBCI 2022	0.13	0.04
Offered SSBCI 2023	0.15	0.05
Offered SSBCI 2024	0.03	0.10
Offered UF 2020	0.03	0.00
Offered UF 2021	0.08	0.02
Offered UF 2022	0.02	0.03

Variable	Unweighted Absolute SMD[a]	Weighted Absolute SMD[b]
Offered UF 2023	0.08	0.07
Offered UF 2024	0.04	0.05
Part C in-network OOP maximum	0.36	0.07
Urban	0.09	0.02
Rural	0.05	0.03
Suburban	0.09	0.03
Percentage over age 65	0.09	0.04
MA penetration	0.06	0.04
Plan type (preferred provider organization = 1; otherwise = 0)	0.07	0.02
Average MA risk score (HCC)	0.49	0.06
Average Part D risk score (RxHCC)	0.68	0.09
Puerto Rico county	0.02	0.00
C-SNP	0.03	0.02
DSNP	0.33	0.08
ISNP	0.04	0.02
Standardized Medicare costs per capita	0.14	0.09
Preparticipation outcome trends	0.13	0.03

SOURCE: Authors' analysis of VBID-participating plan and other data.
NOTE: COVID-19 = coronavirus disease 2019; DSNP = Dual Eligible Special Needs Plan; HCC = hierarchical condition category; ISNP = institutional special needs plan; LIS = low-income subsidy; PDSS = Part D Senior Savings; RxHCC = prescription drug HCC; SSBCI = Special Supplemental Benefits for the Chronically Ill; UF = uniformity flexibility. Absolute SMDs for covariates reported in the table were calculated by averaging the SMD for each covariate across all models (that is, participation patterns and years) for a given outcome variable, weighting patterns by the number of VBID-participating plans in the sample, and then taking the absolute value of the average SMD.
[a] These reflect differences without EB weights.
[b] These reflect differences using EB weights.

Figure A.1 shows the effects of balancing at the level of the outcome variable rather than the balancing variable. The figure is a histogram of the average (turquoise bars) and the maximum (pink bars) absolute SMDs achieved for each outcome, aggregating over all models and balancing variables. The absolute average SMD for all outcome models was 0.112 or below.

Figure A.1. Summary of Mean and Maximum Absolute SMDs After Balancing, Plan-Level Analyses



NOTE: Absolute SMDs are calculated by first averaging the SMD for each covariate across all models (that is, participation patterns and years) for a given outcome variable, weighting patterns by the number of VBID-participating plans in the sample, and then taking the absolute value of the average SMDs. To produce a maximum and mean absolute SMD at the level of the outcome variable, we then took the maximum and the mean of the covariate-specific absolute SMDs for each outcome. The darker turquoise shading indicates overlap between the maximum absolute SMD for some variables and the average absolute SMD for others. Our target balancing threshold, 0.2, is shown with a dotted line.

Change in Methods: Preperiod Years Included in the 2023 and 2024 Outcome Models

We revised the methodology used in this project report for 2023 and 2024 outcomes to reduce the number of preintervention years included in the analysis. For 2022 and earlier outcome years, we included all years from 2017 forward as the preperiod for all participation patterns, meaning that the length of the preperiod was between three and five years (depending on when VBID was first implemented). As we began to analyze later years of the model test and participation patterns with later VBID implementation years, a larger share of plans did not exist in the earlier years of the study period, which increased the sensitivity of the results to our imputation model. It was also undesirable to have a wide range of variation in the length of the preintervention period within a given outcome year. For example, analyses of 2024 outcomes might have had preperiods ranging from three years (for plans adopting VBID in 2020) to seven years (for plans adopting VBID in 2024) under our prior method.

For these reasons, analyses of outcomes in 2023 and later years of the model test used only the three years leading up to VBID implementation as the preperiod for both balancing and DD estimation. As discussed in Chapter 4 of the main project report, this change might have contributed to changes in the enrollment estimate from results published in our 2025 report (Eibner et al., 2025). Previously published results for other outcomes did not change meaningfully, however.

Beneficiary-Level Analyses

The beneficiary-level analyses largely followed the methodology used in our 2025 evaluation report (Eibner et al., 2025). However, we made a substantial change to the way we modeled inpatient stays, and that change had minor downstream impacts on other outcomes.

In previous evaluation reports, we used Poisson regression to model beneficiary-level inpatient utilization because Poisson regression is generally more appropriate for count data than linear regression is. Although Poisson regression is a theoretically appropriate statistical model, the form of the conditional mean function used in Poisson regression made it challenging to combine DD with balancing on pre-VBID trends. Specifically, Poisson regression models the conditional mean of outcomes as an exponential function, meaning that a unit change in an explanatory variable is assumed to result in a constant percentage effect on the expected outcome. When Poisson regression is used for DD estimation, the parallel-trend assumption needed to identify causal effects is that the percentage change from pre- to posttreatment periods in the mean outcome in the absence of treatment (E[Y(0)]) is identical in the treatment and control groups. However, our EB strategy of balancing on within-unit changes in outcomes would lead to violations of this parallel-trend assumption unless the mean untreated outcome were identical in both groups (which was not the case). An appealing alternative strategy (balancing on the beneficiary-level change in the natural log of the outcome) was not feasible because the natural log is not defined when the outcome is 0.

Instead, we used the log of the plan-level average outcomes to construct pretrend variables for use in balancing, in which we computed the plan-level average among all beneficiaries in the plan. This approach was similar to the balancing approach used for other outcomes but was not rigorously justified because it is not guaranteed to achieve the parallel-trend assumption needed for Poisson DD estimation to recover causal effects.

For the current project report, we decided to stop using Poisson regression to model inpatient stays and instead used linear regression to model all outcomes. This means that the parallel-trend assumption required for inpatient stays was defined in terms of the level of the outcome, not in percentages. We applied this change in statistical models to our analysis of inpatient stays in all three outcome years examined in this project report (2020, 2021, and 2022).

As noted above, the change from Poisson regression to linear regression also led us to change the balancing variable used for inpatient stays from the log of plan-level average inpatient stays to the beneficiary-level change in inpatient stays. Importantly, this change affected the balancing weights for all outcomes because we balanced simultaneously on all pretrends for all outcomes. This approach is both more consistent with the approach used for other outcome variables and consistent with the parallel-trend assumption necessary for linear regression to recover causal effects. This change in the included balancing variables also led us to recalculate the EB weights for earlier years using the beneficiary-level change in inpatient stays instead of the log of plan-level average inpatient stays.

Beneficiaries Included in the Analyses

Like we did for the 2025 report (Eibner et al., 2025), we included beneficiaries with any preperiod data—those who were in another VBID-participating plan, another MA plan, or a fee-for-service (FFS) Medicare—in the analysis. We defined the preperiod based on the beneficiary's first exposure to VBID. For example, if a beneficiary entered a VBID-participating plan in 2022, the beneficiary's preperiod would include 2021, even if the plan were a 2021 VBID participant. Table A.4 shows that we retained more than 99% of VBID-targeted beneficiaries in our analysis.

Table A.4. Preperiod Data Patterns, New Beneficiary Selection Approach

Preperiod Data Pattern	Targeted Beneficiaries in 2020	Targeted Beneficiaries in 2021[a]	Targeted Beneficiaries in 2022[a]
2+ years of preperiod data in either MA or FFS	203,588	1,128,169	2,124,183
1 to 2 years of preperiod data in either MA or FFS	32,690	273,644	399,232
Part-year preperiod data in either MA or FFS	26,391	238,497	712,737
No preperiod data	314	3,849	10,390
Total targeted beneficiaries	262,983	1,644,159	3,246,542
Percentage with any preperiod data	99.9	99.8	99.7

NOTE: Estimates include all beneficiaries, whether enrolled in MA-only plans or MAPDs.
[a] Includes newly targeted beneficiaries in the stated year and those observed from 2020 onwards.

To accommodate these newly included beneficiaries in ways that credibly distinguished changes in outcomes associated with VBID from changes associated with plan switching, we stratified the analyses based on the length of preperiod data and whether the beneficiary had been previously enrolled in MA or FFS. We describe this further below.

Stratification

Although including beneficiaries with preperiod data outside their VBID plans helped make the analysis more representative of the MA population, differentiating changes in outcomes related to VBID from changes in outcomes related to coverage transitions was challenging. Beneficiaries moving from FFS to MA can experience changes in utilization due to exposure to plan networks and utilization management processes unrelated to VBID. Furthermore, data collected in FFS can differ systematically from data collected in MA plans, which complicates the analysis. For example, if risk scores are systematically higher in MA because it has higher coding intensity than FFS, then risk scores from FFS are not comparable to risk scores in MA and should not be considered equivalent when balancing, even if they are numerically identical.

Our estimation procedure weighted the comparison group to resemble the VBID group on preperiod outcome trends and other characteristics. As described above, we used the GERBIL algorithm to impute these variables when they were missing. By expanding the sample to include beneficiaries with minimal preperiod data, we increased the reliance on imputation to estimate missing preperiod characteristics and trends, which could have introduced bias. For example, if imputed trends were slightly steeper than actual trends, this imputation might have attributed a post-VBID flattening off to the VBID Model, even though it actually might have stemmed from imperfect imputation.

To address these potential issues, we stratified the data to ensure that VBID and comparison beneficiaries had the same length of preperiod information and the same source of preperiod coverage (MA or FFS). We then ran separate regressions for each stratification group and combined the estimates to develop the overall VBID effect. As a result, we compared only VBID beneficiaries newly joining a VBID plan from FFS with non-VBID beneficiaries newly joining a non-VBID plan from FFS. Similarly, we compared only VBID beneficiaries with limited preperiod data with non-VBID beneficiaries with limited preperiod data. Restricting the comparisons in this way reduced the potential for bias arising from FFS-to-MA transitions or from imperfect imputation. For example, if all beneficiaries transitioning from FFS to MA experienced increases in risk scores, the stratified DD approach would have attributed these increases to the VBID Model only if the changes for beneficiaries who switched from FFS into VBID plans were larger than the changes for beneficiaries who switched from FFS into non-VBID plans. Similarly, even if the imputation approach introduced biases in modeling trends for VBID and non-VBID beneficiaries, our method could still recover unbiased DD estimates, as long as the imputation bias was independent of VBID status.

Below, we list the stratification groups that we use in our analysis:

1. at least two years of prior-period data, MA
2. at least two years of prior-period data, FFS
3. at least one but less than two years of prior-period data, MA
4. at least one but less than two years of prior-period data, FFS
5. partial-year preperiod data, MA

6. partial-year preperiod data, FFS.

Where possible, we further stratified these groups to capture dual and non–dual eligibility status, resulting in 12 groups. This approach ensured that our effect estimates adjusted for any differences in preperiod coverage source (MA or FFS) between VBID and comparison beneficiaries. We excluded beneficiaries with no preperiod data.

In addition, we implemented the stratifications by plan participation pattern described earlier in this appendix. Table A.5 shows the number of VBID-targeted beneficiaries in each stratification group for the 2022 outcome year. Like we did for our 2025 report, we applied these methods to analyze outcomes in 2020 and 2021. The set of stratifications in those years and the number of beneficiaries in the sample are identical to those reported in Tables A.4 and A.5 in Appendix A of our 2025 report (Eibner et al., 2025); see that report for further details.

Table A.5. Number of VBID-Targeted Beneficiaries in Each Stratification Group Used in the 2022 Analysis

Beneficiaries' Data Pattern	Participated in 2020, 2021, and 2022	Participated in 2021 and 2022 only	Participated in 2022 only	Participated in 2020 and 2022 only	Total
At least 2 years of preperiod data, FFS	6,752	58,893	91,384	714	157,743
At least 2 years of preperiod data, MA	96,864	897,859	771,818	72,474	1,839,015
At least 1 but less than 2 years of prior-period data, FFS	1,056	135,587	14,748	156	151,547
At least 1 but less than 2 years of prior-period data, MA	17,019	13,038	186,928	6,912	223,897
Partial-year preperiod data, FFS	4,925	32,387	100,764	1,246	139,322
Partial-year preperiod data, MA	11,035	241,951	283,196	4,729	540,911
Total	**137,651**	**1,379,715**	**1,448,838**	**86,213**	**3,052,435**

NOTE: Sample includes VBID-targeted beneficiaries who were enrolled in MAPDs.

Details of Beneficiary-Level Entropy Balancing

As discussed above, the EB algorithm allows specification of the desired level of covariate balance δ for each covariate. The algorithm attempts to solve for weights that satisfy the balance constraint. However, not all values of δ are feasible; when covariate balance is low, the algorithm can fail to converge.

For all beneficiary-level analyses, we attempted to balance all pretrends, first by setting $\delta = 0.05$ SMDs for pre-VBID outcome trends and $\delta = 0.1$ SMDs for all baseline characteristics.[1] If these initial tolerances were not feasible, we iteratively increased each tolerance by 0.05 until convergence was possible. Finally, for rare (prevalent) binary variables with prevalences less than 0.05 (greater than 0.95), we specified that the EB algorithm apply the tolerance δ to a difference that fixes the standardization in the treatment group to be based on a prevalence of 0.05 (0.95). For example, when the desired SMD was 0.1, we instead set $\delta = 2.2$ percentage points ($0.1 * \sqrt{0.05 * 0.95}$). For these variables, we also calculated SMD standardized with respect to a variable that had a prevalence of 0.05 (0.95).

Several outcomes measuring drug adherence or receipt of recommended care were defined only for subgroups of beneficiaries. For example, adherence to noninsulin diabetes medication was measured only for people with diabetes. Therefore, we created a separate set of balancing weights for each of these outcomes (statin adherence, hypertension drug adherence, diabetes drug adherence, and breast cancer screening) in addition to the weights used for outcomes that were defined for all beneficiaries (risk score, inpatient stays, emergency department [ED] visits, and Part D out-of-pocket [OOP] spending).

To minimize the number of weights we had to derive for this analysis, we included pretrends for all possible beneficiary-level outcomes in a single set of weights.[2] This approach allowed us to use the same weights for each regression within the same stratification patterns (1 to 6, above), regardless of the outcome. When deriving weights for pretrends for outcomes that did not apply to all beneficiaries in the dataset (for example, breast cancer screening), we balanced on the average plan-level outcome trends within the subgroup. Finally, when balancing on pretrends for outcomes used in Poisson models (for example, number of ED visits, number of inpatient stays), we balanced the difference in the log of the plan-level average outcomes, where the plan-level average was computed from data on all beneficiaries in the plan.

Variables Included in EB

Achieving balance across a broader set of covariates proved easier for some participation patterns than for others, which posed a challenge during implementation of the above-described method. For the 2020 analyses, we were able to stratify all analyses on both groups defined above, as well as dual eligibility status (measured in 2019). We also controlled for several plan-

[1] We imposed stricter tolerance for beneficiary-level balancing than for plan-level balancing because we had many more beneficiaries than plans in both the intervention and comparison groups (for example, hundreds of thousands of beneficiaries but hundreds of plans). As a result, we had more flexibility to enforce strict balance in the beneficiary analysis without leading to low effective sample size.

[2] We included all trends in the beneficiary balancing weights, whereas we derived separate weights for each plan-level regression that included the trend for only the outcome under consideration. We had more flexibility to include multiple trends in the beneficiary-level balancing weights because we had many more beneficiaries than plans in our analytic sample.

and contract-level covariates that we did not ultimately include for the 2021 analyses. When we attempted to apply the same approach to the 2021 and 2022 participation patterns, we were unable to achieve a satisfactory level of covariate balance within many of the subgroups. Therefore, we adjusted our stratification variables and covariate sets for the 2021 and 2022 participation patterns relative to 2020. Table A.6 summarizes the differences across each analysis.

Table A.6. Balancing Approach, by Analysis Year and Participation Pattern

Effect Year	Participation Pattern[a]	Stratification	Covariate
2020	1	Group, dual	Beneficiary, plan, contract
2021	01	Group	Beneficiary
2021	11	Group	Beneficiary
2022	111	Group	Beneficiary
2022	101	Group	Beneficiary
2022	011	Group	Beneficiary
2022	001	Group	Beneficiary

NOTE: For 2020, all VBID-targeted beneficiaries in participating plans were assigned a participation pattern of 1. In 2021, beneficiaries who were in VBID plans in both years were assigned a participation pattern of 11, while beneficiaries who were new to the model were assigned a participation pattern of 01. We note that 2021 and 2022 analyses include four plan-level covariates and one contract-level covariate in addition to the beneficiary-level covariates; these are described next.
[a] This is the beneficiaries' history of VBID participation.

Table A.7 provides a more detailed list of variables used in each analysis for each participation pattern for the 2022 results. The balance results for 2021 and 2020 were very similar to the results presented in Table A.7 of the 2025 evaluation report (Eibner et al., 2025). For example, we identified the same ten variables with SMD imbalances greater than 0.1 following both methodologies.

Table A.7. Selected Balancing Variables Included in the 2022 Beneficiary-Level Analyses and Absolute SMDs

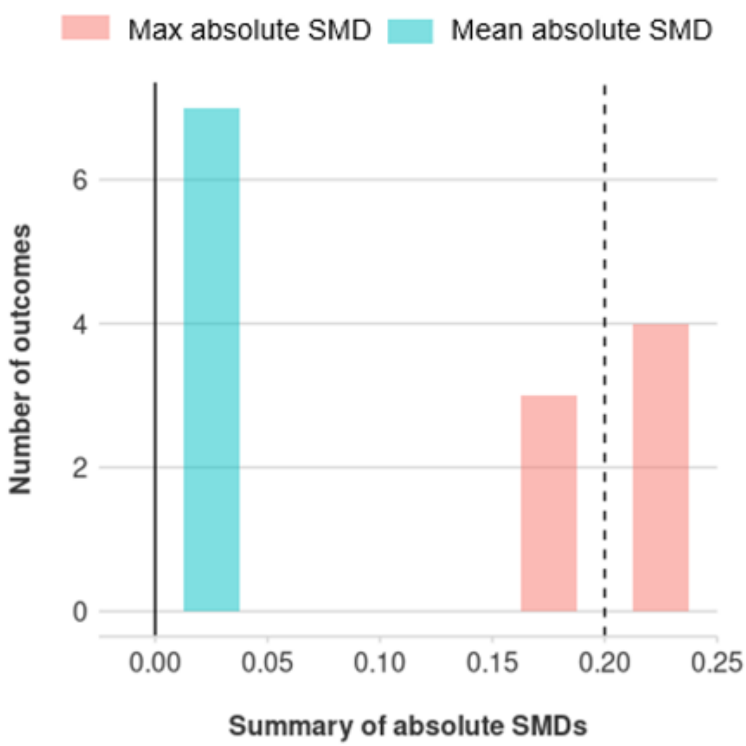
Variable	Unweighted Absolute SMD	Weighted Absolute SMD
Preoutcome trends		
Breast cancer screening	0.04	0.03
OOP cost	0.20	0.02
Diabetes	0.09	0.01
Hypertension	0.14	0.02
Number of ED visits	0.04	0.02
Number of inpatient stays	0.03	0.02

Variable	Unweighted Absolute SMD	Weighted Absolute SMD
Risk score	0.05	0.03
Statin	0.16	0.02
Beneficiary		
≥ $300 in total monthly Part D spend	0.53	0.09
≥ 2 ED visits	0.03	0.01
≥ 2 inpatient stays	0.00	0.00
≥ 8 concurrent medications	1.23	0.20
Age	0.34	0.07
Chronic RxFill	0.47	0.10
Disabled	1.25	0.16
Dually eligible	3.24	0.10
End-stage renal disease	0.00	0.00
Fall risk	0.02	0.04
HCCs	0.12	0.03
LIS level	0.68	0.02
Male	0.22	0.05
Missing outcomes	0.17	0.02
Months in CMS	0.21	0.01
Nonadherent for specific drugs	0.20	0.06
Medication Therapy Management eligible	0.51	0.03
Puerto Rico	0.30	0.09
RxHCCs	0.22	0.03
Social deprivation index	0.55	0.06
Plan		
Health professional shortage area	0.00	0.00
Months enrolled in plan	0.02	0.02
Plan enrollment	0.18	0.09
Special needs	1.20	0.04
Contract		
Star Rating	0.05	0.10
County		
COVID-19 cases per 10,000 population	0.09	0.07
Median income	0.42	0.10
Standardized Medicare costs	0.31	0.02
Urbanicity	0.00	0.00

SOURCE: Authors' analysis of VBID-participating plan and other data.
NOTE: RxFill = prescription fill status notification. Absolute SMDs are calculated by first averaging imbalances across strata within each participation pattern and then taking the absolute weighted mean average imbalance across participation patterns.

Figure A.2 shows the average (turquoise bars) and the maximum (pink bars) absolute SMDs achieved for each beneficiary outcome, aggregating over all models and balancing variables. The absolute average SMD for all outcome models was under 0.05.

Figure A.2. Summary of Mean and Maximum SMDs After Balancing, 2022 Beneficiary-Level Analyses



NOTE: Absolute SMDs are calculated by first averaging imbalances for each covariate across strata within each participation pattern (weighted by the number of treated units) and then taking the maximum and mean absolute weighted mean average imbalance across the four 2022 participation patterns. The target balancing threshold, 0.2, is shown with a dotted line.

Statistical Models and Estimation

The validity of any DD design relies on a parallel-trend assumption; however, there are several approaches to estimating causal effects using this design. For some posttreatment time period t^*, we parameterized the causal quantity $ATT(\boldsymbol{a}, t^*)$ via the term $\beta_{\boldsymbol{a}t*}$ in the following two-way fixed-effect model:

$$E[Y_{it} \mid \zeta_i, \delta_t, X_{it}, A_{it}] = \zeta_i + \eta_t + \beta_{\boldsymbol{a}t*}\, DD_{it} + \delta X_{it}, \qquad \text{(Equation A.5)}$$

where ζ_i and η_t are beneficiary and year fixed effects, respectively; X_{it} is a small set of time-varying controls that we believe to be exogenous to VBID participation (indicating Part D Senior Savings [PDSS] participation, uniformity flexibility (UF), Special Supplemental Benefits for the

Chronically Ill [SSBCI], and primarily health-related supplemental benefit offerings at time t); and DD_{it} denotes an indicator of both treatment status and the posttreatment time period t^*. The data used to estimate this model contain observations from patterns $\boldsymbol{a}$ and not-yet-treated beneficiaries (as of time t^*). Although this specification might allow for multiple posttreatment time periods, we included only the targeted time period t^* in the equation and therefore in the data used to estimate this model. For example, consider the beneficiary-level participation pattern 111. In 2022, we were interested only in the 2022 effect for this participation pattern, meaning that we did not estimate the 2021 effect for these beneficiaries and therefore did not include data from 2021 to estimate the 2022 effect. We then estimated this model using weighted linear regression, in which the weights reweighted the comparison observations to match the covariate distribution of the observations in participation pattern $\boldsymbol{a}$.

To obtain overall effect estimates for each posttreatment year, we averaged each of the participation pattern–specific estimates, weighted by the number of beneficiaries participating in VBID plans in each participation pattern $\boldsymbol{a}$ in that year. To express these effect estimates as percentage changes for each participation pattern $\boldsymbol{a}$, we obtained estimates of the outcome with and without VBID using the model above. We then averaged these estimates across all patterns using the same weights and finally calculated the percentage change using these estimated average outcomes with and without VBID.

Contract-Level Analyses

Contract-level analyses and inference followed the same methodology as the plan-level analyses. We included a contract in the analysis if it had at least one VBID-participating plan or at least one eligible nonparticipating plan. As with the plan- and beneficiary-level analyses, we ran stratified regressions based on participation patterns. Table A.8 shows the stratification groups used in our analysis.

Table A.8. Participation Patterns Used in Contract-Level Analysis

Outcome Year and Inclusion Indicator	Participation History	Number of Contracts
2021, included	01	44
	11	25
	Total	**69**
2022, included	001	54
	011	39
	111	20
	NA01	12
	Total	**125**
2023, included	0001	80
	0011	57
	0111	36
	1111	18
	NA001	14
	NANA01	12
	NA011	9
	Total	**226**
2021, excluded	NA1	1
	Total	**1**
2022, excluded	101	2
	NANA1	2
	NA11	1
	Total	**5**
2023, excluded	1001	2
	1011	2
	NANANA1	9
	NANA11	1
	NA111	1
	Total	**15**

NOTE: Participation history concatenates participation patterns for each year of the model test up to the outcome year. A 0 in a pattern indicates that the contract had no participating plans, a 1 indicates that the contract had at least one participating plan, and NA indicates that the contract did not exist in that year. For example, a pattern of 0111 indicates that the contract was observed in 2023 (four years into the model test) and had no participating plans in 2020 and at least one participating plan in 2021, 2022, and 2023.

Because we had difficulties in achieving balance, the contract analysis used a smaller number of balancing variables. Table A.9 lists the baseline characteristics used for EB in the contract-level analyses and reports the SMDs between VBID and comparison contracts in these outcomes both before and after weighting. Table A.9 shows that EB succeeded in reducing the SMD between VBID and comparison groups below 0.2 for all balancing variables shown. Preparticipation outcome trends had an SMD of 0.02 after balancing.

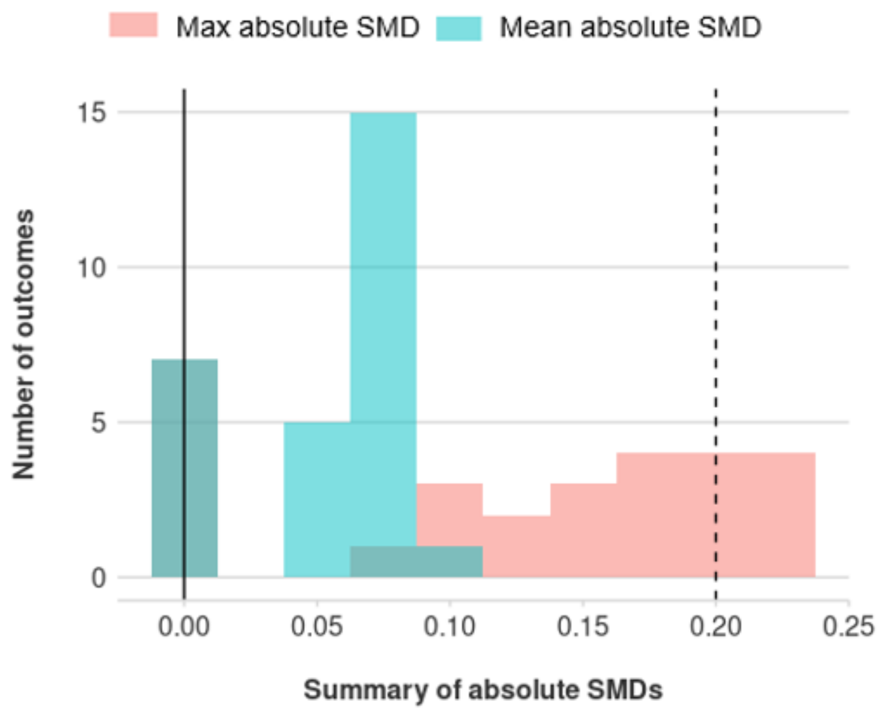
Table A.9. Balancing Variables Included in Contract-Level Analyses and SMDs

Variable	Unweighted Absolute SMD[a]	Weighted Absolute SMD[b]
Average age	0.64	0.08
COVID-19 cases per 10,000 population	0.14	0.06
Percentage disabled	0.66	0.08
Percentage dual eligible	0.51	0.05
For profit (beneficiary months)	0.25	0.04
Percentage LIS status	0.50	0.05
Missing outcomes	0.09	0.03
Part D basic premiums	0.32	0.12
MA penetration	0.25	0.03
Average MA risk score (HCC)	0.51	0.08
Average Part D risk score (RxHCC)	0.50	0.05
C-SNP	0.07	0.02
DSNP	0.41	0.05
ISNP	0.05	0.02
Standardized Medicare costs per capita	0.22	0.08
Preparticipation outcome trends	0.09	0.04

SOURCE: Authors' analysis of VBID-participating plan and other data.
NOTE: Absolute SMDs for covariates reported in the table were calculated by averaging the SMD for each covariate across all models (that is, participation patterns and years) for a given outcome variable, weighting patterns by the number of VBID participant contracts in the sample and then taking the absolute value of the average SMD.
[a] These reflect differences without EB weights.
[b] These reflect differences using EB weights.

Figure A.3 shows the average (turquoise bars) and the maximum (pink bars) absolute SMDs achieved for each beneficiary outcome, aggregating over all models and balancing variables. The absolute average SMD was under 0.10 for most models.

Figure A.3. Summary of Mean and Maximum Absolute SMDs After Balancing, Contract-Level Analyses



NOTE: Absolute SMDs are calculated by first averaging the SMD for each covariate across all models (that is, participation patterns and years) for a given outcome variable, weighting patterns by the number of VBID-participating plans in the sample, and then taking the absolute value of the average SMD. To produce a maximum and mean absolute SMD at the level of the outcome variable, we then took the maximum and the mean of the covariate-specific absolute SMDs for each outcome. The darker turquoise shading indicates overlap between the maximum absolute SMD for some variables and the average absolute SMD for others. The target balancing threshold, 0.2, is shown with a dotted line.

Change in Methods: Preperiod Years Included in 2023 and 2024 Outcome Models

As in the plan-level analyses discussed above, we revised the methodology used in this project report for 2023 contract-level outcomes to reduce the number of preintervention years included in the analysis. Contract-level analyses of outcomes in 2023 and later years of the model test used only the three years leading up to VBID implementation as the preperiod, for both balancing and DD estimation. Further details and discussion of the motivation for this change can be found under “Plan-Level Analyses” above.

We made one additional adjustment to the contract analyses to crosswalk plans between the measure year (which reflects the calendar year of the model test) and the display year for the Star Ratings. We made this change to accommodate an increase in crosswalking activity between the 2023 measure year and 2025 display year that was not as prevalent in prior measure year–display year pairs.

In addition, we added analyses of three Part D domains based on qualitative interview findings that challenged our prior hypothesis that these domains were unrelated to the VBID interventions. In our 2024 interviews, some PO interviewees noted changes in certain domains, such as the Part D Member Experience with the Drug Plan. Including additional outcomes caused changes to the balancing weights, which, in turn, could have caused changes to the point estimates. Therefore, we used separate balancing weights for the main outcomes (overall Star Rating and domain-level Star Ratings analyzed in our 2025 report) and the three additional Part D domains added to this project report.

Appendix B. Qualitative Methods

In this appendix, we describe our approach to collecting and analyzing the primary data from POs, which was reviewed and approved by the RAND Human Subjects Protection Committee. In 2024, we fielded a questionnaire to all POs that participated in the model test and subsequently sampled 27 POs for a virtual interview. The goal for these data collection activities was to provide additional nuance into the implementation experiences in participating in VBID and to describe how and why VBID participation was associated with key model outcomes. All interviews were conducted using an approach similar to the one we described in detail in previous VBID evaluation reports (Eibner et al., 2023a; Eibner et al., 2025; Khodyakov et al., 2022). Some of the text below is copied verbatim from those reports.

To recruit PO representatives, we emailed primary contacts at each PO and provided them with a brief description of our data collection procedures, their purpose, and logistical details. We conducted up to three follow-ups with those who had not responded to our invitations. We used a small-group approach to the interviews. We allowed contacts at each organization to invite colleagues whom they considered to be most knowledgeable about VBID to participate in the interviews. During the interview scheduling phase, we sent the consent form via email. We obtained spoken consent and answered any questions prior to beginning the interview. Each virtual interview was conducted using Zoom for Government software by a team that included up to two researchers and one research assistant who took detailed notes.

Sampling and Data Collection

We invited all 69 POs that participated in the VBID model in 2024 (56 that implemented only the VBID General component, five that implemented only the Hospice Benefit component, and eight that implemented both components) to complete the online questionnaire. We also invited a sample of 27 POs to participate in interviews (14 that implemented only the VBID General component, five that implemented only the Hospice Benefit component, and eight that implemented both components). We prioritized POs that had implemented the Hospice Benefit component, as well as new model test participants and those with more-complex interventions. Both data collection activities were meant to help us understand POs' experiences with specific model components; implementation barriers that they encountered; and the impact that they expect their VBID interventions would have on plan enrollment and retention, utilization of VBID benefits and services, beneficiary health outcomes, and plan and beneficiary costs in 2024. The questionnaires were developed after the review of POs' model test application materials and informed by the results of the PO data collection activities undertaken in previous evaluation years. Although the questionnaire items were primarily close-ended, interview questions were open-ended.

Because this project report focuses only on VBID General, we describe the types of questions we asked POs that implemented this model component. Although our close-ended questionnaire items that covered implementation experiences focused on various aspects of administrative processes and communication, the outcome questions focused on how VBID General affected different plan- and beneficiary-level outcomes, including different measures of care quality and costs, plan bids, and beneficiary care experiences.

During the interviews, we discussed POs' responses to the preinterview questionnaires and asked additional questions covering such topics as

- details of VBID General interventions
- implementation experiences, successes, and challenges
- VBID's impact on plan enrollment, care quality, health and financial outcomes.

We tailored the interview protocols based on whether the PO was a new or continuing model test participant. New POs answered additional questions related to reasons for joining the model test and the rationale behind their interview design.

Of the 69 POs invited to complete the questionnaire, 65 did so. Of these, 52 implemented only the VBID General component, five implemented only the Hospice Benefit component, and eight implemented both components. Of the 27 POs invited to participate in interviews, 25 completed interviews. Of the POs we interviewed, 13 implemented only VBID General, four implemented only the Hospice Benefit component, and the remaining eight implemented both components. We collected all PO data between June and August 2024. Although we spoke with a total of 153 PO representatives across the 25 POs, this project report is based on the interviews we collected with the representatives of 21 POs that implemented one or more VBID General interventions in 2024. Interviews were typically 60 minutes in length.

Data Analysis

All but two interviews were audio recorded and professionally transcribed. Close-to-verbatim notes were taken during the interviews in which PO representatives declined to have their interview recorded. Interviews and notes were coded thematically using the same process described in the previous reports (Khodyakov et al., 2022; Eibner et al., 2023a; Eibner et al., 2025). We used descriptive statistics to analyze the responses from the preinterview PO questionnaires to guide our qualitative analysis of the interview transcripts.

Following completion of data collection, we coded the interview transcripts using a thematic approach to uncover additional nuance not explained by the questionnaire data. We used a team-based approach to qualitatively code the transcripts in Dedoose, a software program for analyzing qualitative data. As in previous years, we refined the codebook used in the previous evaluation year based on new and emerging themes from interviews conducted in 2023 and

2024. We used the same coders as in previous years and relied on the same process for training them on how to use the codebook for team-based coding. Each coder individually coded a set of test transcripts, which was then reviewed by the broader coding team. One researcher resolved questions and discrepancies in coding. We calculated a combined kappa score of 0.76 using the Dedoose feature for interrater reliability (McHugh, 2012). Following the establishment of a reliable kappa score, coders independently coded the remainder of the interview transcripts. Each researcher assigned to each section of the project report then reviewed all the relevant codes for their section for consistency prior to analysis and write-up.

We used a thematic analysis (Guest, MacQueen, and Namey, 2012) to compare themes and explore patterns and variation in PO perspectives on and experiences with the model test. We also compared emerging themes from this round of data collection with the findings from previous rounds. Lastly, for consistency with the mixed-methods nature of this evaluation, we integrated quantitative and qualitative analytic techniques to ensure the rigor of our findings.

Appendix C. Descriptive Information on Participating Plans

Table C.1 presents descriptive statistics for participating and eligible nonparticipating POs in 2024.

Table C.1. Descriptive Statistics for Participating and Eligible Nonparticipating POs, 2024

Characteristic	VBID	Eligible Nonparticipating POs
Number of POs	64	92
Blue Cross affiliate (%)	17.19	17.39
PO's geographic reach (%)		
1–2 states	68.75	78.26
3–8 states	18.75	20.65
9 or more states	12.50*	1.09
For-profit status (%)	46.88	43.48
MA penetration (%)[a]	57.7*** (8.0)	51.8 (10.2)
Median income ($)[a]	31,025 (5,799)	30,988 (3,230)
Enrollment[a]	352,779* (1,172,290)	27,130 (38,842)

SOURCE: Authors' analysis of CMS and other data.
NOTE: ***, **, and * represent statistical significance at the 0.1%, 1%, and 5% levels, respectively.
[a] Means are provided, with standard errors in parentheses.

Table C.2 presents descriptive statistics for 2024 VBID participating and eligible comparison plans.

Table C.2. Descriptive Statistics for VBID-Participating and Eligible Comparison Plans, 2024

Characteristic	VBID-Participating Plans	Comparison Plans
N	1,400	3,100
Offers Part D (%)	99.64***	89
DSNP (%)	53.86***	2.55
C-SNP (%)	4.64	5.06
ISNP (%)	0.00***	4.55
$0 premium plan (%)	29.29***	66.77
Offers UF (%)	10.79**	8.1
Offers SSBCI (%)	22.5	21.03
Offers non–primarily health-related benefit (%)	94.79	93.42
Preferred provider organization (%)	30.79***	40.03
Total premium ($ PMPM)[a]	26.0 (22.6)***	18.5 (37.5)
Maximum OOP limit ($)[a]	6,445.9 (2434.0)***	5,067.3 (2,000)
Rural (%)[a]	9.2 (12.1)	8.5 (13.3)
Suburban (%)[a]	21.1 (14.9)***	19.3 (16.8)
Urban (%)[a]	69.7 (22.5)**	72.3 (25.2)
Dually eligible enrollees (%)[a]	58.5 (40.8)***	15.4 (22.9)
Part D LIS enrollees (%)[a]	66.5 (38.8)***	23.8 (24.5)
Age (years)[a]	67.8 (4.6)***	71.9 (3.6)
Male (%)[a]	42.8 (5.8)***	47.5 (8.9)
MA bid ($)[a]	884.7 (112.8)**	874.1 (101.9)
Part D bid ($)[a]	68.2 (16.5)***	59.6 (20.8)
MA premium ($)[a]	2.0 (11.4)***	7.2 (24.8)
Part D premium ($)[a]	24.0 (18.0)***	12.7 (20.1)
MSB costs ($)[a]	117.5 (86.3)***	52.6 (32.8)
MA rebate ($)[a]	219.8 (90.1)***	179.5 (81.5)
Administrative costs ($)[a]	184.8 (66.9)***	155.1 (81.6)
Star Rating[a]	4.0 (0.6)***	3.8 (0.6)
Total enrollment[a]	7,242.7 (12,857.1)***	4,796.2 (10,390.1)

SOURCE: Authors' analysis of CMS and other data.
NOTE: MSB = mandatory supplemental benefit. ***, **, and * represent statistical significance at the 0.1%, 1%, and 5% levels, respectively.
[a] Means are provided, with standard errors in parentheses.

Table C.3 presents selected descriptive statistics for VBID participating plans from 2020 through 2024.

Table C.3. Descriptive Statistics for VBID-Participating Plans, 2020 to 2024

Characteristic	2020	2021	2022	2023	2024
Number of plans	144	376	859	1,218	1,400
PO size (enrollees)[a]	1,071,379 (979,047)	3,039,389 (1,116,647)	3,390,326 (1,784,005)	3,732,206 (2,254,301)	3,694,777 (2,434,932)
DSNPs (%)	27.8	38.0*	43.5	49.7**	53.9*
PO's geographic reach (%)					
Offered in 1–2 states	19.4	7.2*	13.4***	6.5***	7.5
Offered in 3–8 states	0	0.3	0.1	5.7***	4.3
Offered in 9+ states	80.6	92.6*	86.5*	87.8	88.2
Plans in Puerto Rico (%)	4.9	2.1	4.1	2.6	2.6
Offers UF (%)	4.9	2.9	10.8***	15.9**	10.8***
Offers SSBCI (%)	9.0	19.7**	28.2**	32.8*	22.5***
Offers non–primarily health-related supplemental benefits (%)	93.1	97.3	98.1	97.4	94.8**

SOURCE: Authors' analysis of CMS and other data.
NOTE: ***, **, and * represent statistical significance at the 0.1%, 1%, and 5% levels, respectively. Statistical differences are shown for the comparison of two adjacent years, and the asterisk or asterisks appear on the last year referenced. For example, differences between 2023 and 2024 are shown on the 2024 variables.
[a] Means are provided, with standard errors in parentheses.

Table C.4 presents selected descriptive statistics for VBID participating plans that newly entered the VBID model test in each year.

Table C.4. Descriptive Statistics for New VBID-Participating Plans, 2020 to 2024

Characteristic	2020	2021	2022	2023	2024
Number of plans	144	279	489	462	361
PO size (enrollees)[a]	1,071,379 (979,047)	3,301,270 (937,460)	2,793,575 (1,866,183)	2,886,272 (2,322,573)	2,759,812 (2,268,854)
DSNPs (%)	27.8	41.6	46.4	58.4	47.1
PO's geographic reach (%)					
Offered in 1–2 states	19.4	2.5	18.6	7.8	9.1
Offered in 3–8 states	0	0	0	9.1	11.1
Offered in 9+ states	80.6	97.5	81.4	83.1	79.8
Plans in Puerto Rico (%)	4.9	2.5	4.7	3.5	2.5

SOURCE: Authors' analysis of CMS and other data.
NOTE: All VBID-participating plans in 2020 were considered as new model participants.
[a] Means are provided, with standard errors in parentheses.

Table C.5 identifies POs that offered VBID benefits in each year.

Table C.5. VBID Participants, by Year

PO ID	2020	2021	2022	2023	2024
PO B	✓	✓	✓	✓	✓
PO C	✓	✓	—	✓	✓
PO E	—	—	✓	✓	✓
PO G	✓	✓	✓	✓	✓
PO H	—	—	—	—	✓
PO J	✓	✓	✓	—	—
PO L	✓	✓	✓	✓	✓
PO M	—	—	—	—	✓
PO N	✓	✓	✓	✓	✓
PO O	✓	✓	✓	✓	—
PO P	✓	✓	✓	✓	✓
PO Q	✓	✓	✓	✓	✓
PO R	—	✓	✓	✓	✓
PO S	—	✓	✓	✓	✓
PO T	—	—	✓	✓	✓
PO U	✓	✓	✓	✓	✓
PO Y	—	✓	✓	✓	–
PO AA	✓	—	✓	✓	✓
PO AB	✓	—	—	✓	✓
PO AC	—	—	✓	✓	✓
PO AD	—	—	✓	✓	✓
PO AE	—	—	✓	✓	✓
PO AF	—	—	✓	✓	✓
PO AG	—	—	✓	✓	✓
PO AH	✓	✓	✓	✓	✓
PO AK	—	—	✓	✓	✓
PO AL	—	—	✓	✓	✓
PO AO	—	—	✓	✓	✓
PO AP	—	—	✓	✓	✓
PO AQ	✓	—	✓	✓	✓
PO AR	—	—	✓	—	—
PO AS	—	—	—	✓	✓
PO AT	—	—	—	✓	✓
PO AU	—	—	—	✓	✓
PO AV	—	—	—	✓	✓

PO ID	2020	2021	2022	2023	2024
PO AW	—	—	—	✓	✓
PO AX	—	—	—	✓	—
PO AY	—	—	—	✓	✓
PO AZ	—	—	—	✓	✓
PO BB	—	—	—	✓	✓
PO BC	—	—	—	✓	✓
PO BD	—	—	—	✓	✓
PO BE	—	—	—	✓	✓
PO BF	—	—	—	✓	✓
PO BG	—	—	—	✓	✓
PO BH	—	—	—	✓	✓
PO BI	—	—	—	✓	✓
PO BJ	—	—	—	✓	✓
PO BK	—	—	—	✓	✓
PO BL	—	—	—	✓	✓
PO BM	—	—	—	—	✓
PO BN	—	—	—	—	✓
PO BO	—	—	—	—	✓
PO BP	—	—	—	—	✓
PO BQ	—	—	—	—	✓
PO BR	—	—	—	—	✓
PO BS	—	—	—	—	✓
PO BT	—	—	—	—	✓
PO BU	—	—	—	—	✓
PO BV	—	—	—	—	✓
PO BW	—	—	—	—	✓
PO BX	—	—	—	—	✓
PO BY	—	—	—	—	✓
PO BZ	—	—	—	—	✓
PO CA	—	—	—	—	✓
PO CD	—	—	—	—	✓
PO CE	—	—	—	—	✓
PO CF	—	—	—	—	✓

SOURCE: Authors' analysis of CMS and other data.

Table C.6 provides some additional information about VBID General interventions that each participating PO offered in 2024. It focuses specifically on beneficiary targeting and types of benefits POs offered as part of the model.

Table C.6. Summary of 2024 VBID General Interventions, by PO

PO ID	Chronic Condition Targeting	SES Targeting	VBID Flexibilities	VBID Flexibilities: Part C Reduced Cost Sharing	VBID Flexibilities: Part D Reduced Cost Sharing	VBID Flexibilities: Participation Requirements	VBID Flexibilities: Supplemental Benefits	RI
PO B	—	✓	✓	—	✓	—	✓	—
PO C	—	✓	✓	—	✓	—	✓	—
PO E	—	✓	✓	—	✓	—	✓	—
PO G	✓	✓	✓	✓	✓	✓	✓	✓
PO H	—	✓	✓	—	✓	—	✓	—
PO L	—	✓	✓	—	✓	—	✓	—
PO M	—	✓	✓	—	✓	—	—	—
PO N	✓	✓	✓	—	✓	—	✓	✓
PO P	✓	✓	✓	✓	✓	✓	✓	—
PO Q	—	✓	✓	✓	✓	✓	✓	—
PO R	—	✓	✓	—	✓	—	✓	—
PO S	—	✓	✓	—	✓	—	✓	—
PO T	—	✓	✓	—	✓	—	—	—
PO U	✓	—	—	—	—	—	—	✓
PO AA	—	✓	✓	—	✓	—	—	—
PO AB	—	✓	✓	—	✓	—	✓	—
PO AC	—	✓	✓	—	✓	—	✓	—
PO AE	✓	✓	✓	—	✓	—	✓	✓
PO AF	—	✓	✓	—	✓	—	✓	—
PO AG	—	✓	✓	—	✓	—	✓	✓
PO AH	✓	✓	✓	—	✓	—	✓	✓
PO AK	—	✓	✓	—	✓	—	✓	—

PO ID	Chronic Condition Targeting	SES Targeting	VBID Flexibilities	VBID Flexibilities: Part C Reduced Cost Sharing	VBID Flexibilities: Part D Reduced Cost Sharing	VBID Flexibilities: Participation Requirements	VBID Flexibilities: Supplemental Benefits	RI
PO AL	—	✓	✓	—	✓	—	✓	—
PO AO	✓	✓	✓	—	✓	—	✓	✓
PO AP	—	✓	✓	—	✓	—	✓	—
PO AQ	—	✓	✓	—	✓	—	✓	—
PO AS	✓	✓	✓	—	✓	—	—	✓
PO AT	—	✓	✓	—	✓	—	✓	—
PO AU	—	✓	✓	—	✓	—	✓	—
PO AV	—	✓	✓	—	✓	—	✓	—
PO AW	✓	✓	✓	—	✓	—	✓	—
PO AY	—	✓	✓	—	✓	—	✓	—
PO AZ	—	✓	✓	—	✓	—	—	—
PO BB	—	✓	✓	—	—	—	✓	—
PO BC	—	✓	✓	—	✓	—	✓	—
PO BD	—	✓	✓	—	✓	—	—	—
PO BE	—	✓	✓	—	✓	—	—	—
PO BF	—	✓	✓	—	✓	—	✓	—
PO BG	—	✓	✓	—	✓	✓	✓	—
PO BH	—	✓	✓	—	✓	—	✓	—
PO BI	—	✓	✓	—	✓	—	✓	—
PO BJ	—	✓	✓	—	✓	—	—	—
PO BK	—	✓	✓	—	✓	—	—	—
PO BL	—	✓	✓	—	✓	—	—	—
PO BM	—	✓	✓	—	✓	—	—	—
PO BN	—	✓	✓	—	✓	—	—	—

PO ID	Chronic Condition Targeting	SES Targeting	VBID Flexibilities	VBID Flexibilities: Part C Reduced Cost Sharing	VBID Flexibilities: Part D Reduced Cost Sharing	VBID Flexibilities: Participation Requirements	VBID Flexibilities: Supplemental Benefits	RI
PO BO	—	✓	✓	—	✓	—	—	—
PO BP	—	✓	✓	—	✓	—	✓	—
PO BQ	—	✓	✓	✓	—	✓	—	—
PO BR	—	✓	✓	—	✓	—	✓	—
PO BS	—	✓	✓	—	✓	—	✓	—
PO BT	✓	✓	✓	—	✓	—	—	✓
PO BU	—	✓	✓	—	✓	—	—	—
PO BV	—	✓	✓	—	✓	—	✓	—
PO BW	✓	✓	✓	—	✓	—	✓	✓
PO BX	—	✓	✓	—	✓	—	✓	—
PO BY	—	✓	✓	—	✓	—	—	—
PO BZ	—	✓	✓	—	—	—	✓	—
PO CA	—	✓	✓	—	✓	—	✓	—
PO CD	—	✓	✓	—	—	—	✓	—
PO CE	—	✓	✓	—	✓	—	—	—
PO CF	✓	✓	✓	—	✓	—	—	✓
PO AD	—	✓	✓	—	✓	—	✓	—

SOURCE: Authors' analysis of CMS and other data.
NOTE: SES = socioeconomic status.

Appendix D. Plan-Level Outcomes

This appendix provides detailed regression results for our analyses of plan-level outcomes, including all estimates referenced in Chapter 4 (“Estimates with VBID Plans Equally Weighted”). The appendix also contains estimates for plan-level outcomes in which VBID plans were weighted by enrollment instead of weighted equally (“Estimates with VBID Plans Weighted by Enrollment”). The enrollment-weighted estimates might be of interest because they reflect the average effect per beneficiary in the participant plans. For financial outcomes that are defined on a PMPM basis, such as bids or costs to CMS, the enrollment-weighted estimates could thus offer greater insight than the plan-weighted estimates into the aggregate effect of VBID on costs.

Estimates with VBID Plans Equally Weighted

Table D.1 shows the estimated associations between VBID and log plan enrollment. These estimates are used to derive the estimates presented in the main text.

Table D.1. Estimated Association Between Participation in VBID and Log Enrollment (Unweighted)

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	0.07	0.07	–0.08	0.21	0.403	313.88
2021	0.08	0.07	–0.04	0.22	0.212	869.59
2022	0.04	0.07	–0.09	0.17	0.580	1,539.69
2023	0.06	0.07	–0.08	0.20	0.388	1,871.86
2024	0.04	0.07	–0.10	0.18	0.539	2,105.77

SOURCE: Authors’ analysis of CMS and other data.
NOTE: CI = confidence interval.

Table D.2 shows the association between VBID and total MAPD bids. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.2. Estimated Association Between Participation in VBID and Total MAPD Bid, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–4.38	3.91	–12.18	3.31	0.249	305.88
2021	–3.51	2.71	–8.59	1.78	0.200	837.60
2022	–5.63	2.76	–11.13	–0.34	0.040	1,630.95
2023	–13.79	3.41	–20.34	–7.01	<0.001	1,871.86
2024	–8.98	3.46	–15.67	–1.77	0.009	2,119.43

SOURCE: Authors' analysis of CMS and other data.

Table D.3 shows the association between VBID and MA bids.

Table D.3. Estimated Association Between Participation in VBID and MA Bid, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–6.59	4.20	–14.64	1.48	0.110	309.36
2021	–12.95	2.67	–17.99	–7.70	<0.001	851.99
2022	–9.83	2.84	–15.38	–4.24	<0.001	1,623.74
2023	–15.78	3.61	–23.03	–8.67	<0.001	1,826.08
2024	–22.88	3.67	–29.76	–15.16	<0.001	2,117.93

SOURCE: Authors' analysis of CMS and other data.

Table D.4 shows the association between VBID and Part D bids.

Table D.4. Estimated Association Between Participation in VBID and Part D Bid, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	2.53	1.30	0.10	5.19	0.040	316.73
2021	8.31	0.79	6.81	9.96	<0.001	798.75
2022	4.23	0.95	2.19	6.06	<0.001	1,650.60
2023	1.70	0.88	–0.00	3.38	0.052	1,921.20
2024	14.03	1.15	11.76	16.29	<0.001	2,107.47

SOURCE: Authors' analysis of CMS and other data.

Table D.5 shows the association between VBID and total MAPD costs. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.5. Estimated Association Between Participation in VBID and Total MAPD Cost, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	5.23	10.31	–15.15	25.75	0.606	306.68
2021	23.35	9.11	6.03	41.42	0.012	876.61
2022	31.82	11.45	9.65	55.07	0.005	1,531.91
2023	36.10	16.59	0.65	66.07	0.044	1,866.82

SOURCE: Authors' analysis of CMS and other data.

Table D.6 shows the association between VBID and Part D costs. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.6. Estimated Association Between Participation in VBID and Part D Cost (Unweighted)

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–0.67	5.82	–12.68	10.48	0.942	306.68
2021	11.35	5.51	0.52	21.59	0.038	869.66
2022	10.15	7.10	–3.81	24.52	0.142	1,626.68
2023	36.80	8.12	21.28	52.88	<0.001	1,821.23

SOURCE: Authors' analysis of CMS and other data.

Table D.7 shows the association between VBID and specific Part D cost components. We present results for two of these components (low-income subsidy [LIS] and reinsurance) in the main text.

Table D.7. Estimated Association Between Participation in VBID and Part D Cost Components, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
Direct subsidy						
2020	1.35	0.70	–0.04	2.70	0.057	320.73
2021	2.02	0.49	1.10	2.95	<0.001	844.70
2022	0.02	0.93	–2.03	1.57	0.910	1,519.08
2023	–0.29	1.11	–2.22	2.21	0.717	1,829.13
LIS (low-income cost-sharing subsidy + low-income premium subsidy)						
2020	–1.99	2.15	–6.15	2.23	0.360	318.42
2021	3.13	1.86	–0.61	6.85	0.087	870.03
2022	6.02	2.30	1.48	10.35	0.003	1,538.58
2023	17.83	2.75	12.56	23.17	<0.001	1,860.88
Reinsurance						
2020	1.25	4.20	–7.17	9.15	0.761	304.23
2021	4.41	3.96	–3.43	11.94	0.253	865.96
2022	3.22	7.84	–11.75	17.91	0.702	1,630.27
2023	17.29	6.30	4.85	29.38	0.010	1,860.32

SOURCE: Authors' analysis of CMS and other data.

Table D.8 shows the association between VBID and MA costs. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.8. Estimated Association Between Participation in VBID and MA Cost (Unweighted)

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	1.55	7.72	–13.68	17.06	0.847	318.59
2021	13.61	6.36	1.18	26.18	0.031	868.55
2022	19.36	8.09	3.43	35.36	0.018	1,538.31
2023	–5.06	14.84	–36.85	19.81	0.792	1,834.13

SOURCE: Authors' analysis of CMS and other data.

Table D.9 shows the association between VBID and plan-level MA risk scores. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.9. Estimated Association Between Participation in VBID and MA Plan-Level MA Risk Score, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	0.007	0.010	–0.013	0.025	0.436	314.99
2021	0.014	0.008	–0.000	0.029	0.059	872.17
2022	0.023	0.009	0.006	0.040	0.008	1,535.77
2023	–0.000	0.015	–0.033	0.024	0.951	1,832.88

SOURCE: Authors' analysis of CMS and other data.

Table D.10 shows the association between VBID and MA rebates. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.10. Estimated Association Between Participation in VBID and MA Rebate, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	4.09	3.45	–2.17	11.08	0.231	303.23
2021	18.65	2.81	13.29	24.19	<0.001	836.34
2022	14.85	2.74	9.43	20.15	<0.001	1,595.95
2023	23.15	3.85	15.32	30.50	<0.001	1,861.39

SOURCE: Authors' analysis of CMS and other data.

Table D.11 shows the association between VBID and total MAPD premiums. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.11. Estimated Association Between Participation in VBID and Total MAPD Premium, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	0.04	0.87	–1.67	1.69	0.922	313.45
2021	1.44	0.97	–0.23	3.57	0.098	832.99
2022	0.59	0.66	–0.64	1.95	0.342	1,642.13
2023	0.35	0.65	–0.82	1.73	0.595	1,857.72
2024	2.56	0.85	1.01	4.27	<0.001	2,127.60

SOURCE: Authors' analysis of CMS and other data.

Table D.12 shows the association between VBID and MA premiums.

Table D.12. Estimated Association Between Participation in VBID and MA Premium, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–0.06	0.81	–1.69	1.51	0.987	314.92
2021	–1.88	0.66	–3.17	–0.56	0.005	857.48
2022	–0.54	0.57	–1.60	0.58	0.358	1,611.15
2023	–0.18	0.40	–0.93	0.62	0.652	1,928.64
2024	0.26	0.49	–0.66	1.24	0.567	2,194.06

SOURCE: Authors' analysis of CMS and other data.

Table D.13 shows the association between VBID and Part D premiums.

Table D.13. Estimated Association Between Participation in VBID and Part D Total Premium, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	0.46	0.49	–0.51	1.43	0.356	310.74
2021	2.65	0.54	1.65	3.66	<0.001	798.49
2022	1.21	0.47	0.30	2.16	0.009	1,551.07
2023	0.59	0.51	–0.34	1.65	0.231	1,848.12
2024	2.12	0.62	0.99	3.35	<0.001	2,163.38

SOURCE: Authors' analysis of CMS and other data.

Table D.14 shows the association between VBID and the number of mandatory supplemental benefits (MSBs) offered. These estimates are identical to those presented in the main text but contain additional information, such as standard errors.

Table D.14. Estimated Association Between Participation in VBID and Number of MSB Offerings, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–0.43	0.22	–0.87	0.00	0.052	309.31
2021	–1.64	0.19	–2.03	–1.28	<0.001	828.93
2022	–1.02	0.16	–1.34	–0.72	<0.001	1,670.95
2023	–0.91	0.19	–1.30	–0.56	<0.001	1,927.43
2024	–0.99	0.20	–1.37	–0.61	<0.001	2,123.33

SOURCE: Authors' analysis of CMS and other data.

Table D.15 shows the association between VBID and total MSB costs (PMPM).

Table D.15. Estimated Association Between Participation in VBID and PMPM MSB Cost, Unweighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	1.99	2.55	–2.68	7.21	0.454	308.08
2021	14.60	2.46	9.97	19.47	<0.001	842.83
2022	15.49	2.21	11.14	19.81	<0.001	1,607.35
2023	22.13	3.05	15.93	27.91	<0.001	1,871.39
2024	25.71	3.91	17.79	33.02	<0.001	2,095.59

SOURCE: Authors' analysis of CMS and other data.

Estimates with VBID Plans Weighted by Enrollment

Table D.16 shows the association between VBID and total MAPD bids, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.16. Estimated Association Between Participation in VBID and Total MAPD Bid. Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–10.18	6.37	–22.27	2.17	0.109	223.27
2021	–2.25	4.00	–10.10	5.64	0.539	414.22
2022	–8.01	3.67	–15.45	–0.96	0.025	639.51
2023	–21.91	5.58	–32.64	–10.94	<0.001	841.25
2024	–16.92	5.42	–27.13	–5.99	0.003	1,034.65

SOURCE: Authors' analysis of CMS and other data.

Table D.17 shows the association between VBID and total MAPD costs, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.17. Estimated Association Between Participation in VBID and Total MAPD Cost, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–12.69	14.53	–41.20	14.20	0.412	217.83
2021	41.99	16.14	12.54	74.20	0.007	417.94
2022	35.35	14.29	7.38	63.75	0.011	644.35
2023	51.65	19.56	12.52	90.42	0.009	1,045.86

SOURCE: Authors' analysis of CMS and other data.

Table D.18 shows the association between VBID and total MAPD premiums, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.18. Estimated Association Between Participation in VBID and Total MAPD Premium, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–0.40	0.95	–2.43	1.27	0.751	221.90
2021	3.01	1.78	0.20	7.04	0.036	405.29
2022	0.80	0.89	–0.91	2.51	0.340	667.07
2023	0.78	0.87	–0.83	2.56	0.361	1,041.28
2024	2.63	1.41	0.11	5.57	0.038	1,017.17

SOURCE: Authors' analysis of CMS and other data.

Table D.19 shows the association between VBID and MSB offerings, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.19. Estimated Association Between Participation in VBID and Number of MSB Offerings, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–0.19	0.43	–1.00	0.66	0.656	218.02
2021	–1.34	0.33	–2.03	–0.71	<0.001	410.25
2022	–0.67	0.21	–1.08	–0.28	<0.001	607.96
2023	–0.62	0.26	–1.18	–0.15	0.008	1,056.71
2024	–0.77	0.28	–1.34	–0.25	0.004	1,061.18

SOURCE: Authors' analysis of CMS and other data.

Table D.20 shows the association between VBID and MA bids, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.20. Estimated Association Between Participation in VBID and MA Bid, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–11.83	6.24	–23.47	0.63	0.062	224.69
2021	–10.53	4.00	–18.31	–2.43	0.009	416.83
2022	–12.99	3.75	–20.49	–5.57	<0.001	627.85
2023	–22.18	6.18	–34.26	–10.23	0.001	828.94
2024	–30.34	5.64	–40.76	–18.98	<0.001	1,047.28

SOURCE: Authors' analysis of CMS and other data.

Table D.21 shows the association between VBID and Part D bids, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.21. Estimated Association Between Participation in VBID and Part D Bid, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	1.85	1.73	–1.53	5.18	0.291	222.76
2021	7.37	1.07	5.27	9.49	<0.001	358.07
2022	5.07	1.23	2.67	7.49	<0.001	583.10
2023	0.23	1.52	–2.56	3.33	0.819	1,055.63
2024	13.84	1.65	10.56	16.89	<0.001	1,053.83

SOURCE: Authors' analysis of CMS and other data.

Table D.22 shows the association between VBID and MA costs, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.22. Estimated Association Between Participation in VBID and MA Cost, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–9.50	11.60	–31.95	13.43	0.414	222.24
2021	26.30	13.68	2.08	55.13	0.033	425.08
2022	19.40	9.76	0.19	37.89	0.049	645.89
2023	14.52	14.49	–15.17	41.33	0.351	1,046.75

SOURCE: Authors' analysis of CMS and other data.

Table D.23 shows the association between VBID and MA plan-level risk scores, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.23. Estimated Association Between Participation in VBID and MA Plan-Level Risk Score, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	0.002	0.011	–0.019	0.025	0.886	222.61
2021	0.022	0.016	–0.007	0.056	0.154	420.97
2022	0.019	0.010	–0.001	0.039	0.062	647.36
2023	0.018	0.014	–0.010	0.045	0.225	854.36

SOURCE: Authors' analysis of CMS and other data.

Table D.24 shows the association between VBID and Part D costs, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.24. Estimated Association Between Participation in VBID and Part D Cost, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	–8.14	9.02	–27.97	6.01	0.377	218.79
2021	13.01	6.50	0.03	25.62	0.050	419.18
2022	15.32	8.62	–1.48	31.89	0.081	641.58
2023	35.46	8.99	18.36	53.61	<0.001	933.23

SOURCE: Authors' analysis of CMS and other data.

Table D.25 shows the association between VBID and Part D cost components, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.25. Estimated Association Between Participation in VBID and Part D Cost Components, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	p-Value	ESS
Direct subsidy						
2020	0.23	0.92	–1.74	1.84	0.737	185.39
2021	2.35	0.76	0.79	3.76	0.002	502.89
2022	1.44	0.74	–0.06	2.81	0.062	669.73
2023	–0.48	1.18	–2.74	1.82	0.723	723.67
LIS (low-income cost-sharing subsidy + low-income premium subsidy)						
2020	–4.12	2.33	–8.63	0.37	0.075	185.39
2021	3.59	2.84	–1.94	9.16	0.211	502.89
2022	6.82	2.97	0.77	12.53	0.024	669.73
2023	19.56	3.46	12.97	26.22	<0.001	723.67
Reinsurance						
2020	10.56	8.78	–5.40	29.28	0.199	185.39
2021	6.94	5.68	–4.27	18.05	0.236	502.89
2022	10.05	8.20	–6.35	25.28	0.253	669.73
2023	28.90	11.51	8.79	53.82	0.003	723.67

SOURCE: Authors' analysis of CMS and other data.

Table D.26 shows the association between VBID and MA premiums, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.26. Estimated Association Between Participation in VBID and MA Premium, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	p-Value	ESS
2020	–0.68	1.10	–2.94	1.33	0.588	221.35
2021	–0.71	0.72	–2.16	0.66	0.312	402.56
2022	–0.01	0.54	–1.05	1.06	0.981	644.22
2023	0.34	0.39	–0.43	1.13	0.378	1,036.07
2024	1.14	0.67	–0.10	2.56	0.072	1,023.28

SOURCE: Authors' analysis of CMS and other data.

Table D.27 shows the association between VBID and MA rebates, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.27. Estimated Association Between Participation in VBID and MA Rebate, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	6.52	5.03	–2.25	17.25	0.172	222.13
2021	17.07	3.88	9.48	24.77	<0.001	392.53
2022	19.69	3.73	12.06	26.80	<0.001	602.15
2023	32.41	5.86	20.17	42.95	<0.001	851.42

SOURCE: Authors' analysis of CMS and other data.

Table D.28 shows the association between VBID and Part D premiums, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.28. Estimated Association Between Participation in VBID and Part D Total Premium, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	0.96	0.86	–0.70	2.70	0.270	219.74
2021	2.58	0.84	0.97	4.27	0.001	378.72
2022	0.47	0.66	–0.82	1.74	0.449	615.32
2023	0.29	0.66	–0.95	1.62	0.646	1,037.58
2024	1.14	0.97	–0.68	3.07	0.234	1,024.10

SOURCE: Authors' analysis of CMS and other data.

Table D.29 shows the association between VBID and MSB costs, with enrollment weighting (as opposed to the equally weighted approach used in the main text).

Table D.29. Estimated Association Between Participation in VBID and PMPM MSB Cost, Enrollment Weighted

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2020	3.24	3.78	–2.90	11.36	0.419	221.36
2021	14.07	3.40	7.41	20.67	<0.001	404.24
2022	17.05	3.26	10.65	23.55	<0.001	602.49
2023	31.02	4.81	21.24	39.97	<0.001	1,019.61
2024	31.78	4.96	21.96	41.57	<0.001	1,023.87

SOURCE: Authors' analysis of CMS and other data.

Appendix E. Beneficiary-Level Outcomes

Table E.1 contains the analytic results for the beneficiary-level outcomes presented in Chapter 5. These results are identical to those presented in the chapter but presented in tabular form with some additional information, including the standard error and the ESS for the analysis.

Table E.1. Estimated Associations Between Participation in VBID and Beneficiary-Level Outcomes, 2020–2022

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
Adherence to diabetes medication (ppt changes)						
2020	0.011	0.004	0.004	0.019	0.003	90,986
2021	0.005	0.002	0.001	0.010	0.023	654,098
2022	0.005	0.005	–0.005	0.015	0.321	888,337
Adherence to hypertension medication (ppt changes)						
2020	0.001	0.007	–0.013	0.014	0.904	170,521
2021	0.004	0.002	–0.001	0.008	0.092	1,263,610
2022	0.003	0.004	–0.006	0.012	0.463	1,990,034
Adherence to cholesterol medication (ppt changes)						
2020	0.013	0.004	0.005	0.021	0.001	228,852
2021	0.004	0.003	–0.000	0.009	0.074	1,274,305
2022	0.011	0.004	0.004	0.018	0.002	2,529,919
Adherence to breast cancer screening recommendations (ppt changes)						
2020	0.028	0.013	0.003	0.052	0.027	43,826
2021	0.017	0.010	–0.002	0.036	0.075	473,272
2022	–0.005	0.016	–0.035	0.026	0.770	680,170
Inpatient utilization						
2020	0.048	0.006	0.036	0.061	<0.001	326,831
2021	0.024	0.006	0.012	0.035	<0.001	2,693,001
2022	0.011	0.009	–0.005	0.028	0.181	3,822,599
Beneficiary OOP drug costs ($)						
2020	0.592	5.01	–9.23	10.41	0.906	326,831

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2021	–24.92	4.57	–33.87	–15.97	<0.001	2,693,001
2022	–35.86	3.17	–42.07	–29.65	<0.001	3,822,599
Targeted beneficiary risk scores						
2020	0.052	0.013	0.026	0.078	<0.001	326,831
2021	0.072	0.014	0.045	0.099	<0.001	2,693,001
2022	0.086	0.016	0.054	0.118	<0.001	3,822,599

SOURCE: Authors' analysis of CMS and other data.
NOTE: Analyses correspond to those reported in the main text. All estimates are from stratified, entropy-balanced DD regressions estimated separately for each year. Inpatient regressions were fitted using ordinary-least-squares models.

Appendix F. Contract-Level Outcomes

This appendix provides additional detail on the changes in Star Rating methodologies and contract-level Star Rating results from Chapter 6. Star Ratings are compiled and published each year by CMS to inform consumers about the quality of MA, MAPD, and stand-alone Part D plans and to determine MA rebates. Star Ratings published in a given display year reflect data from roughly two years prior. As we did for prior reports, we analyzed Star Ratings based on the year the quality information was measured (the measurement year), as opposed to the year in which it was displayed (the display year). For example, to analyze the association between VBID and Star Ratings in 2023, we used Star Rating data from measurement year 2023, which correspond to the 2025 display year. Because CMS made substantial changes to the Star Rating methodology for the 2020 measurement year to account for the coronavirus disease 2019 (COVID-19) pandemic, we excluded the 2020 measurement year from our analysis.

Detailed Regression Results

Overall Star Rating

Table F.1 shows the associations between VBID and overall Star Rating for 2021–2023 that are also presented in the main project report.

Table F.1. Estimated Association Between Participation in VBID and Contract-Level Overall Star Rating

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
2021	0.25	0.08	0.09	0.41	0.002	152
2022	0.14	0.06	0.03	0.27	0.021	258
2023	–0.02	0.07	–0.16	0.10	0.815	459

SOURCE: Authors' analysis of CMS data.

Domain-Level Star Ratings

In Table F.2, we report associations between VBID and domain-level Star Ratings that are presented in the main text.

Table F.2. Estimated Association Between Participation in VBID and Domain-Level Star Ratings

Effect and Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
Part C domains						
Staying Healthy						
2021	0.08	0.07	–0.06	0.22	0.337	151
2022	–0.17	0.16	–0.53	0.10	0.265	260
2023	–0.22	0.08	–0.36	–0.07	0.003	455
Managing Chronic Conditions						
2021	0.23	0.08	0.07	0.38	0.005	156
2022	0.15	0.09	–0.02	0.33	0.098	280
2023	0.03	0.08	–0.12	0.19	0.665	444
Member Experience with Health Plan						
2021	0.23	0.12	0.00	0.47	0.047	182
2022	0.29	0.13	0.03	0.55	0.032	257
2023	0.19	0.11	–0.02	0.41	0.081	455
Member Complaints and Changes in the Health Plan's Performance						
2021	0.04	0.12	–0.20	0.29	0.731	151
2022	–0.05	0.11	–0.28	0.17	0.650	257
2023	0.08	0.09	–0.10	0.26	0.382	449
Health Plan Customer Service						
2021	0.31	0.11	0.10	0.53	<0.001	153
2022	–0.02	0.11	–0.24	0.20	0.897	260
2023	–0.04	0.08	–0.19	0.14	0.680	462
Part D domains						
Member Experience with the Drug Plan						
2021	0.45	0.15	0.20	0.77	0.002	150
2022	0.53	0.16	0.22	0.84	<0.001	257
2023	0.06	0.12	–0.18	0.30	0.654	448
Drug Safety and Accuracy of Drug Pricing						
2021	–0.01	0.09	–0.17	0.16	0.907	146
2022	–0.05	0.08	–0.22	0.11	0.551	254
2023	–0.20	0.09	–0.38	–0.04	0.012	461

SOURCE: Authors' analysis of CMS data.

Select Results Star by Level of Contract Exposure to VBID

For Table F.3, we changed the definition of *VBID participation* to focus on contracts in which a minimum share of beneficiaries (25%, 50%, or 75%) were exposed to VBID. We consider a beneficiary to be exposed to VBID if the beneficiary was enrolled in a VBID-participating plan, irrespective of whether the beneficiary was targeted for VBID or received VBID benefits. Regardless of how we define *contract-level VBID exposure*, the association between VBID and the overall Star Rating was not statistically significant in 2023, while the associations with the Staying Healthy and Drug Safety and Accuracy of Drug Pricing domains were negative and statistically significant. The negative association between VBID and Drug Safety and Accuracy of Drug Pricing increased somewhat as the share of beneficiaries exposed to VBID increased.

Table F.3. Estimated Association Between Participation in VBID and Overall Star Rating, by Levels of VBID Exposure Within the Participating Contracts in 2023

VBID Exposure	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*–value	ESS
Overall Star Rating						
25%	–0.06	0.08	–0.21	0.09	0.489	289
50%	–0.08	0.08	–0.24	0.08	0.331	224
75%	–0.13	0.09	–0.31	0.05	0.141	177
Staying Healthy						
25%	–0.27	0.09	–0.44	–0.10	0.001	291
50%	–0.22	0.10	–0.41	–0.02	0.020	220
75%	–0.22	0.11	–0.44	0.00	0.057	178
Drug Safety and Accuracy of Drug Pricing						
25%	–0.23	0.09	–0.42	–0.06	0.011	295
50%	–0.25	0.10	–0.45	–0.06	0.011	220
75%	–0.35	0.11	–0.56	–0.14	0.001	167

SOURCE: Authors' analysis of CMS data.

To further analyze the unexpected Star Rating results for 2023, we stratified the analysis to separately assess contracts with at least one Dual Eligible Special Needs Plan (DSNP) and contracts without any DSNPs. We thought that this distinction might be important given that quality outcomes might be systematically different in DSNPs and non-DSNPs. Table F.4 shows that—for both types of contracts—the association with the overall Star Rating was not statistically significant in 2023. Furthermore, the negative associations with the Staying Healthy and Drug Safety and Accuracy of Drug Pricing domains were present for both types of contracts.

Table F.4. Estimated Association Between Participation in VBID and Selected Star Rating Measures, by DSNP Status, 2023

Measure and DSNP Status	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-Value	ESS
Overall Star Rating						
With DSNPs	–0.10	0.07	–0.24	0.02	0.127	347
Without DSNPs	0.02	0.16	–0.26	0.36	0.973	57
Staying Healthy						
With DSNPs	–0.21	0.09	–0.37	–0.03	0.015	375
Without DSNPs	–0.36	0.21	–0.75	0.05	0.076	57
Drug Safety and Accuracy of Drug Pricing						
With DSNPs	–0.28	0.09	–0.46	–0.11	0.001	385
Without DSNPs	–0.35	0.14	–0.60	–0.08	0.016	59

SOURCE: Authors' analysis of CMS data.
NOTE: Contracts with DSNPs contain at least one DSNP, while contracts without DSNPs have no DSNPs.

References

Austin, Peter 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, September 15, 2009.

Callaway, Brantly, and Pedro H. C. Sant'Anna, "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, Vol. 225, No. 2, December 2021.

Cohen, Jacob, *Statistical Power Analysis for the Behavioral Sciences*, Academic Press, 1977.

de Chaisemartin, Clément, and Xavier d'Haultfoeuille, "Difference-in-Differences Estimators of Intertemporal Treatment Effects," *Review of Economics and Statistics*, February 9, 2024.

Eibner, Christine, Dmitry Khodyakov, Erin Audrey Taylor, Denis Agniel, Rebecca Anhang Price, Julia Bandini, Marika Booth, Lane F. Burgette, Christine Buttorff, Catherine C. Cohen, Stephanie Dellva, Michael Dworsky, Natalie Ernecoff, Priya Gandhi, Alice Y. Kim, Julie Lai, Monique Martineau, Nabeel Qureshi, Jessica Randazzo, Afshin Rastegar, Lucy B. Schulson, Daniel Schwam, Joan M. Teno, Anagha Alka Tolpadi, Asa Wilks, and Shiyuan Zhang, *Evaluation of Phase II of the Medicare Advantage Value-Based Insurance Design Model Test: First Three Years of Implementation (2020–2022)*, Centers for Medicare & Medicaid Services, September 2023a.

Eibner, Christine, Dmitry Khodyakov, Erin Audrey Taylor, Denis Agniel, Rebecca Anhang Price, Julia Bandini, Marika Booth, Lane F. Burgette, Christine Buttorff, Catherine C. Cohen, Stephanie Dellva, Michael Dworsky, Natalie C. Ernecoff, Priya Gandhi, Alice Y. Kim, Julie Lai, Monique Martineau, Nabeel Qureshi, Jessica Randazzo, Afshin Rastegar, Lucy Schulson, Daniel Schwam, Joan M. Teno, Anagha Alka Tolpadi, Asa Wilks, and Shiyuan Zhang, *Appendices: Evaluation of Phase II of the Medicare Advantage Value-Based Insurance Design Model Test: First Three Years of Implementation (2020–2022)*, September 2023b.

Eibner, Christine, Dmitry Khodyakov, Erin A. Taylor, Denis Agniel, Rebecca Anhang Price, Julia Bandini, Marika Booth, Lane F. Burgette, Christine Buttorff, Catherine C. Cohen, Stephanie Dellva, Michael Dworsky, Natalie C. Ernecoff, Alice Y. Kim, Julie Lai, Monique Martineau, Nabeel Qureshi, Afshin Rastegar, Max Rubinstein, Daniel Schwam, Joan M. Teno, Anagha Tolpadi, and Shiyuan Zhang, *Evaluation of the Medicare Advantage Value-Based Insurance Design Model Test: 2020 to 2023*, Centers for Medicare & Medicaid Services, March 2025.

Goodman-Bacon, Andrew, “Difference-in-Differences with Variation in Treatment Timing,” *Journal of Econometrics*, Vol. 225, No. 2, December 2021.

Guest, Greg, Kathleen M. MacQueen, and Emily E. Namey, *Applied Thematic Analysis*, SAGE Publications, 2012.

Khodyakov, Dmitry, Christine Eibner, Erin Audrey Taylor, Rebecca Anhang Price, Christine Buttorff, Matthew Cefalu, Brian G. Vegetabile, Julia Bandini, Monique Martineau, Catherine C. Cohen, Michael Dworsky, Marika Booth, Alice Y. Kim, Julie Lai, Shiyuan Zhang, Afshin Rastegar, Stephanie Dellva, Nabeel Qureshi, Priya Gandhi, Courtney Armstrong, Daniel Schwam, Natalie Ernecoff, and Anagha Alka Tolpadi, *Evaluation of Phase II of the Medicare Advantage Value-Based Insurance Design Model Test: First Two Years of Implementation (2020–2021)*, Centers for Medicare & Medicaid Services, October 2022.

Kish, Leslie, *Survey Sampling*, Wiley, 1965.

McHugh, Mary L., “Interrater Reliability: The Kappa Statistic,” *Biochemica Medica (Zagreb)*, Vol. 22, No. 3, 2012.

Public Law 74-271, Social Security Act, August 4, 1935.

Robbins, Michael W., “The Utility of Nonparametric Transformations for Imputation of Survey Data,” *Journal of Official Statistics*, Vol. 30, No. 4, December 2014.

Robbins, Michael W., “Joint Imputation of General Data,” *Journal of Survey Statistics and Methodology*, Vol. 12, No. 1, February 2024.

Robbins, Michael W., Sujit K. Ghosh, and Joshua D. Habiger, “Imputation in High-Dimensional Economic Data as Applied to the Agricultural Resource Management Survey,” *Journal of the American Statistical Association*, Vol. 108, No. 501, 2013.

Rubin, Donald B., “Causal Inference Using Potential Outcomes: Design, Modeling, Decisions,” *Journal of the American Statistical Association*, Vol. 100, No. 469, 2005.

Stuart, Elizabeth A., Brian K. Lee, and Finbarr P. Leacy, “Prognostic Score-Based Balance Measures Can Be a Useful Diagnostic for Propensity Score Methods in Comparative Effectiveness Research,” *Journal of Clinical Epidemiology*, Vol. 66, No. 8 Suppl., August 2013.

U.S. Code, Title 42, The Public Health and Welfare; Chapter 7, Social Security; Subchapter XVIII, Health Insurance for Aged and Disabled; Part E, Miscellaneous Provisions; Section 1395mm, Payments to Health Maintenance Organizations and Competitive Medical Plans.

Wang, Yixin, and Jose R. Zubizarreta, “Minimal Dispersion Approximately Balancing Weights: Asymptotic Properties and Practical Considerations,” *Biometrika*, Vol. 107, No. 1, 2020.