Project Report

Appendices: Evaluation of the Medicare Advantage Value-Based Insurance Design Model Test: Implementation Years 2020 Through 2023

Christine Eibner, 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, Alice Y. Kim, Julie Lai, Monique Martineau, Nabeel Qureshi, Afshin Rastegar, Max Rubinstein, Daniel Schwam, Joan M. Teno, Anagha Tolpadi, Shiyuan Zhang

RAND Health Care

PR-A1881-3
March 2025
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

About These Appendices

These appendices supplement the Evaluation of Phase II of the Medicare Advantage (MA) Value-Based Insurance Design (VBID) Model test, initiated by the Center for Medicare & Medicaid Innovation (Innovation Center), for the years 2020 through 2023. Because the analyses are similar year over year, these appendices contain some of the same descriptions as those presented in the 2023 Evaluation Report (Eibner et al., 2023a).

The evaluation was funded by the Innovation Center under Research, Measurement, Assessment, Design, and Analysis Contract Number 75FCMC19D0093, Order Number 75FCMC20F0001, for which Julia Driessen is the contracting officer's representative. It was carried out within the Payment, Cost, and Coverage Program in RAND Health Care.

For more information, see www.rand.org/health-care, or contact

RAND Health Care Communications
1776 Main Street
P.O. Box 2138 Santa Monica, CA 90407-2138
(310) 393-0411, ext. 7775
RAND_Health-Care@rand.org

Abbreviations

ASMD	absolute standardized mean difference
ATT	average treatment effect on the treated
BCBS	Blue Cross Blue Shield
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CCN	CMS Certification Number
CHF	congestive heart failure
CI	confidence interval
CKD	chronic kidney disease
CMS	Centers for Medicare & Medicaid Services
COPD	chronic obstructive pulmonary disease
COVID-19	coronavirus disease 2019
C-SNP	chronic condition special needs plan
DD	difference-in-differences
DM	disease management
DSNP	dual eligible special needs plan
EB	entropy balancing
ED	emergency department
ESRD	end-stage renal disease
ESS	effective sample size
FFS	fee-for-service
HCC	Hierarchical Condition Category
HIV/AIDS	human immunodeficiency virus/acquired immunodeficiency syndrome
HPSA	Health Professional Shortage Area
ICH	intracerebral hemorrhage
I-SNP	institutional special needs plan
LICS	Low-Income Cost-Sharing Subsidy
LIPS	Low-Income Premium Subsidy
LIS	Low-Income Subsidy
LOS	length of stay
MA	Medicare Advantage
MAPD	Medicare Advantage Prescription Drug
MI	myocardial infarction
MICE	Multiple Imputation by Chained Equations
MSB	mandatory supplemental benefits

MTM	Medication Therapy Management
NPHRB	non–primarily health-related supplemental benefit
OLS	ordinary least squares
OON	out of network
OOP	out of pocket
PBP	plan benefit package
PDP	Part D plan
PDSS	Part D Senior Savings
PHRSB	Primarily Health-Related Supplemental Benefits
PMPM	per member per month
PO	parent organization
PPO	preferred provider organization
RA	rheumatoid arthritis
RI	Rewards and Incentives
RxHCC	Prescription Drug Hierarchical Condition Category
SDI	Social Deprivation Index
SES	socioeconomic status
SMD	standardized mean difference
SNP	Special Needs Plan
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: VBID General

The main report includes quantitative analyses intended to estimate changes in outcomes associated with Value-Based Insurance Design (VBID) General at three levels of analysis: the plan level, the beneficiary level, and the contract level. As in our 2022 and 2023 Evaluation Reports, all these analyses use a difference-in-differences (DD) design to identify the effects of VBID General. For the plan- and contract-level analyses, our methods closely follow those used in our 2022 and 2023 Evaluation Reports (Khodyakov et al., 2022; Eibner et al., 2023b), with some refinements to the statistical methods described below. For the beneficiary-level analyses, we made more substantial revisions to the methods from our previous reports that were necessitated by the longer time horizon of the extended Model Test, but we continued to use similar statistical models within a DD framework.

In this appendix, we describe these methods. We draw heavily on descriptions provided in the appendices to our 2023 Evaluation Report (Eibner et al., 2023a), as the basic approach is similar to the approach described there.

Overview of Statistical Methods

As described in the main report, VBID General encompasses a variety of intervention types, such as reduced cost sharing for high-value services, reduced cost sharing for high-value drugs, VBID supplemental benefits, and Rewards and Incentives (RI). Our primary analyses do not differentiate these different subcomponents, but we also include select analyses of the individual subcomponents in Appendix D.

There are a few challenges to successfully estimating the average impact of VBID participation, given the observational nature of the VBID Model test. First, plans' fidelity of implementation and beneficiaries' uptake of the proposed intervention may 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 proposed interventions, regardless of fidelity or uptake. This allows us to estimate the effectiveness of VBID participation under real-world implementation of the interventions. We do not estimate the efficacy of the interventions, which would measure the effect of VBID participation under the ideal circumstances of perfect fidelity and uptake. Second, plans were allowed to join and leave the VBID Model test on a year-to-year basis, leading 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 these analytic concerns, this evaluation combines entropy balancing (EB) on observables with the difference-in-differences (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 test, and addresses concerns that staggered adoption of the treatment (in this case, VBID) could lead to a biased result if associations between VBID and key outcomes varied over time (Goodman-Bacon, 2021; de Chaisemartin and De Haultfoeuille, 2020). EB serves to bolster the DD design, which allows for differences (in both observable and unobservable characteristics) between VBID and comparison plans under certain assumptions.

This analytic strategy can be summarized into 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 step, we group plans into "participation patterns" defined based on each plan's history of participation in VBID General (or, for subgroup analyses, each plan's history of implementing specific types of VBID General interventions). In the second step, we identify eligible nonparticipant plans that can serve as 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 observable characteristics that might predict differences in how outcomes of interest will evolve over time.

In the third step, we therefore construct a set of weights chosen to ensure that the covariate distribution of the comparison group (including both baseline characteristics and trends in outcomes prior to VBID implementation) approximately equals the covariate distribution of the VBID participating group.

Finally, in the fourth step, we estimate a weighted two-way fixed effects regression model for each outcome and participation pattern, using the weights constructed in the third step to make the weighted comparison group as similar as possible to the VBID participating group. Estimation of separate models for each outcome and participation pattern is critical to ensure that two-way fixed effects 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 adopted the model test at different points in time, a situation (often referred to as "staggered adoption") that requires some care to obtain unbiased estimates if the changes in outcomes associated with VBID (which we refer to as "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 combine 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, or 2023). 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 number of outcomes and years considered in our evaluation. The policy question answered by each year's estimate—questions of the form "What was the average effect of all the VBID interventions implemented in 2023?"—is likely of interest to the Centers for Medicare & Medicaid Services (CMS) because it 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 due both to rapid growth in the number of participating plans and 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. The average effects by calendar year that we report therefore 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 test (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. To provide greater insight into the effects of specific interventions or targeting approaches, we conducted subgroup analyses for selected key outcomes, which are reported in Appendix D.

In the sections below, 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 general methods are introduced focusing on analyses that use the plan as the unit of analysis. 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 below. 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.

Defining Groups of Participating Plans

As in our 2022 and 2023 Evaluation Reports, we limited our analyses to Medicare Advantage Prescription Drug (MAPD) plans because very few Medicare Advantage (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 test, we do not attempt 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 Test, which began in 2020. Beginning with 2017 (the earliest year of pre-VBID data used in our analyses), a plan might be observed annually for up to seven years (2017 through 2023), and VBID participation could begin in 2020 or a later year. 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). The history of a plan's participation in VBID General in 2020 and later years can then be encoded as vector (denoted by $\boldsymbol{a}$) of the four indicator variables $\boldsymbol{a} = (a_{2020}, a_{2021}, a_{2022}, a_{2023})$. Table A.1 lists the participation patterns used in our plan-level analyses, as well as participation patterns that were excluded.

Table A.1. Number of Plans Participating in VBID General by Year and Participation History

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	240	0	240
	11	97	7	90
	Total	**337**	**7**	**330**
2022, included	001	318	0	318
	011	252	0	252
	101	8	1	7
	111	87	4	83
	NA01	68	0	68
	Total	**733**	**5**	**728**
2023, included	0001	141	0	141
	0011	279	0	279
	0111	236	0	236
	1111	78	3	75
	NANA01	65	0	65
	NA001	38	0	38
	NA011	77	0	77
	Total	**913**	**3**	**911**
2021, excluded	NA1	17	17	0
	Total	**17**	**17**	**0**
2022, excluded	NA11	17	17	0
	NANA1	80	80	0
	Total	**97**	**97**	**0**
2023, excluded	1001	4	0	4
	1011	4	1	3
	NANANA1	211	211	0
	NANA11	61	61	0
	NA111	16	16	0
	Total	**296**	**289**	**7**

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. Plans that did not exist in a year prior to an outcome year of interest have "NA" indicated in the participation pattern. For example, "NANA01" in outcome year 2023 indicates plans that did not exist in 2020 or 2021, then were offered without participating in VBID in 2022, and then participated in VBID in 2023. "Pattern Included in Analysis" indicates whether a pattern was used 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 re-entered VBID, but had too few participating plans to be appropriate for our estimation methods.

Plans that entered VBID in their first year of existence were excluded since it is not possible to distinguish the effects of VBID from other factors affecting outcomes if a plan never existed without VBID. Exclusion of plans entering VBID in their first year of existence reduces the sample of plans included in our estimates by 7 plans in 2020, by 17 plans in 2021, by 97 plans in 2022, and by 288 plans in 2023. We also exclude two small participation patterns (N = 4 plans for pattern 1001 and N = 3 plans for pattern 1011) containing plans that participated in 2020, exited in 2021, and later re-entered the model test.

To evaluate changes in outcomes associated with VBID General implementation, we sought to estimate the average treatment effect on the treated (ATT). Following the potential outcomes framework proposed by Rubin (2005) and 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 where the plans implementing VBID had not implemented VBID) among the plans that implemented VBID. The ATT can be defined at different points in time, since—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 General: $Y_{it}(\boldsymbol{a})$ 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 General.

We can then define the ATT for each VBID General 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 below), these $ATT(\boldsymbol{a}, t)$ can be estimated using comparisons between a group of participating plans defined by the VBID General participation pattern $\boldsymbol{A} = \boldsymbol{a}$ and a group with $\boldsymbol{A} = \boldsymbol{0}$.

Once we have estimated the $ATT(\boldsymbol{a}, t)$ for each participation pattern $\boldsymbol{a}$ and year t, we aggregate the estimates to obtain an average change in outcomes for each calendar year of the VBID Model test from 2020 through 2023. That is, for each year *t* = 2020, 2021, 2022, or 2023, we define the ATT of VBID General 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, 75 of the 911 VBID-General-participating plans had participated continuously since 2020, so the weight $w(\boldsymbol{1111}, 2023)$ was defined to be (0.082 = 75 / 911).

Identifying Nonparticipating Plans Eligible for VBID

Our prior reports have documented in detail the plan-level eligibility criteria for the VBID model test, and thus we summarize the process here. Key eligibility criteria include limiting to specific MA plans 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 3-star rating or higher and not under sanction).

We make several other exclusions to the pool of eligible comparison plans including the end-stage renal disease (ESRD) chronic condition special needs plans (C-SNPs) due to their very different patient populations, plans that transitioned from 1876 Cost plans (since these plans are not eligible for the model) and Part B only plans (since they are missing key outcome data). In a given calendar year, new or discontinued participating VBID plans contributed data for descriptive analyses, but only contributed data for DD analyses if they have at least one year of pre and post data for the particular model year.

Entropy Balancing 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 to improve comparability between the groups and justify the key assumptions of our DD regression models.

As described in our 2023 Evaluation Report, we used EB to derive weights for use in our regression estimates. The weights increase comparability on observables 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 pre-intervention characteristics between VBID participants and the weighted comparison group to be small. For a particular covariate Z, the SMD is defined as the mean in the treated group minus the weighted mean in the control group, all divided by the standard deviation in the treated group. In other words, a SMD of 0 means 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 would indicate that the difference in means is equal to 0.1 standard deviations. A rule of thumb is that a SMD below 0.2 indicates acceptable balance between treatment and comparison groups, while a SMD above 0.2 indicates unacceptable balance (Cohen, 1977).

As in Khodyakov et al. (2022), we modified the standard EB algorithm to produce weights that balance the covariates within a pre-specified range (or "tolerance") of SMDs. For example, we can estimate weights that 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), which 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 may 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 this comes at the cost of balance between the groups. In practice, SMD values lower than δ = 0.1 are customarily used (Austin, 2009; Stuart, Lee, and Leacy, 2013) when the goal of balancing is to fully control for confounding from observable characteristics. Because our empirical strategy also uses DD (which does not require balance on baseline characteristics to deliver unbiased estimates), we have emphasized 0.2 as a threshold value for the SMD between VBID and comparison plans after weighting.

As noted above, EB weights were derived separately for each participation pattern and outcome. To allow a unified approach to estimation for 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. The smallest tolerance for which the ESS of the comparison group met this threshold was used to define the weights.

Variables Included in Entropy Balancing

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 2023 Evaluation Report and balanced simultaneously on two groups of variables:

1. Characteristics of VBID participants and comparison plans observed prior to VBID plans' participation in the VBID Model test ("baseline characteristics")
2. Trends in the outcome variable observed prior to the first year of VBID implementation ("pre-VBID outcome trends")

A comprehensive set of baseline characteristics was used, including beneficiary demographics, plan characteristics, and characteristics of the local health care market. The set of baseline characteristics included varied depending on the level of analysis (plan-, beneficiary-, or contract-level), but the same set of baseline characteristics was used for all analyses at a given level of analysis: detail on the included variables and levels of balance achieved for each level of analysis (plan-, beneficiary-, and contract-level) are presented below.

The pre-VBID outcome trends 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. As an illustration, the EB weights for our analysis of an outcome variable Y_{it} (for example, MAPD bids) among participation pattern $\boldsymbol{a} = \mathbf{1111}$, which denotes

plans that implemented VBID General in 2020 and continued through 2023, 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 ($Y_{i2019} - Y_{i2017}$).

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 test—for at least some observations, yet EB and subsequent steps in our analysis require a data set with no missing data.

We therefore imputed missing covariate and outcome information jointly using a Markov Chain Monte Carlo method known as GERBIL (Robbins, 2024). GERBIL has some theoretical and computational advantages over older alternatives such as Multiple Imputation by Chained Equations (MICE): These advantages are important given the large scale of the data used in these analyses. This approach also uniformly improves on imputation approaches used in previous reports by using more information.

This imputation strategy further ensures that all plans contribute to all analyses. In previous reports, plans with missing outcome data in the period after VBID implementation were not included in analyses. As the VBID Model test has extended over time, the number of plans that had missing data in any post-period year has grown. Thus, an approach that keeps all such plans in the analysis was required.

Difference-in-Differences

To identify the $ATT(\boldsymbol{a}, t)$ using observed data, we use a DD design. We estimate these models using the EB weights described above, using weighted least squares for the contract- and plan-level models, and using comparable weighted estimators (for example, generalized linear models [GLMs]) in some 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 simultaneously affect outcomes across all plans during the post-intervention period. Specifically, 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 post-intervention period and 0 otherwise). For each participation pattern, we estimate a weighted DD models of the form

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

for each outcome year where 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 zero 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, a separate DD models was fit for each of the participation patterns ($\boldsymbol{a}$), so the β_t are estimates of $ATT(\boldsymbol{a}, t)$ for those groups of plans. Unless otherwise noted, potential confounders were included in the balancing weights; hence, we do not include any time-varying controls in Equation (A.3). The β_t for each group of participating plans are then aggregated to obtain calendar year effects as described above. Variance estimates were derived 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 trends" assumption to identify causal effects. This assumption states that the post-participation trend in the outcomes for the comparison group is equal to the trend for each VBID-participation pattern had they not participated in VBID. To bolster the plausibility of this assumption, we assume that parallel trends hold within levels of observed variables $\boldsymbol{X_{it}}$ after applying the EB weights. We write this assumption formally in Equation A.4:

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

where t^* is some time period post VBID implementation, and t is some time period prior to VBID implementation, and the expectation on the right-hand side (for $\boldsymbol{A} = \boldsymbol{0}$) is taken with respect 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 the treatment group (which is inherently unobservable) can be expressed in terms of the pre-treatment outcomes among the treated group plus the observed trend in the comparison group.

The DD methodology does not require that the balancing characteristics are perfectly balanced or that they are sufficient to control for confounding, as long as the parallel trends assumption described in Equation A.4 holds. Rather, a DD model works by assuming that the post-participation 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 compares the change in the pre-participation outcome with the post-participation outcome between participating and comparison plans.

In Tables A.1, A.3, and A.4, we summarize balance on preparticipation outcome trends by reporting the average SMD for each outcome pre-trend variable that was used in balancing. The tables show that weighting was generally able to achieve average SMDs in pre-participation outcome trends below 0.1 for each level of analysis. Readers interested in figures that illustrate the impact of balancing on pre-participation trends for selected outcomes should consult Appendix C of our 2023 evaluation report (Eibner et al., 2023a).

Inference

We use a smooth version of the bootstrap that accounts for dependencies across time and within plans by generating two hundred sets of plan-level weights, where the weights are generated from the beta distribution. Holding the balancing weights fixed, we multiply the balancing weights by the bootstrap weights and recompute the two-way fixed effect estimates for each new set of weights and conducting the relevant aggregation. We then calculate the empirical standard deviation of the bootstrap using a normal approximation to generate confidence across estimates intervals.

Additional Estimation Details

In the subsections below, we provide details about methods that were specific to analyses at the three different levels of analysis in this evaluation: plan-level, beneficiary-level, and contract-level, emphasizing:

1. Details of balancing variables
2. Results of balancing
3. Departures from the overall methodology described above
4. Differences from last year's report specific to the level of analysis

Plan-Level Analyses

Table A.2 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. Appendix B contains further information about these variables, including their sources.

Table A.2 shows that EB succeeded in reducing the SMD between VBID and comparison groups below 0.2 for all balancing variables shown. Pre-participation outcome trends had an SMD of 0.03.

Table A.2. Selected Balancing Variables Included in Plan-Level Analyses

Variable	Unweighted ASMD	Weighted ASMD
Age	0.59	0.06
Star Rating (overall)	0.21	0.08
COVID-19 cases per 10,000	0.09	0.02
Percentage disabled	0.75	0.08
Percentage dual eligible	0.59	0.07
For-profit (beneficiary months)	0.22	0.02
For-profit (enrollment)	0.22	0.02
Percentage cancer	0.15	0.04
Percentage CHF	0.30	0.02
Percentage COPD	0.55	0.08
Percentage diabetes	0.44	0.08
Hospice participant 2021	0.02	0.01
Hospice participant 2022	0.02	0.01
Hospice participant 2023	0.03	0.02
HPSA	0.01	0.00
Newly transitioned into bonus	0.02	0.01
Percentage LIS status	0.57	0.06
Sex	0.31	0.05
Area-level income	0.28	0.04
Missing outcomes	0.06	0.01
Newly transitioned out of bonus	0.03	0.01
Part D basic premium	0.19	0.02
PDSS participant 2021	0.02	0.02
PDSS participant 2022	0.04	0.05
PDSS participant 2023	0.01	0.06
Part D supplemental premium	0.25	0.02
Part D total premium	0.18	0.02
Offered SSBCI 2020	0.02	0.01
Offered SSBCI 2021	0.09	0.04
Offered SSBCI 2022	0.11	0.02
Offered SSBCI 2023	0.12	0.07
Offered UF 2020	0.03	0.00
Offered UF 2021	0.08	0.01
Offered UF 2022	0.02	0.02
Offered UF 2023	0.08	0.06
Part C in-network OOP maximum	0.37	0.05
Urban	0.10	0.02
Rural	0.06	0.03
Suburban	0.10	0.03
Percentage over age 65	0.13	0.04
MA penetration	0.07	0.03
Plan type (PPO = 1; otherwise = 0)	0.06	0.02
Average MA risk score (HCC)	0.39	0.05
Average Part D risk score (RxHCC)	0.60	0.05
Puerto Rico county	0.03	0.00
C-SNP	0.03	0.01

Variable	Unweighted ASMD	Weighted ASMD
DSNP	0.29	0.04
I-SNP	0.04	0.01
Standardized Medicare costs per capita	0.11	0.04
Pre-Participation Outcome Trends	0.13	0.02

SOURCE: RAND analysis of VBID-participating plan and other data.
NOTE: ASMD = absolute standardized mean difference; CHF = congestive heart failure; COPD = chronic obstructive pulmonary disease; COVID-19 = coronavirus disease 2019; DSNP = dual eligible special needs plan; HCC = Hierarchical Condition Category; HPSA = health professional shortage area; I-SNP = institutional special needs plan; LIS = Low-Income Subsidy; OOP = out of pocket; PDSS = Part D Senior Savings; PPO = preferred provider organization; RxHCC = Prescription Drug Hierarchical Condition Category; SSBCI = Special Supplemental Benefits for the Chronically Ill; UF = uniformity flexibility. ASMDs for covariates reported in the table are 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 plans in the sample, and then taking the absolute value of the average SMD. "Unweighted" ASMDs reflect differences without EB weights. "Weighted" ASMDs reflect differences using EB weights.

Figure A.1 (next page) 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 (light green bars) and the maximum (pink bars) absolute SMDs achieved for each outcome, aggregating over all models and balancing variables. The figure shows that the average SMD for all outcome models was 0.1 or below.

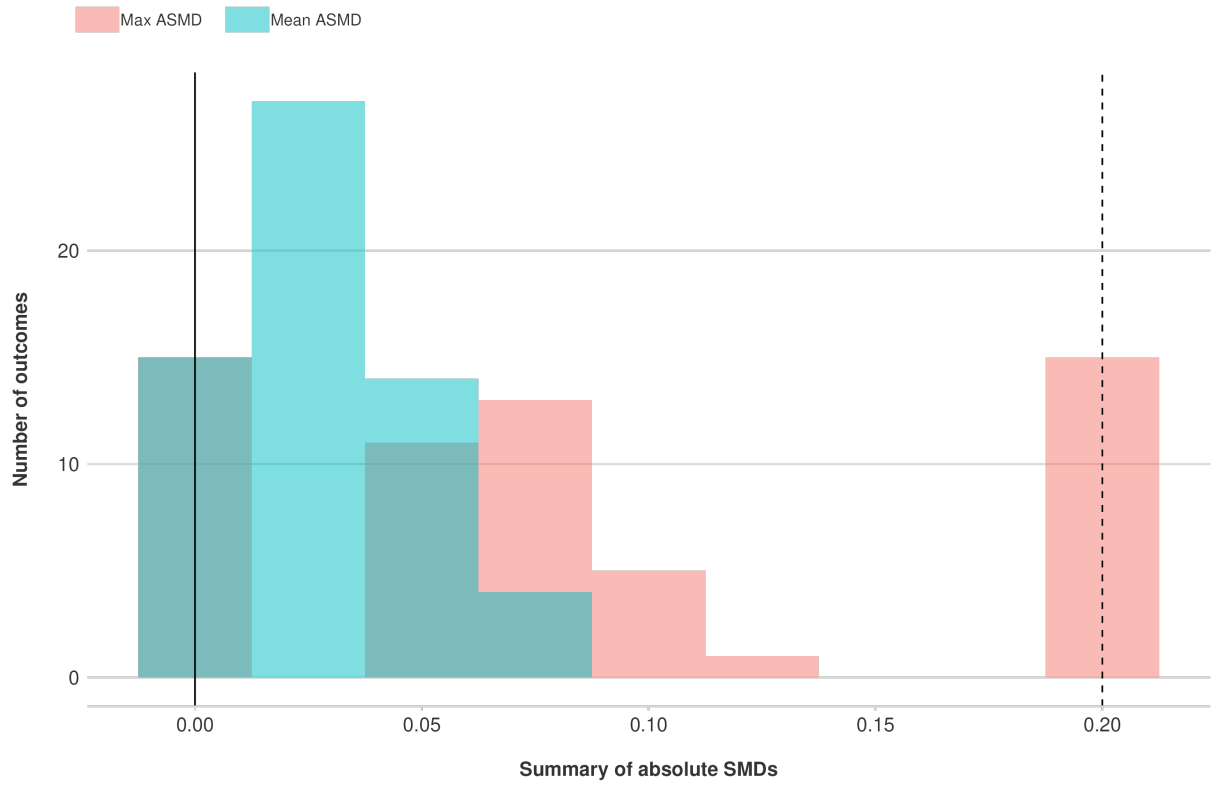
Beneficiary-Level Analyses

We have made several modifications to the methods used for beneficiary-level analyses in this Report in comparison to the 2023 Evaluation Report.

Inclusion of Additional Beneficiaries in Analysis

In our 2023 report, we restricted our beneficiary-level analysis to a "stable cohort" of MA beneficiaries who were enrolled in their VBID plan for at least one full year prior to VBID implementation. By restricting the sample to beneficiaries who were enrolled in the same plan before and after VBID, we avoided potential confounding that could occur if beneficiaries were exposed to VBID at the same time that they experienced other large changes in their health benefits (such as changes related to transitioning from fee-for-service [FFS] to MA). Further, this approach allowed us to have at least one year of pre-period data within the same plan to estimate pre-period trends.

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



NOTE: ASMDs 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 participant plans in the sample, and then taking the absolute value of the average SMDs. To produce a maximum and mean ASMD at the level of the outcome variable, we then take the maximum and the mean of the covariate-specific ASMDs for each outcome.

A limitation of the stable cohort approach, however, was that the sample was restricted to a subset of all VBID beneficiaries. Those who switched MA plans, who moved from FFS Medicare to MA, or who became newly eligible for Medicare were excluded from the sample. These exclusions required us to drop a nontrivial share of beneficiaries from our analysis. For example, on January 1, 2020, there were 262,983 VBID-targeted beneficiaries enrolled in VBID-participating plans; of those 192,735 (73.3%) had at least one year of pre-period data in the same plan and were included in our analysis. In other words, for our prior report, the sample restriction required us to ignore more than 25% of VBID-targeted beneficiaries in our analysis. If we had retained this sample restriction going into 2021, only 53.1% of VBID-targeted beneficiaries in plans that participated in 2020 and 2021 would have met our inclusion criteria, further limiting the representativeness of our sample.

For this year's evaluation, we therefore allowed beneficiaries with any pre-period data to contribute to the analysis. This includes beneficiaries who were in another VBID participating

plan, another MA plan, or FFS. Further, we no longer required that the beneficiary had a full year of pre-period data, as long as they had at least some pre-period data. We defined the pre-period based on the beneficiary's first exposure to VBID. For example, if a beneficiary entered a VBID participating plan in 2021, the beneficiary's pre-period would include 2020, even if the plan was a 2020 VBID participant. Table A.3 shows that with the new approach we retain more than 99 percent of VBID-targeted beneficiaries in our analysis.

Table A.3. Pre-Period Data Patterns, New Beneficiary Selection Approach

Pre-Period Data Pattern	Targeted Beneficiaries in 2020	Targeted Beneficiaries in 2021 (includes those observed in 2020 and those new in 2021)
2+ years of pre-period data in either MA or FFS	203,588	1,128,169
1 to 2 years of pre-period data in either MA or FFS	32,690	273,644
Part-year pre-period data in either MA or FFS	26,391	238,497
No pre-period data	314	3,849
Total targeted beneficiaries	262,983	1,644,159
Share with any pre period data	99.9%	99.8%

NOTE: Estimates include all beneficiaries, whether enrolled in MA-only or MAPD plans.

To accommodate these newly included beneficiaries in ways that credibly distinguished changes in outcomes associated with VBID from changes associated with plan switching, we made some modifications to prior methods. These changes are described below.

Stratification

While adding beneficiaries with pre-period data outside of their VBID plan increases the representativeness of our analysis, it makes it more difficult to differentiate changes in outcomes that are related to VBID from changes in outcomes that are related to coverage transitions. For example, someone who moves from FFS to MA may experience changes in utilization due to exposure to plan networks and utilization management processes unrelated to VBID. Further, data collected in FFS may differ systematically from data collected in MA plans, complicating the analysis. For example, if risk scores are systematically higher in MA due to higher coding intensity relative to 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 also weights the comparison group to resemble the VBID group on pre-period outcome trends and other characteristics. As described above, we use the GERBIL algorithm to impute these variables when they are missing. By expanding the sample to include beneficiaries with minimal pre-period data, we increase our reliance on imputation to estimate

missing pre-period characteristics and trends. The increase in imputation could lead to biases. For example, if imputed trends are slightly steeper than actual trends would have been, our method could attribute a post-VBID "flattening off" to the VBID model, when in fact this stemmed from imperfect imputation.

To address these issues, we stratified the data to ensure that VBID and comparison beneficiaries had the same length of pre-period information and the same source of pre-period 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, VBID beneficiaries who newly joined a VBID plan from FFS are only ever compared to non-VBID beneficiaries who newly joined a non-VBID plan from FFS. Similarly, VBID beneficiaries with limited pre-period data are only ever compared to non-VBID beneficiaries with limited pre-period data. Restricting the comparisons in this manner reduces the potential for bias that could arise due to transitions from FFS into MA or due to imperfect imputation. For example, if all beneficiaries who transition from FFS to MA experience an increase in risk score, the stratified DD approach will only attribute this increase to the VBID model if the change for beneficiaries who switched from FFS into VBID plans was larger than the change for beneficiaries who switched from FFS into non-VBID plans. Similarly, even if the imputation approach introduces biases in modeling trends for VBID and non-VBID beneficiaries, our method can 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 2 years of prior period data, MA
2. At least 2 years of prior period data, FFS
3. At least 1 but less than 2 years of prior period data, MA
4. At least 1 but less than 2 years of prior period data, FFS
5. Partial year pre-period data, MA
6. Partial year pre-period data, FFS.

Where possible, we further stratify these groups to capture dual and non–dual eligibility status, resulting in 12 groups. This approach ensures that our effect estimates adjust for any differences in pre-period coverage source (MA or FFS) between VBID and comparison beneficiaries. We exclude beneficiaries with no pre-period data.

We implement the stratifications above in addition to the stratifications by plan participation pattern described earlier in this appendix. Because we analyze beneficiary level outcomes for just two years (2020 and 2021), there are only three participation patterns of interest: participated in 2020, participated in 2020 and 2021, and participated only in 2021.

Tables A.4 and A.5 show the number of VBID-targeted beneficiaries in each stratification group for each year of analysis (2020 and 2021).

Table A.4. Number of VBID-Targeted Beneficiaries in Each Stratification Group Used in 2020 Analysis

Beneficiaries' Data Pattern	Dual Eligible	Non–Dual Eligible	Total
At least 2 years of pre-period data, MA	77,073	113,768	190,841
At least 2 years of pre-period data, FFS	6,311	3,403	9,714
At least 1 but less than 2 years of prior period data, MA	19,146	12,020	31,166
At least 1 but less than 2 years of prior period data, FFS	986	492	1,478
Partial year pre-period data, MA	12,616	5,997	18,613
Partial year pre-period data, MA	5,744	2,018	7,762
Total	121,876	137,698	259,574

NOTE: Sample includes VBID targeted beneficiaries enrolled as of January 1, 2020, who were not in an MA-only plan.

Table A.5. Number of VBID-Targeted Beneficiaries in Each Stratification Group Used in 2021 Analysis

Beneficiaries' Data Pattern	Participated in 2020 and 2021	Participated in 2021 only	Total
At least 2 years of pre-period data, MA	161,572	876,227	1,037,799
At least 2 years of pre-period data, FFS	8,840	64,302	73,142
At least 1 but less than 2 years of prior period data, MA	26,781	231,220	258,001
At least 1 but less than 2 years of prior period data, FFS	1,499	10,897	12,396
Partial year pre-period data, MA	18,380	151,110	169,490
Partial year pre-period data, MA	7,503	58,608	66,111
Total	224,575	1,392,364	1,616,939

NOTE: Sample includes VBID targeted beneficiaries enrolled as of January 1, 2020, who were not in an MA-only plan.

Details of Beneficiary-Level Entropy Balancing

As discussed above, the EB algorithm allows the user to specify the desired level of covariate balance δ for each covariate. The algorithm will then attempt to solve for weights that satisfy the balance constraint; however, not all values of δ are feasible when covariate balance is low, so that the algorithm may fail to converge.

For all beneficiary-level analyses, we attempt to balance all pre-trends first 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 binary variables with prevalences less than 0.05, we specified the EB algorithm to apply the tolerance δ to a difference that fixes the standardization in the treatment group to be based on a prevalence of 0.05. For example, when the desired SMD is 0.1, we instead set $\delta = 2.2$ percentage points $(0.1 * \sqrt{0.05 * 0.95})$. For these variables we also calculate the SMD standardized with respect to a variable that has a prevalence of 0.05.

Several outcomes measuring drug adherence or receipt of recommended care are defined only for subgroups of beneficiaries. For example, adherence to non-insulin diabetes medication is measured only for people with diabetes. We therefore 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 are 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 need to derive for this analysis, we include pre-trends for all possible beneficiary-level outcomes in a single set of weights.[2] This approach allows 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 pre-trends for outcomes that do not apply to all beneficiaries in the dataset (for example, breast cancer screening), we balance on the average plan-level outcome trends within the subgroup. Finally, when balancing on pre-trends for outcomes used in Poisson models (for example, number of ED visits, number of inpatient stays), we balance the difference in the log of the plan-level average outcomes, where the plan-level average is computed among all beneficiaries in the plan.

---

[1] We imposed stricter tolerance for beneficiary-level balancing than for plan-level balancing because we had many more benficairies than plans in both the treatment and comparison groups (for example, hundreds of thousands of beneficiaries versus 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 only the trend for 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.

Variables Included in Entropy Balancing

A challenge that arose when implementing this general method is that it was easier to achieve balance across a wider variety of covariates for some participation patterns than others. In particular, for the 2020 analyses, we were able to stratify all analysis on both the groups defined above as well as dual eligibility status (measured in 2019). We also controlled for several plan and contract level covariates that we did not ultimately include for the 2021 analyses. When we attempted to apply this same approach to the 2021 participation patterns, we were unable to achieve a satisfactory level of covariate balance within many of the subgroups. We therefore adjusted our stratification variables and covariate sets for the 2021 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	Stratification	Covariates
2020	1	Group, Dual	Bene, Plan, Contract
2021	01	Group	Bene
2021	11	Group	Bene

NOTE: Participation pattern describes the beneficiaries' history of VBID participation. For 2020, all VBID-targeted beneficiaries in participating plans are assigned a participation pattern of 1. In 2021, beneficiaries who were in VBID plans in both years are assigned a participation pattern of 11, while beneficiaries who are new to the model are assigned a participation pattern of 01.

Table A.7 provides a more detailed list of variables used in each analysis for each participation pattern.

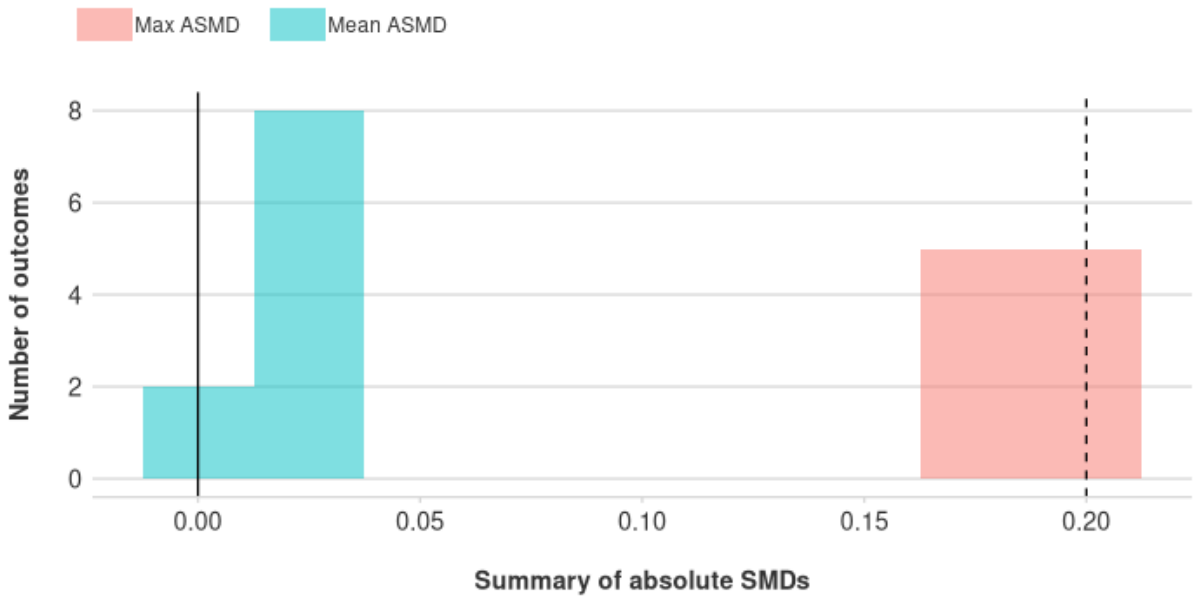
Table A.7. Balancing Variables Included in Beneficiary-Level Analyses and Standardized Mean Differences

Variable	Unweighted ASMD	Weighted ASMD
Pre-participation outcome trends		
Breast cancer screening trend	0.12	0.04
Diabetes trend	0.30	0.01
Hypertension trend	0.37	0.02
Number ED visits trend	0.26	0.01
Number inpatient stays trend	0.04	0.00
Part D OOP cost trend	0.08	0.00
Risk score trend	0.04	0.03
Statin trend	0.41	0.02
Beneficiary		
≥ $300 in total monthly Part D spend	0.55	0.17
≥ 2 ED visits	0.03	0.01
≥ 2 inpatient stays	0.10	0.03
≥ 8 concurrent medications	0.85	0.17
Age	0.38	0.08
Chronic RxFill	0.08	0.09

Variable	Unweighted ASMD	Weighted ASMD
Disabled	1.15	0.18
Dual eligible	3.14	0.13
ESRD	0.00	0.00
Fall risk	0.03	0.02
HCCs	0.11	0.04
LIS level	0.71	0.03
Male	0.23	0.01
Missing outcomes	0.11	0.06
Months in CMS	0.05	0.00
Nonadherent for specific drugs	0.11	0.03
MTM eligible	0.41	0.10
Part C premium	0.37	0.07
Puerto Rico	0.08	0.06
RxHCCs	0.06	0.00
SDI	0.37	0.07
Plan		
MA bid	0.03	0.06
Cost of MSB	0.19	0.01
Part D bid	0.71	0.11
HPSA	0.00	0.00
Into bonus	0.28	0.10
MA rebate	0.13	0.04
Months enrolled in plan	0.06	0.09
Out of bonus	0.08	0.05
PPO	0.83	0.13
Part D OOP amount	0.15	0.01
Part D total premium	0.08	0.08
Plan enrollment	0.08	0.07
SNP	1.12	0.04
Total nonbeneficiary expenditure	0.22	0.11
Contract		
Star Rating	0.13	0.10
PO		
BCBS	0.70	0.11
For profit enrollment	1.04	0.12
PO enrollment	0.57	0.12
County		
COVID-19 cases per 10,000	0.22	0.08
Median income	0.19	0.03
Penetration	0.04	0.10
Percentage over age 65	0.23	0.07
Standardized Medicare costs	0.23	0.02
Urbanicity	0.00	0.00

SOURCE: RAND analysis of VBID-participating plan and other data. SDI=social deprivation index.
NOTE: BCBS = Blue Cross Blue Shield; MSB = mandatory supplemental benefits; MTM = Medication Therapy Management; SDI = Social Deprivation Index; SNP = Special Needs Plan. ASMDs 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. Summary of Mean and Maximum Standardized Mean Differences After Balancing, Beneficiary-Level Analyses



NOTE: ASMDs 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 participation patterns.

Statistical Models and Estimation

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

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

where ζ_i and η_t are beneficiary and year fixed effects, respectively; X_{it} are a small set of time-varying controls that are we believe to be exogenous to VBID participation (indicating Part D Senior Service [PDSS] participation, uniform flexibility, Special Supplemental Benefits for the Chronically Ill [SSBCI], and Primarily Health-Related Supplemental Benefits [PHRSB] offerings at time t); and DD_{it} denotes an indicator of both treatment status and the post treatment time period t^*. For g we use the identity function for continuous and binary outcomes, and the natural logarithm for count outcomes. Finally, to estimate this model, the data contain observations from patterns $\boldsymbol{a}$ and not yet treated beneficiaries (as of time t^*). While this specification may allow for multiple post-treatment time periods, we only include 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 11. In 2021, we are only interested in the

2021 effect for this participation pattern, meaning that we need not estimate the 2020 effect for these beneficiaries, and therefore need not include data from 2020 to estimate the 2021 effect. We then estimate this model using weighted generalized linear models (specifically linear and Poisson regression), where the weights reweight the comparison observations to match the covariate distribution of the observations in participation pattern $\boldsymbol{a}$.

Subcomponent Analyses

To estimate the VBID effects for specific subcomponents, we estimate a weighted first-differences model (described below) using the comparison beneficiaries and the relevant VBID subcomponent group, using the balancing weights generated for the primary analysis. While the primary balancing weights balance the comparison group to the VBID General group, these do not necessarily balance the comparison group to each subcomponent group.

To overcome these limitations of the balancing weights, we control for the balancing characteristics within a regression model. A challenge, however, is that the model we use for our primary analysis – a two-way fixed effects model – includes beneficiary fixed effects, which are collinear with any time-invariant covariates – and therefore the time-invariant covariates cannot be included in this model specification. We therefore instead use weighted ordinary leats squares (OLS) to estimate a first-differences model for these analyses, as described in Equation A.6:

$$Y_{it} - Y_{it-1} = \delta_t + \theta A_{it} + \beta X_{it} + \gamma Z_i + \epsilon_{it} \qquad \text{(Equation A.6)}$$

where θ represents the treatment effect, A_{it} is a post-treatment indicator of the subcomponent, Z_i are the balancing characteristics, and X_{it} are the time-varying covariates representing whether the plan participated in other initiatives, including the PDSS Model, SSBCI, uniform flexibility (UF), and new PHRSB.[3] Moreover, because we stratify our estimation by treatment year, none of the first-differences are taken with respect to multiple post-treatment years. This model thus controls for beneficiary level heterogeneity without the collinearity problems that a fixed effects model presents.

While effect estimates from this first-differences model will not be identical to estimates from a two-way fixed effects model, both models are motivated by the parallel-trends assumption. Moreover, the first-differences specification allows us to correct for imbalances between the treatment and comparison groups with respect to time-invariant covariates using regression adjustment. By contrast, the balancing weights effectively control for these same time-invariant factors in our two-way fixed effects specification.

For inference for these analyses, to speed computation we simply take the weighted sum over the outer product of all the subgroup level standard error estimates, where the weights are proportional to the number of treated units. Each subgroup variance estimate, in turn, is based on

[3] In the plan- and contract-level models, we control for participation in other initiatives via EB weights.

a cluster-robust covariance matrix from the estimated model above. The resulting quantity is an upper bound on the true variance (conditional on the covariates and treatment assignment) that allows the correlation between all subgroup and year estimates to be one (the maximum possible correlation across models, and thus the most conservative possible assumption). For the estimates of the variance of the percent change estimates, we additionally assume that the correlation between the expected outcomes under treatment and control are equal to 0.5. These results are all available in Appendix D.

Contract-Level Analyses

Contract-level analyses and inference followed the same methodology as the plan-level analyses. Contracts were included in the analysis if they had at least one VBID General-participating plan or at least one eligible nonparticipating plan. Additionally, not all years of data in our study period were analyzed due to changes in the Star Ratings methodology; we discuss these issues in the contract results appendix, Appendix I. 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**
2021, excluded	NA1	1
	Total	**1**
2022, excluded	101	2
	NANA1	2
	NA11	1
	Total	**5**

NOTE: Participation history concatenates participation patterns for each year of the model test up to the outcome year. Patterns of 0 indicate the contract had no participating plans, patterns of 1 indicated the contract had at least one participating plan, and NA indicates the contract did not exist in that year. For example, a pattern of 011 indicates that the contract was observed in 2022 (three years into the model test) and had no participating plans in 2020 and at least one participating plan in 2021 and 2022.

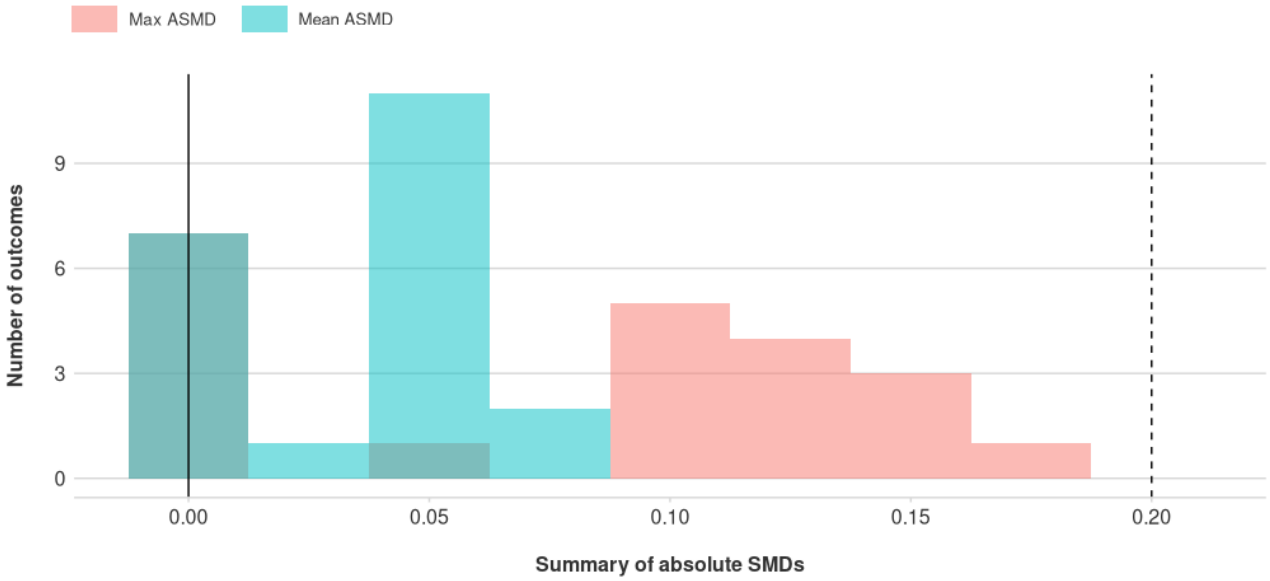
Due to difficulties in achieving balance, the contract analysis used a smaller number of balancing variables. Table A.9 lists the baseline characteristics used for EB in the contract-level analyses and reports the SMDs between VBID and comparison contracts in these outcomes both before and after weighting. Appendix B contains further information about these variables, including their sources. Table A.9 shows that EB succeeded in reducing the SMD between VBID and comparison groups below 0.2 for all balancing variables shown. Pre-participation outcome trends had an SMD of 0.02 after balancing.

Table A.9. Balancing Variables Included in Contract-Level Analyses and Standardized Mean Differences

Variable	Unweighted ASMD	Weighted ASMD
Average age	0.59	0.08
COVID-19 cases per 10,000	0.11	0.04
Percentage disabled	0.62	0.05
Percentage dual eligible	0.52	0.04
For-profit (beneficiary months)	0.21	0.02
Percentage LIS status	0.48	0.05
Missing outcomes	0.07	0.02
Part D basic premiums	0.22	0.07
MA penetration	0.27	0.02
Average MA risk score (HCC)	0.48	0.06
Average Part D risk score (RxHCC)	0.54	0.03
C-SNP	0.04	0.02
DSNP	0.34	0.03
I-SNP	0.02	0.02
Standardized Medicare costs per capita	0.18	0.06
Pre-participation outcome trends	0.08	0.02

SOURCE: RAND analysis of VBID-participating plan and other data.
NOTE: ASMDs for covariates reported in the table are 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. “Unweighted” ASMDs reflect differences without EB weights. “Weighted” ASMDs reflect differences using EB weights.

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



NOTE: ASMDs 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 participant plans in the sample, and then taking the absolute value of the average SMD. To produce a maximum and mean ASMD at the level of the outcome variable, we then take the maximum and the mean of the covariate-specific ASMDs for each outcome.

Appendix B. Variables Used in This Report's Analyses

Variables are used in this report for descriptive analyses, EB, and as outcomes in impact assessments and are described in more detail in prior reports (Eibner et al., 2023a; Eibner et al., 2023b; Khodyakov et al., 2022). Variables are aggregated at a variety of levels, including beneficiary, plan, contract, PO, and county. Beneficiary variables, including sociodemographic characteristics and utilization, are derived from FFS claims and encounter data, as well as other sources, such as Consumer Assessment of Healthcare Providers and Systems (CAHPS) surveys and data submitted by VBID participants. Plan, contract, and PO variables are sourced from CMS files. including those related to Star Ratings and plan benefit package (PBP) benefits, as well as information submitted to the Office of the Actuary. County-level variable sources include the American Community Survey and Area Health Resources Files. Additional detail is available upon request.

Appendix C. Methods for Statistical Analysis: Hospice Benefit Component

We used a similar set of analytic tools to quantitatively analyze the outcomes of the Hospice Benefit component as in the 2023 report. This Appendix summarizes those methods, and highlights changes. As with VBID General, are main tools are DD combined with EB; to avoid repetition we focus on differences from the VBID General analyses. For Hospice Benefit component plan-level financial outcomes (analyzed in Chapter 12), methods are identical to those described above for analysis of changes in plan-level financial outcomes associated with VBID General interventions.

Because the beneficiary-level outcomes of interest for Hospice can be observed only once for an individual at or near the end of their life, we are not able to estimate the effects of interest using data that track individuals over multiple years. As a result, there is a concern that the composition of the treatment or comparison group might shift over time in a way that degrades the parallel trends assumption underlying DD analyses. Moreover, whereas it makes sense to balance on pre-period trends when it is possible to track observational units over time, we do not have that advantage in the setting Hospice beneficiary-level outcomes—balancing on pre-period trends for individuals whose outcomes are observed in one year does not determine weights for individuals whose outcomes are observed in later years.

As in the previous report, because we are unable to accurately model beneficiary death in a prospective fashion (as is relevant for the denominator of many outcomes that are typically experienced near end-of-life), several of our key hospice outcomes use the cohort of decedents in a given year. For example, we are not able to identify beneficiaries prospectively who are eligible for hospice admission in a given year, so we instead analyze hospice admission rates in the decedent cohort, for example, the cohort of beneficiaries who passed away in a given year. Other outcomes relate to the cohort of beneficiaries admitted to Hospice, such as CAHPS Hospice measures. Beneficiaries in these analyses may or may not have been included in the decedent cohort in a given year.

If the treatment were applied in a randomized manner, we would expect the distributions of covariates to be approximately balanced between each of the groups defined by time period and treatment group. Our analytic approach attempts to create such balance, as described by Stuart et al. (2014). In contrast to the prior report, we now have observations of Hospice Benefit component participants in two periods: 2021 and 2022. To avoid bias that can result from application of two-way fixed effects for time-varying DD, we now generate 10 sets of weights. For plans that adopted the Hospice Benefit component in 2021, we weight the following groups to look like the 2021 Hospice-participating group: 2019 beneficiaries in a plan that never participates; 2019 beneficiaries in a plan that adopted the Hospice Benefit component in 2021;

2021 beneficiaries in a plan that never adopted the Hospice Benefit component; 2022 beneficiaries in a plan that never adopted the Hospice Benefit component; and 2022 beneficiaries in a plan that adopted the Hospice Benefit component in 2021. Similarly, we weighted the following groups to look like the 2022 beneficiaries who were in a plan that first adopted the Hospice Benefit component in 2022: 2019 beneficiaries in a plan that never adopted the Hospice Benefit component; 2019 beneficiaries in a plan that adopted the Hospice Benefit component in 2022; 2021 beneficiaries in a plan that never adopted the Hospice Benefit component; 2021 beneficiaries in a plan that first adopted the Hospice Benefit component in 2022; and 2022 beneficiaries in a plan that never adopted the Hospice Benefit component. Said more succinctly, for each year when the Hospice Benefit component was first adopted by some plans, we weight beneficiaries in "never participating" plans to look like the beneficiaries in the participating plans (in the first year they participated) and we weight the beneficiaries in the plans that would participate (or already had been participating for more than one year) to look like the beneficiaries in the first year of Hospice Benefit component participation. Thus, we are neither estimating changes in DD models comparing against "already-treated" beneficiaries nor are we allowing compositional changes over time in the treated or control groups. We use a fractional bootstrap to estimate standard errors for the 2021 treatment effect, and to combine the estimated treatment effects in 2022 for new and established Hospice Benefit component participants, as is done with the VBID General beneficiary-level models.

In EB weighted DD analyses, we do not wish to weight on variables that are measured after the intervention and may have been impacted by the intervention itself. For the longitudinal VBID General analyses, this is easier to achieve, where we can focus on variables that are measured before VBID started. For the Hospice analyses, we wish to balance on variables that are potentially associated with the outcomes of interest, but to exclude measures that may have been impacted by the intervention itself. The variables included in the analyses attempt to walk this line, but inevitably there will be cases where the groups will differ compositionally in important ways, though the DD analysis should also help to resolve any lingering differences between the treated and control groups. Also, another change from the prior report is that we now include additional Hierarchical Condition Category (HCC) variables in our weighting approach, which gives us information to make the various groups more similar to each other (and assess how different they are after weighting).

Because access to health care services was so strongly impacted in the early phases of the coronavirus disease 2019 (COVID-19) pandemic, we do not include 2020 data in our analyses.

Achieving good balance was more challenging for the Hospice regressions than for VBID General. Therefore, for the beneficiary-level hospice outcomes, we removed covariates that we were unable to balance from the EB procedure so that we could bring all remaining covariates within a SMD of at most 0.2. The variables that could not be balanced were instead included as covariates in the outcomes model. While we generally prefer the weighting approach due to its better alignment with estimating ATT effects and making fewer assumptions regarding the

relationship between potential confounders and the outcome of interest, controlling for covariates that we could not bring into acceptable balance is better than ignoring them altogether. While the models that included regression covariates were our primary specification, we used the alternative versions in which all characteristics were included in balancing (despite poor balance) as a validity check on the primary results.

Limitations

The Hospice analyses face two primary difficulties. First, few beneficiaries are in Hospice-participating plans, and a small fraction of plans are Hospice-participating, both in 2021 and in 2022. Second, those who are in Hospice-participating plans have very different characteristics on average than those who are not in Hospice-participating plans. In 2021 strong differences in beneficiary characteristics between the intervention and control groups seemed to be driven in part by the fact that a majority of beneficiaries in Hospice-participating plans lived in Puerto Rico. The Puerto Rico-heavy nature of the participating group eased in 2022, but there were still many differences between participating and nonparticipating groups and from year to year that were generally difficult to resolve through weighting. This results in lingering imbalances between the various groups described above, which could result in bias of treatment effect estimates if it corresponds to violations of parallel trends. Because the same individuals are not tracked over time, changes in beneficiary characteristics within a plan could be expected to be particularly problematic as it relates to parallel trends. (Note that this is not a concern with VBID General analyses.) At the same time, the weighting produces low ESSs in some cases, which can result in relatively wide confidence intervals (CIs).

In Appendix J, we present additional detail on the quality of the balance for the hospice analysis, including pre-VBID absolute standardized mean difference (ASMDs), ESSs after balancing, and quality of the balance. In all, the data limitations for Hospice Benefit component beneficiary-level analyses are more acute than for VBID General analyses, so the level of evidence they produce is lower.

Appendix D. Subgroup Analysis for VBID General Outcomes

Through VBID General, POs have many options to tailor PBPs to promote beneficiary health, patient-centeredness, and high-value care. In the main text of our report, we grouped all VBID General interventions together for the purposes of evaluating their impact. However, it is possible that impacts vary by the type of interventions offered. In this appendix, we consider whether the relationship between VBID General and the outcomes of interest varied with intervention type. In consultation with CMS, we selected six outcomes and five intervention types for subgroup analyses (Table D.1). Three of the outcomes were measured at the plan level and three at the beneficiary level. We analyzed VBID flexibilities separately from RI, and also conducted analyses of four subtypes of VBID flexibilities (socioeconomic status [SES] targeting, chronic conditions targeting, Part C cost sharing reductions, and Part D cost sharing reductions).

Table D.1. VBID General Subgroup Analyses Included, and Unit of Observation Considered

Intervention Type	Total Costs to CMS	Total MAPD Premiums	Number of MSB	Inpatient Stays	Risk Score	Part D OOP Costs
VBID Flexibilities	Plan	Plan	Plan	Beneficiary	Beneficiary	Beneficiary
SES targeting	Plan	Plan	Plan	Beneficiary	Beneficiary	Beneficiary
Part C cost-sharing reductions	Plan	Plan	Plan	Beneficiary	Beneficiary	Beneficiary
Part D cost-sharing reductions	Plan	Plan	Plan	Beneficiary	Beneficiary	Beneficiary
Chronic conditions targeting	Plan	Plan	Plan	Beneficiary	Beneficiary	Beneficiary
RI	Plan	Plan	Plan	Beneficiary	Beneficiary	Beneficiary

NOTE: We analyzed Total Costs to CMS for 2022, Total MAPD Premiums and Number of MSB for 2023, and all beneficiary-level outcomes for 2021.

We report subgroup analyses for the most recent year of data available: 2022 for costs to CMS, 2023 for premiums and mandatory supplemental benefits (MSB), and 2021 for beneficiary-level outcomes (inpatient utilization, risk scores, and Part D OOP costs).

Our methodology for the subgroup analyses is described in Appendix A. While the plan level subgroup analyses use a similar approach to the analyses presented in the main text, for the beneficiary analysis, we used first-difference models to conduct subcomponent analyses. First-difference models assess whether the change in the outcome among targeted beneficiaries is associated with plans' participation in VBID in the post-treatment period conditional on covariates. Conceptually, first-difference models are similar to DD models because both types of models assess whether changes over time for treated beneficiaries diverged from changes over time for comparison beneficiaries. A significant advantage of the first-difference approach is that

we can estimate the model while adjusting for imbalances among time-invariant covariates that are imbalanced between the VBID subcomponent and comparison groups without recalculating a separate set of balancing weights for each subcomponent. While results from the first-difference models were of the same sign and statistical significance as the DD models shown in the main text, point estimates were sometimes different. For the beneficiary-level analysis, we therefore report the coefficients for the full sample, along with the results for subgroups, to enable readers to compare to the original results reported in the main text (and also found in Appendix H).

Results

Costs to CMS

In Table D.2, we report effects on costs to CMS by intervention type for 2022. We found statistically significant increases for all subtypes of VBID General interventions analyzed, with the largest effects for plans with SES targeting ($43 per member per month [PMPM] increase, 95% CI: $2 to $82), and Part D cost sharing reductions ($42 PMPM increase, 95% CI: $6 to $61).

Table D.2. Estimated Associations Between VBID General Interventions and Total Costs to CMS, by Subgroup, 2022

Intervention Type	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
VBID Flexibilities	39.40	13.44	10.56	64.19	<0.001	960
SES targeting	43.37	20.90	1.78	81.80	0.037	554
Part C cost-sharing reductions	34.84	13.84	6.16	60.90	0.025	418
Part D cost-sharing reductions	41.64	14.42	12.76	69.96	0.003	869
Chronic conditions targeting	28.69	13.07	2.38	53.23	0.041	467
RI	24.98	10.47	4.23	45.17	0.015	1,130

SOURCE: RAND analysis of CMS and other data.

MAPD Premium

In the main text, we found no statistically significant association between VBID General implementation and MAPD premiums (Chapter 5). In Table D.3, we found no statistically significant associations with premiums across the subgroups.

Table D.3. Estimated Associations Between VBID General Interventions and Total Medicare Advantage Part D Premiums, by Subgroup, 2023

Intervention Type	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
VBID Flexibilities	–0.22	1.44	–3.00	2.50	0.903	1641
SES targeting	0.40	1.65	–2.79	3.57	0.805	1012
Part C cost-sharing reductions	–1.05	1.60	–4.32	1.96	0.550	571
Part D cost-sharing reductions	0.07	1.46	–2.71	2.91	0.918	1528
Chronic conditions targeting	0.04	1.55	–3.03	3.01	0.927	852
RI	–0.81	1.69	–4.08	2.48	0.730	969

SOURCE: RAND analysis of CMS and other data.

Mandatory Supplemental Benefits

VBID General was associated with statistically significant decreases in the number of MSBs offered to enrollees for all subtypes of interventions that we analyzed (Table D.4). Decreases were largest for plans with chronic conditions targeting (–0.77 change in MSBs offered, 95% CI: –1.26 to –0.30).

Table D.4. Estimated Associations Between VBID General Interventions and Number of Mandatory Supplemental Benefits, by Subgroup, 2023

Intervention Type	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
VBID Flexibilities	–0.48	0.25	–1.01	–0.04	0.038	1652
SES targeting	–0.64	0.33	–1.35	–0.04	0.039	1040
Part C cost-sharing reductions	–0.59	0.29	–1.16	–0.05	0.035	553
Part D cost-sharing reductions	–0.70	0.28	–1.31	–0.20	0.002	1540
Chronic conditions targeting	–0.77	0.25	–1.26	–0.30	0.002	886
RI	–0.54	0.23	–0.98	–0.10	0.012	997

SOURCE: RAND analysis of CMS and other data.

Inpatient Stays

Because inpatient stays are estimated at the beneficiary level, we use the first difference methodology described above for subgroup analyses. The first two rows of Table D.5 compare the original estimate from the main report to the results of a FD model. Both estimation approaches indicate that VBID was associated with a statistically significant increase in inpatient stays. The subsequent rows, all estimated using the first difference approach, show that VBID was associated with a statistically significant increase in inpatient stays for the VBID Flexibilities subgroup as a whole, and for all subcomponents of VBID Flexibilities considered in

the analysis. The largest effects were for Part C cost sharing and chronic conditions interventions, which were both associated with an increase in inpatient utilization of 9% ($p < 0.001$ in both cases, 95% CIs reported in the table). There was no statistically significant association between RI interventions and inpatient utilization.

Table D.5. Estimated Associations Between VBID General Interventions and Inpatient Stays, by Subgroup, 2021

Intervention Type	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
All targeted beneficiaries, (original finding from main text)	0.04	0.01	0.02	0.05	<0.001	3,683,191
All targeted beneficiaries, FD	0.06	0.00	0.05	0.07	<0.001	3,683,191
VBID Flexibilities subgroup, FD	0.06	0.00	0.05	0.07	<0.001	3,576,412
SES targeting subgroup, FD	0.05	0.01	0.04	0.06	<0.001	3,392,197
Part C cost-sharing reductions subgroup, FD	0.09	0.01	0.06	0.11	<0.001	2,099,530
Part D cost-sharing reductions subgroup, FD	0.05	0.01	0.04	0.06	<0.001	2,935,354
Chronic conditions targeting subgroup, FD	0.09	0.01	0.08	0.11	<0.001	2,252,188
RI subgroup, FD	0.06	0.04	–0.02	0.13	0.146	2,077,625

SOURCE: RAND analysis of CMS and other data. Inpatient stays with a COVID-19 diagnosis are excluded. FD indicates results are from first-difference models.

Risk Score

Table D.6 shows subgroup analyses that estimate the relationship between VBID and targeted beneficiaries' risk scores. The first and second rows show the result from the main text compared to the result derived from a first difference models. Again, the sign and statistical significance of the estimates are the same regardless of approach, although the magnitude of the association is somewhat smaller with the first difference approach. Similar to the inpatient results above, we found that VBID General was associated with increases in risk scores for all VBID Flexibilities subgroups considered, but not for RI interventions. Risk score increases were particularly large among plans with Part C cost sharing reductions, for which we estimated an increase of 0.17 points (95% CI: 0.12 to 0.22).

Table D.6. Estimated Associations Between VBID General Interventions and Risk Score, by Subgroup, 2021

Intervention Type	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
All targeted beneficiaries, (original finding from main text)	0.07	0.02	0.04	0.10	<0.001	3,683,191
All targeted beneficiaries, FD	0.04	0.02	0.01	0.08	0.009	3,683,191
VBID Flexibilities subgroup, FD	0.05	0.02	0.01	0.08	0.006	3,576,412
SES targeting subgroup, FD	0.04	0.02	0.01	0.08	0.015	3,392,197
Part C cost-sharing reductions subgroup, FD	0.17	0.03	0.12	0.22	<0.001	2,099,530
Part D cost-sharing reductions subgroup, FD	0.05	0.02	0.00	0.09	0.031	2,935,354
Chronic conditions targeting subgroup, FD	0.06	0.02	0.02	0.10	0.004	2,252,188
RI subgroup, FD	0.02	0.04	–0.05	0.09	0.641	2,077,625

SOURCE: RAND analysis of CMS and other data. FD indicates results are from first-difference models.

Part D Out-of-Pocket Costs

Table D.7 shows subgroup analyses that consider associations between VBID General and changes in OOP spending. As with the prior charts, the first two rows compare the original report estimate to the first difference estimate—results are very similar across the two approaches. In subgroup analyses, we find that VBID General implementation was associated with statistically significant decreases in Part D OOP costs in VBID Flexibilities plans, SES targeting plans, and plans with Part D cost-sharing reductions (Table D.7). Unsurprisingly, the decrease is largest in plans with Part D cost sharing reductions (-$41.84, 95% CI: –$53.04 to –$30.63).

Table D.7. Estimated Associations Between VBID General Interventions and Part D Out-of-Pocket Costs, by Subgroup, 2021

Intervention Type	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
All targeted beneficiaries, (original finding from main text)	–24.59	4.28	–32.99	–16.20	<0.001	3,683,191
All targeted beneficiaries, FD	–22.72	5.53	–33.56	–11.87	<0.001	3,683,191
VBID Flexibilities subgroup, FD	−21.59	5.68	−32.71	−10.46	<0.001	3,576,412
SES targeting subgroup, FD	−24.19	5.89	−35.74	−12.65	<0.001	3,392,197
Part C cost-sharing reductions subgroup, FD	18.72	17.27	−15.12	52.56	0.278	2,099,530
Part D cost-sharing reductions subgroup, FD	−41.84	5.72	−53.04	−30.63	<0.001	2,935,354
Chronic conditions targeting subgroup, FD	−6.58	8.10	−22.46	9.30	0.417	2,252,188
RI subgroup, FD	−30.81	25.94	−81.65	20.03	0.235	2,077,625

SOURCE: RAND analysis of CMS and other data. FD indicates results are from first-difference models.

Appendix E. Qualitative Methods

This appendix describes our approach to collecting and analyzing the primary data from POs and hospices. All data collection procedures were reviewed and approved by the RAND Human Subjects Protection Committee. In 2023, we fielded a questionnaire to all POs that participated in the model test; we then sampled 42 POs for either an in-person site visit or a virtual interview. In addition, we surveyed and interviewed representatives of 10 hospices either in-person during site visits or virtually. The goal of these data collection activities was to provide additional nuance into the implementation experiences in participating in VBID and to describe how and why VBID implementation was associated with key model outcomes. All interviews were conducted using an approach similar to the one we described in detail in our 2023 VBID evaluation report (Eibner et al., 2023b). Some of the text below is copied verbatim from Appendix A of the 2023 report (Eibner et al., 2023a).

To recruit PO and hospice representatives, we reached out to contacts at each organization via email and provided them with a brief description of the interview, its purpose, and logistical details. We conducted follow-up outreach activities by email and phone with up to three attempts to reach 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 who they considered to be most knowledgeable about VBID to participate in the interviews. During the scheduling phase, we sent the consent form via email. We obtained verbal 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. We also conducted in-person site visits with six POs, including four that participated in the Hospice Benefit component. As part of site visits, we were able to interview representatives of six in-network hospices working with these four POs.

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. We provide additional descriptions of our sampling and data collection processes in the following section.

Sampling and Data Collection

PO Survey and Interviews

We invited all 52 POs that participated in the VBID model in 2023 to complete an online questionnaire. We also invited a sample of 42 POs to participate in a follow-up interview. We prioritized POs that have 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 understand POs' experiences with specific model components; implementation barriers that they encountered; and the impact that they expect their VBID interventions will have on plan enrollment and retention, utilization of VBID benefits and services, beneficiary health outcomes, and plan and beneficiary costs in 2023. 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 2022. While the questionnaire items were primarily closed ended, interview questions were open ended.

Survey questions varied based on whether POs implemented VBID General or the Hospice Benefit component, but generally included the same type of questions. For example, while rating questions about VBID General implementation that were focused on various aspects of administrative processes and communication, Hospice Benefit component participants also answered questions about challenges related to training, care delivery, and creating and maintaining a hospice network. Similarly, while all participants answered close-ended questions about how VBID will affect (or has already affected) a variety of plan- and beneficiary-level outcomes, Hospice Benefit component participants also rated the impact of their interventions on utilization outcomes. In addition, POs implementing the Hospice Benefit component answered questions about their model test interventions.

During the interviews, we discussed POs' responses to the pre-interview surveys and asked additional questions covering such topics as

- details of VBID interventions
- implementation experiences, successes, and challenges
- intervention uptake among beneficiaries
- 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 52 POs invited to complete the questionnaire, 51 did so. Of the 42 POs invited to participate in an interview, 37 participated. Representatives of six POs did so in-person during a site visit; representatives of the remaining 31 POs participated in a virtual interview. Of the POs we interviewed, 27 implemented only VBID General, five implemented only the Hospice Benefit component, and the remaining five implemented both components). We collected all PO data between May and September 2023 and spoke with a total of 255 PO representatives across the 37 POs. The interviews varied in length based on whether they were conducted in-person or virtually and based on whether a PO implemented one or both model test components. In-person interviews typically lasted for 4 hours; virtual interviews lasted for one or two hours, depending on the number of implemented intervention components.

Hospice Interviews

We administered pre-interview questionnaires and conducted interviews with representatives of hospices that were part of PO hospice networks ("in-network" hospices) to understand which services they were contracted to provide as part of the Hospice component, challenges they faced implementing the Hospice Benefit component and factors that supported them to implement it, impacts of the Hospice component on their hospice, and their intent to participate in PO hospice networks in the future.

Hospice interview protocols varied based on whether we had interviewed representatives from the hospice in previous years. During interviews, we discussed hospices' responses to the pre-interview questionnaires. In all interviews, we asked open-ended questions covering such topics as

- reasons for joining or not joining the hospice network of all POs participating in the Hospice Benefit component in their service area
- the process of negotiating contracts and working with POs
- implementation experiences, successes, and challenges
- experiences working with the POs as an in-network or out of network (OON) hospice (if relevant)
- changes in care delivery as a result of the Hospice Benefit component
- thoughts about model achieved and expected future outcomes, including any unintended outcomes
- plans for participating in the Hospice Benefit component in the future.

We assembled a diverse sample of in-network hospices to achieve thematic saturation. To do so, we asked POs that implemented the Hospice Benefit component to share contact information of their in-network hospices that provided care to the largest proportion of their beneficiaries. In sampling hospices, we prioritized those that provided care to VBID beneficiaries from more than one PO participating in the Hospice Benefit component, either as in-network or OON.

Of the 16 hospices invited to participate in our evaluation, ten participated in an interview, two emails were undeliverable, and four remaining did not respond to our invitation. Of the ten participants, nine also completed the pre-interview questionnaire. Five hospice participants were for-profit, three were not-for-profit, and two were classified as government or other in CMS Provider of Services files. Two participants were chain organizations; hospices within these chains ranged in size, with some having an average daily census of 101 and 249, some 250 and 499, and some 500 or more. The remaining eight participants were independent hospices, one with average daily census between 250 and 499 and the other seven with census of 500 or more. We interviewed representatives of two hospices in prior years; one of these two hospices was an OON hospice in 2021 but joined a PO network in 2023. Four hospices that we interviewed were in the networks of two or three POs. We conducted interviews with the representatives of five hospices in-person during site visits; we conducted virtual interviews with representatives of the remaining five hospices. All hospice interviews were conducted between May and October 2023.

During this time, we spoke with 23 hospice representatives. The interviews varied in length based on whether they were conducted in-person or virtually, but on average were one hour.

Data Analysis

We followed the same qualitative data analysis process as in previous reports (Eibner et al., 2023b; Khodyakov et al., 2022). We used descriptive statistics to analyze the responses from the pre-interview questionnaires sent to POs and hospices 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 qualitative software program. 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. We used the same coders as in previous years and relied on the same process for training coders in the codebook for team-based coding. Coders individually coded a set of test transcripts, which were then reviewed by the broader coding team. One researcher resolved questions and discrepancies in coding. We calculated a combined kappa score of 0.79 using the Dedoose feature for interrater reliability for the PO interviews (McHugh, 2012). Following establishment of a reliable kappa score, coders independently coded the remainer of the transcripts. Each researcher assigned to each section of the 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 and hospice perspectives on and experiences with the model test. We also compared emerging themes from this evaluation with the findings from previous evaluations. Lastly, in line with the mixed-methods nature of this evaluation, we integrated quantitative and qualitative analytic techniques to ensure the rigor of our findings.

Appendix F. Descriptive Information on Participating Plans

Table F.1 presents descriptive statistics for participating and eligible nonparticipating POs in 2023.

Table F.1. Descriptive Statistics for Participating and Eligible Nonparticipating POs, 2023

Characteristic	VBID General	Hospice Benefit Component	Eligible Nonparticipating POs
Number of POs	46	15	104
Blue Cross affiliate (%)	17.39	40	16.35
PO geographic reach (%)			
1–2 states	67.39	40*	75.96
3–8 states	19.57	26.67	22.12
9 or more states	13.04*	33.33*	1.92
For-profit status	52.17	53.33	43.27
MA penetration	55.5** (7.8)	54.7** (4.4)	50.2 (10.3)
Median income	$30,358 (5,931)	$31,114 (4,117)	$31,442 (3,718)
Enrollment	443,603* (1,296,502)	1,209,585* (2,108,613)	25,202 (38,682)

NOTE: ***, **, and * represent statistical significance at the 0.1%, 1%, and 5% levels, respectively.

Table F.2 presents descriptive statistics for 2023 VBID participating and eligible comparison plans, separated by participation in VBID General and VBID Hospice. We also present descriptive statistics for Hospice Benefit component participants located on the mainland U.S.

Table F.2. Descriptive Statistics for VBID Participating and Eligible Comparison Plans, 2023

Characteristic	VBID General Participating Plans	Hospice Benefit Component Participating Plans	Hospice Benefit Component Participating Plans —Mainland Only	Comparison Group
N	1218	112	81	3093
Offers Part D (%)	99.43***	96.43**	96.3**	89.91
DSNP (%)	49.67***	22.32***	16.05**	4.33
C-SNP (%)	4.93	3.57	0***	5.59
I-SNP (%)	0.41***	4.46	6.17	4.66
$0 premium plan (%)	34.65***	67.86	56.79	62.92
PDSS participant (%)	44.91	60.71***	61.73**	42.74
Offers UF (%)	15.85***	27.68***	30.86***	9.12
Offers SSBCI (%)	32.76***	29.46	35.8*	21.66

Characteristic	VBID General Participating Plans	Hospice Benefit Component Participating Plans	Hospice Benefit Component Participating Plans —Mainland Only	Comparison Group
Offers NPHRB (%)	97.37***	97.32*	96.3	93.6
PPO (%)	30.54***	26.79*	37.04	36.89
Total premium	22.3 (23.1)*	15.5 (30.5)	21.2 (34.1)	20.4 (39.2)
Maximum OOP limit	5931.5 (2271.4)***	4665.4 (1773.6)*	5143 (1874.9)	5054.1 (1927.8)
Rural (%)	8.5 (11.7)	10.9 (14.7)*	14.3 (16)**	8.1 (13.1)
Suburban (%)	19.3 (15)	17.9 (13.9)	21.5 (14.8)	19.3 (17)
Urban (%)	72.1 (22.5)	71.2 (22.3)	64.2 (22.6)**	72.6 (25.1)
Dual eligible enrollees (%)	57 (42.7)***	30.1 (38.8)**	25.6 (35)	19.5 (26)
Part D LIS enrollees (%)	60.8 (40)***	23.8 (31.3)	31.4 (33.8)	25.8 (26.8)
Age	67.9 (4.6)***	71.8 (4.5)	71.8 (4.7)	71.5 (3.9)
Male (%)	43.3 (6.1)***	44.6 (7.2)***	42.9 (6.2)***	47.1 (9)
MA bid	883 (113.3)	743.2 (234.3)***	880 (76.9)	878 (94.7)
PDB	35.5 (16.3)***	41.2 (16.8)	43.9 (18)*	38.9 (24)
MA premium	3 (13.6)***	4.6 (18)	6.2 (20.9)	7.5 (24.8)
Part D premium	19.4 (16.8)***	11.2 (18.8)	15.6 (20.6)	14.4 (21.8)
MSB costs	102.8 (77.2)***	84.6 (70)***	64.6 (58.1)**	46.4 (30.8)
MA rebate	212 (91.8)***	248.2 (112.9)***	201.3 (90.1)*	178.6 (85.3)
Administrative costs	174.8 (53.3)***	158.7 (92.4)	168 (105.5)	151.1 (68.3)
Star Rating	4.1 (0.6)***	4.2 (0.4)***	4.2 (0.4)***	3.8 (0.7)
Total enrollment	7330.5 (12919.1)***	10485.7 (12815.5)***	10394.2 (11598.1)***	4683.5 (10461.3)

NOTE: NPHRB = non–primarily health-related supplemental benefit. ***, **, and * represent statistical significance at the 0.1%, 1%, and 5% levels, respectively.

Table F.3 presents selected descriptive statistics for VBID participating plans from 2020 through 2023.

Table F.3. Descriptive Statistics for VBID Participating Plans, 2020 to 2023

Characteristic	VBID General	VBID General	VBID General	VBID General	Hospice Benefit Component	Hospice Benefit Component	Hospice Benefit Component
	2020	2021	2022	2023	2021	2022	2023
Number of plans	144	376	859	1,218	52	109	112
PO size	1,071,379 (979,047)	3,039,389 (1,116,647)	3,390,326 (1,784,005)	3,732,206 (2,254,301)	622,562 (1,086,971)	2,188,365 (2,283,324)	2,898,513 (2,777,129)
DSNPs (%)	27.8	38.0*	43.5	49.7**	28.8	18.3	22.3
PO's Geographic reach							

Characteristic	VBID General	VBID General	VBID General	VBID General	Hospice Benefit Component	Hospice Benefit Component	Hospice Benefit Component
Offered 1–2 states	19.4	7.2*	13.4***	6.5***	69.2	28.4***	23.2
Offered 3–8 states	0	0.3	0.1	5.7***	11.5	14.7	13.4
Offered 9+ states	80.6	92.6*	86.5*	87.8	19.2	56.9***	63.4
Plans in Puerto Rico (%)	4.9	2.1	4.1	2.6	50	25.7**	27.7
Offers UF (%)	4.9	2.9	10.8***	15.9**	51.9	30.3*	27.7
Offers SSBCI (%)	9.0	19.7**	28.2**	32.8*	34.6	44.0	29.5*
Offers NPHRB (%)	93.1	97.3	98.1	97.4	92.3	93.6	97.3

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 '*' appears on the last year referenced. For example, differences between 2022 and 2023 are shown on the 2023 variables.

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

Table F.4. Descriptive Statistics for New VBID Participating Plans, 2020 to 2023

Characteristic	VBID General	VBID General	VBID General	VBID General	Hospice Benefit Component	Hospice Benefit Component	Hospice Benefit Component
	2020	2021	2022	2023	2021	2022	2023
Number of plans	144	279	489	462	52	61	27
PO size	1,071,379 (979,047)	3,301,270 (937,460)	2,793,575 (1,866,183)	2,886,272 (2,322,573)	622,562 (1,086,971)	2,927,306 (2,487,359)	2,505,152 (2,632,847)
DSNPs (%)	27.8	41.6	46.4	58.4	28.8	13.1	29.6
Percentage of plans offered at state level	19.4	2.5	18.6	7.8	69.2	16.4	14.8
Percentage of plans offered at regional level	0	0	0	9.1	11.5	16.4	25.9
Percentage of plans offered nationally	80.6	97.5	81.4	83.1	19.2	67.2	59.3
Percentage of plans in Puerto Rico	4.9	2.5	4.7	3.5	50	8.2	29.6

Table F.5 displays the components that VBID-participating plans implemented in each year.

Table F.5. VBID Model Test Components Implemented, by PO and Year

PO ID	VBID General	VBID General	VBID General	VBID General	Hospice Benefit Component	Hospice Benefit Component	Hospice Benefit Component
	2020	2021	2022	2023	2021	2022	2023
PO B	✓	✓	✓	✓	—	—	✓
PO C	✓	✓	—	✓	—	—	—
PO E	—	—	✓	✓	—	—	—
PO G	✓	✓	✓	✓	—	✓	✓
PO J	✓	✓	✓	—	—	—	—
PO L	✓	✓	✓	✓	—	✓	✓
PO M	—	—	—	—	✓	✓	✓
PO N	✓	✓	✓	✓	—	—	—
PO O	✓	✓	✓	✓	—	—	—
PO P	✓	✓	✓	✓	✓	✓	✓
PO Q	✓	✓	✓	✓	—	—	—
PO R	—	✓	✓	✓	✓	✓	✓
PO S	—	✓	✓	✓	—	—	—
PO T	—	—	✓	✓	✓	—	—
PO U	✓	✓	✓	✓	—	✓	—
PO V	—	—	—	—	✓	✓	✓
PO X	—	—	—	—	✓	✓	✓
PO Y	—	✓	✓	✓	✓	✓	✓
PO Z	—	—	—	—	✓	✓	—
PO AA	✓	—	✓	✓	—	—	—
PO AB	✓	—	—	✓	—	—	—
PO AC	—	—	✓	✓	—	—	—
PO AD	—	—	✓	✓	—	—	—
PO AE	—	—	✓	✓	—	—	—
PO AF	—	—	✓	✓	—	—	—
PO AG	—	—	✓	✓	—	—	—
PO AH	✓	✓	✓	✓	✓	✓	✓
PO AI	—	—	—	—	—	✓	✓
PO AJ	—	—	—	—	—	✓	✓
PO AK	—	—	✓	✓	—	—	—

PO ID	VBID General	VBID General	VBID General	VBID General	Hospice Benefit Component	Hospice Benefit Component	Hospice Benefit Component
	2020	2021	2022	2023	2021	2022	2023
PO AL	—	—	✓	✓	—	—	—
PO AO	—	—	✓	✓	—	—	—
PO AP	—	—	✓	✓	—	—	—
PO AQ	✓	—	✓	✓	—	—	—
PO AR	—	—	✓	—	—	—	—
PO AS	—	—	—	✓	—	—	—
PO AT	—	—	—	✓	—	—	—
PO AU	—	—	—	✓	—	—	—
PO AV	—	—	—	✓	—	—	—
PO AW	—	—	—	✓	—	—	—
PO AX	—	—	—	✓	—	—	—
PO AY	—	—	—	✓	—	—	—
PO AZ	—	—	—	✓	—	—	—
PO BA	—	—	—	—	—	—	✓
PO BB	—	—	—	✓	—	—	—
PO BC	—	—	—	✓	—	—	—
PO BD	—	—	—	✓	—	—	—
PO BE	—	—	—	✓	—	—	✓
PO BF	—	—	—	✓	—	—	—
PO BG	—	—	—	✓	—	—	—
PO BH	—	—	—	✓	—	—	—
PO BI	—	—	—	✓	—	—	✓
PO BJ	—	—	—	✓	—	—	—
PO BK	—	—	—	✓	—	—	—
PO BL	—	—	—	✓	—	—	—

Table F.6. on the next page provides some additional detail on the VBID General targeted populations and subcomponents offered by each participating PO in 2023.

Table F.6. 2023 VBID General Model Test Subcomponents Implemented, 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 L	—	✓	✓	—	✓	—	✓	—
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 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 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	—	✓	✓	—	✓	—	—	—

Appendix G. Plan-Level VBID General and Hospice Benefit Component Outcomes

This appendix provides detailed regression results for our analyses of plan-level outcomes, including all estimates referenced in Chapters 5 and 12 of this Report. In addition to results discussed in Chapters 5 and 12, the Appendix contains some supplementary results on outcomes not discussed in the Chapters that offer insight into mechanisms for changes in outcomes. The Appendix also contains estimates for the plan-level outcomes discussed in Chapters 5 and 12 in which VBID plans were weighted by enrollment (instead of weighted equally). These enrollment-weighted estimates may be of interest because they reflect the average effect per beneficiary in the participant plans. For financial outcomes, such as bids or costs to CMS, that are defined on a PMPM basis, the enrollment-weighted estimates may thus offer greater insight than the plan-weighted estimates into the aggregate effect of VBID on costs.

Tables G.1 through G.13 provide full regression results for outcomes reported in Chapter 5.

Table G.1. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Enrollment

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	0.05	0.10	–0.15	0.27	0.750	304
2021	0.03	0.05	–0.06	0.13	0.485	766
2022	0.05	0.06	–0.08	0.16	0.392	1,500
2023	0.24	0.09	0.08	0.41	0.003	1,841
Hospice						
2021	0.05	0.16	–0.25	0.36	0.866	91
2022	0.07	0.10	–0.14	0.27	0.516	211
2023	0.14	0.17	–0.20	0.47	0.453	369

SOURCE: RAND analysis of CMS and other data.
NOTE: Estimates reflect regression results for natural log of enrollment.

Table G.2. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Medicare Advantage Part D Bids

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	–4.38	4.51	–13.14	4.71	0.318	300
2021	–2.61	3.13	–8.89	3.38	0.361	755
2022	–4.76	2.91	–10.58	0.83	0.092	1,465
2023	–10.53	4.25	–18.61	–1.93	0.016	1,862
Hospice						
2021	–20.39	7.74	–36.03	–5.36	0.005	90
2022	–15.51	7.58	–31.32	–0.83	0.037	193
2023	–10.98	8.37	–27.44	4.98	0.177	240

SOURCE: RAND analysis of CMS and other data.

Table G.3. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Medicare Advantage Bids

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	–6.89	4.49	–15.81	2.20	0.122	301
2021	–12.20	3.18	–18.59	–6.15	<0.001	750
2022	–10.17	3.00	–16.26	–4.47	<0.001	1,476
2023	–12.34	4.47	–21.08	–3.61	0.009	1,856
Hospice						
2021	–26.93	7.17	–41.26	–12.65	<0.001	90
2022	–16.85	6.97	–31.27	–3.27	0.012	192
2023	–9.75	8.37	–26.43	5.86	0.224	246

SOURCE: RAND analysis of CMS and other data.

Table G.4. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Part D Bids

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	2.97	1.24	0.59	5.41	0.015	311
2021	8.52	0.80	6.99	10.13	<0.001	804
2022	4.83	1.07	2.72	6.83	<0.001	1,555
2023	1.48	1.18	–0.86	3.79	0.208	1,812
Hospice						
2021	3.24	1.71	0.12	6.65	0.045	91
2022	1.34	1.63	–1.84	4.55	0.395	204
2023	–0.06	2.15	–4.19	4.25	0.999	242

SOURCE: RAND analysis of CMS and other data.

Table G.5. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Total Costs to CMS

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	6.75	8.33	–9.59	23.16	0.448	309
2021	28.56	8.85	11.40	45.37	<0.001	773
2022	24.64	9.99	5.80	44.29	0.014	1,473
Hospice						
2021	6.69	19.14	–28.56	45.14	0.701	90
2022	–12.53	17.59	–45.31	22.25	0.525	197

SOURCE: RAND analysis of CMS and other data.

Table G.6. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Medicare Advantage Costs to CMS

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	1.73	7.92	–14.08	16.85	0.850	308
2021	21.96	7.30	7.32	35.94	0.002	776
2022	19.69	8.62	3.08	36.27	0.021	1,482
Hospice						
2021	9.72	17.85	–23.79	46.18	0.561	90
2022	–11.37	15.46	–40.43	19.01	0.505	188

SOURCE: RAND analysis of CMS and other data.

Table G.7. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Medicare Advantage Rebate

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	6.19	3.38	–0.42	12.73	0.065	299
2021	17.28	3.00	11.73	23.20	<0.001	737
2022	16.57	2.70	11.22	21.85	<0.001	1,598
2023	22.94	4.18	14.44	30.80	<0.001	1,830
Hospice						
2021	25.78	8.44	10.32	43.32	0.001	92
2022	9.12	5.88	–2.37	21.08	0.102	195
2023	12.48	8.39	–3.39	29.53	0.124	235

SOURCE: RAND analysis of CMS and other data.

Table G.8. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Final Medicare Advantage Risk Score

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	0.01	0.01	–0.01	0.03	0.431	306
2021	0.02	0.01	0.00	0.03	0.039	776
2022	0.02	0.01	0.00	0.04	0.025	1,467
Hospice						
2021	0.02	0.02	–0.02	0.07	0.292	91
2022	0.00	0.02	–0.03	0.03	0.830	200

SOURCE: RAND analysis of CMS and other data.

Table G.9. Estimated Association Between VBID Participation and Part D Costs to Centers for Medicare & Medicaid Services

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	1.23	2.89	–4.44	6.89	0.684	306
2021	7.85	2.32	3.31	12.24	<0.001	768
2022	5.10	3.45	–1.58	11.86	0.137	1,493
Hospice						
2021	–3.68	3.02	–9.96	1.90	0.216	90
2022	1.59	3.57	–5.02	9.11	0.638	325

SOURCE: RAND analysis of CMS and other data.

Table G.10. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Total Premiums

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	0.02	0.89	–1.72	1.77	0.962	304
2021	2.14	1.09	0.16	4.36	0.036	825
2022	0.80	0.84	–0.81	2.47	0.326	1,474
2023	0.65	1.00	–1.23	2.68	0.513	1,850
Hospice						
2021	–4.46	2.34	–9.24	–0.12	0.043	91
2022	–1.00	1.60	–4.14	2.09	0.570	206
2023	–1.24	2.71	–6.34	4.13	0.665	237

SOURCE: RAND analysis of CMS and other data.

Table G.11. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Medicare Advantage Premiums

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	–0.11	0.84	–1.80	1.42	0.941	309
2021	–1.86	0.66	–3.18	–0.58	0.004	746
2022	–0.56	0.47	–1.44	0.41	0.246	1,583
2023	0.84	0.70	–0.50	2.27	0.232	1,854
Hospice						
2021	–1.41	1.99	–5.35	2.36	0.468	90
2022	–0.01	1.13	–2.14	2.25	0.988	208
2023	1.93	2.05	–2.06	5.99	0.346	568

SOURCE: RAND analysis of CMS and other data.

Table G.12. Estimated Association Between VBID Participation and Part D Premiums

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Part D Basic Premium						
Year						
VBID General						
2020	1.03	0.78	–0.35	2.67	0.168	304
2021	2.87	0.53	1.86	3.87	<0.001	815
2022	0.70	0.81	–1.03	2.09	0.404	1,564
2023	0.19	0.80	–1.29	1.82	0.770	1,882
Hospice						
2021	–3.61	1.71	–7.13	–0.26	0.032	99
2022	–1.63	1.36	–4.28	1.07	0.244	222
2023	–4.07	1.73	–7.48	–0.77	0.014	234
Part D Supplemental Premium						
Year						
VBID General						
2020	0.31	0.67	–1.13	1.45	0.549	304
2021	0.10	0.21	–0.30	0.51	0.629	815
2022	0.34	0.57	–0.35	1.67	0.704	1,564
2023	–0.02	0.16	–0.32	0.30	0.866	1,882
Hospice						
2021	1.06	1.59	–1.98	4.32	0.518	99
2022	0.77	1.01	–1.13	2.88	0.438	222
2023	–0.55	0.68	–1.89	0.84	0.407	234
Part D Total Premium						
Year						
VBID General						
2020	0.36	0.57	–0.79	1.47	0.516	305
2021	3.05	0.57	1.93	4.18	<0.001	800
2022	1.43	0.59	0.28	2.57	0.009	1,528
2023	0.56	0.79	–0.92	2.17	0.463	1,939
Hospice						
2021	–3.41	1.80	–6.93	–0.05	0.046	90
2022	–1.85	1.41	–4.50	0.84	0.187	204
2023	–3.40	1.54	–6.35	–0.43	0.027	348

SOURCE: RAND analysis of CMS and other data.

Table G.13. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Number of Supplemental Benefits Offered

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	–0.40	0.26	–0.93	0.10	0.128	300
2021	–1.54	0.21	–1.95	–1.13	<0.001	739
2022	–1.10	0.18	–1.45	–0.77	<0.001	1,662
2023	–0.68	0.24	–1.16	–0.24	0.003	1,850
Hospice						
2021	0.12	0.52	–0.92	1.13	0.852	92
2022	0.74	0.43	–0.12	1.57	0.094	205
2023	0.57	0.43	–0.28	1.44	0.193	231

SOURCE: RAND analysis of CMS and other data.

Supplementary Results on Components of Plan-Level Outcomes

Tables G.14 though G.16 report results for additional outcomes, including several outcomes that are not described in the main text.

Table G.14 shows changes in four quantities that determine the standardized MA bid. The standardized MA bid is the sum of three components (after adjustment from the plan's risk factor to a risk factor of 1.0): the net PMPM cost to the plan of Medicare-covered services ("Medicare-covered net PMPM"), nonbenefit expenses allocated to Medicare-covered services, and the gain/loss amount allocated to MA-covered services. The table also shows the projected MA risk score submitted with the MA bid, which is used to convert the plan's bid at its risk factor to the standardized bid.

Table G.14. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Medicare Advantage Bid Components

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Medicare-covered net PMPM						
Year						
VBID General						
2020	–9.82	8.73	–27.29	6.80	0.258	304
2021	20.91	8.04	5.37	36.61	0.006	815
2022	2.98	7.06	–11.16	16.50	0.714	1,564
2023	–24.22	4.97	–33.52	–14.08	<0.001	1,882
Hospice						
2021	–30.93	11.45	–52.05	–8.15	0.004	99
2022	–16.22	11.33	–38.32	5.28	0.147	222
2023	–7.48	8.95	–25.57	9.45	0.359	234
Nonbenefit expenses allocated to Medicare-covered services						
Year						
VBID General						
2020	3.62	3.82	–2.98	12.17	0.340	304
2021	–1.01	2.05	–4.90	3.26	0.594	815
2022	0.88	1.95	–2.92	4.56	0.622	1,564
2023	0.34	2.68	–4.64	5.95	0.958	1,882
Hospice						
2021	–0.38	5.07	–10.20	9.75	0.992	99
2022	–7.76	2.76	–12.95	–2.11	0.008	222
2023	–4.15	4.01	–12.22	3.48	0.308	234
Gain/loss allocated to Medicare-covered services						
Year						
VBID General						
2020	–2.88	6.96	–17.37	10.03	0.683	304
2021	2.05	4.93	–7.24	12.03	0.679	815
2022	16.07	4.20	7.78	24.36	<0.001	1,564
2023	16.28	4.19	8.27	24.60	<0.001	1,882
Hospice						
2021	–15.06	13.36	–40.49	11.12	0.305	99
2022	8.59	6.85	–3.94	22.64	0.193	222
2023	–3.62	7.70	–18.18	11.69	0.679	234
Projected MA risk score						
Year						
VBID General						
2020	0.03	0.01	0.01	0.05	0.003	268
2021	0.04	0.01	0.02	0.06	<0.001	822
2022	0.03	0.01	0.01	0.04	<0.001	1,457
2023	0.02	0.02	–0.01	0.05	0.212	1,848

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Hospice						
2021	–0.03	0.04	–0.10	0.05	0.550	90
2022	–0.01	0.02	–0.05	0.03	0.668	210
2023	–0.01	0.03	–0.06	0.05	0.839	256

SOURCE: RAND analysis of CMS and other data.

Table G.15 shows changes in four quantities that determine the standardized Part D bid. The standardized Part D bid is the sum of three components (after adjustment from the plan's risk factor to a risk factor of 1.0): the net PMPM cost to the plan of standard Part D coverage, nonbenefit expenses allocated to standard Part D coverage, and the gain/loss amount allocated to standard Part D coverage. The table also shows the projected Part D risk score submitted with the Part D bid, which is used to convert the plan's bid at its risk factor to the standardized bid.

Table G.15. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Part D Bid Components

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Part D standard coverage net PMPM						
Year						
VBID General						
2020	3.46	1.31	0.87	5.97	0.008	304
2021	9.01	1.22	6.74	11.49	<0.001	815
2022	5.49	2.78	0.69	11.46	0.013	1,564
2023	–1.45	2.55	–6.20	3.77	0.571	1,882
Hospice						
2021	2.20	1.82	–1.17	5.85	0.214	99
2022	3.77	3.57	–2.19	11.92	0.261	222
2023	0.99	2.58	–4.02	6.41	0.671	234
Part D basic nonbenefit expense						
Year						
VBID General						
2020	1.81	0.48	0.88	2.77	<0.001	304
2021	1.58	0.28	1.02	2.13	<0.001	815
2022	0.81	0.82	–1.04	2.03	0.352	1,564
2023	2.31	0.73	0.88	3.64	<0.001	1,882
Hospice						
2021	–0.02	0.53	–1.12	0.98	0.994	99
2022	0.12	0.58	–0.87	1.38	0.876	222
2023	1.32	1.21	–0.92	3.81	0.269	234
Part D basic gain-loss						
Year						
VBID General						

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
2020	0.15	0.29	–0.41	0.74	0.625	304
2021	–0.94	0.21	–1.37	–0.54	<0.001	815
2022	0.10	0.21	–0.31	0.52	0.667	1,564
2023	–1.38	0.37	–2.12	–0.68	<0.001	1,882
Hospice						
2021	0.16	0.84	–1.53	1.68	0.754	99
2022	1.03	0.37	0.34	1.76	0.002	222
2023	–0.92	0.58	–2.13	0.13	0.091	234
Part D projected risk score						
Year						
VBID General						
2020	0.00	0.01	–0.02	0.01	0.704	304
2021	–0.03	0.01	–0.04	–0.01	<0.001	815
2022	0.01	0.01	–0.01	0.02	0.230	1,564
2023	0.01	0.02	–0.02	0.04	0.521	1,882
Hospice						
2021	0.00	0.02	–0.04	0.04	0.874	99
2022	0.01	0.02	–0.02	0.04	0.586	222
2023	0.00	0.03	–0.05	0.05	0.984	234

SOURCE: RAND analysis of CMS and other data.

Table G.16 shows changes associated with VBID in four variables that determine the Part D cost to CMS. The Direct Subsidy is a monthly capitation payment to the plan based on the plan's bid and beneficiary risk scores. The Low-Income Subsidy (LIS), which comprises the Low-Income Cost-Sharing Subsidy (LICS) and the Low-Income Premium Subsidy (LIPS), represents the PMPM amount paid by CMS to subsidize coverage and prescription drug utilization for low-income beneficiaries. Reinsurance reflects the PMPM amount paid to plans via individual reinsurance for beneficiary drug costs in the catastrophic phase of the Part D benefit. The realized Part D risk score is the average risk score for plan enrollees.

Table G.16. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Part D Cost to CMS Components

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Direct subsidy						
Year						
VBID General						
2020	0.73	0.76	–0.78	2.15	0.359	304
2021	2.26	0.58	1.14	3.41	<0.001	815
2022	0.77	0.91	–1.22	2.26	0.414	1,564
Hospice						
2021	–1.52	1.10	–3.69	0.61	0.171	99
2022	0.49	0.73	–0.90	1.88	0.486	222
LIS (LICS + LIPS)						
Year						
VBID General						
2020	–2.15	1.71	–5.47	1.12	0.220	304
2021	3.54	1.70	0.11	6.79	0.043	815
2022	3.99	1.85	0.56	7.63	0.028	1,564
Hospice						
2021	–2.79	1.12	–4.99	–0.62	0.013	99
2022	0.74	1.87	–2.72	4.47	0.714	222
Reinsurance						
Year						
VBID General						
2020	1.74	2.12	–2.16	6.15	0.442	304
2021	1.28	1.27	–1.20	3.77	0.326	815
2022	0.28	2.26	–4.29	4.69	0.865	1,564
Hospice						
2021	0.35	2.08	–3.57	4.35	0.868	99
2022	1.56	2.31	–2.70	6.39	0.474	222
Realized Part D Risk Score						
Year						
VBID General						
2020	0.00	0.01	–0.02	0.01	0.682	304
2021	0.01	0.01	0.00	0.02	0.016	815
2022	0.02	0.01	0.01	0.04	0.002	1,564
Hospice						
2021	0.01	0.02	–0.02	0.04	0.566	99
2022	0.01	0.02	–0.02	0.04	0.390	222

SOURCE: RAND analysis of CMS and other data.

Enrollment-Weighted Results

Tables G.17 through G.20 provide enrollment weighted results for the main outcomes reported in Chapter 5.

Table G.17. Estimated Association Between VBID Participation and Medicare Advantage Part D Bids (enrollment-weighted)

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
MAPD Bid						
Year						
VBID General						
2020	–10.08	6.72	–23.16	2.85	0.140	216
2021	–3.06	4.27	–11.56	5.11	0.456	506
2022	–6.72	3.82	–14.46	0.71	0.070	648
2023	–22.73	6.90	–35.56	–7.96	0.002	892
Hospice						
2021	–18.66	8.59	–34.75	–0.94	0.040	41
2022	–15.18	7.43	–29.94	–0.89	0.038	135
2023	–16.56	8.32	–33.16	0.30	0.053	170
MA Bid						
Year						
VBID General						
2020	–12.00	6.40	–24.36	0.58	0.062	216
2021	–12.61	4.28	–21.15	–4.51	0.005	493
2022	–12.61	3.89	–20.71	–5.02	0.001	646
2023	–24.61	7.24	–37.95	–9.83	0.001	901
Hospice						
2021	–22.19	7.71	–36.37	–6.12	0.006	41
2022	–14.82	7.06	–29.05	–1.21	0.032	138
2023	–15.30	8.95	–32.40	3.33	0.093	167
Part D Bid						
Year						
VBID General						
2020	2.11	1.64	–1.03	5.39	0.188	219
2021	7.68	0.99	5.70	9.63	<0.001	440
2022	5.10	1.25	2.70	7.58	<0.001	613
2023	0.68	1.83	–2.67	4.36	0.649	955
Hospice						
2021	1.45	2.01	–2.26	5.62	0.428	40
2022	0.79	1.88	–2.92	4.48	0.633	144
2023	–0.10	2.23	–4.35	4.31	0.995	179

SOURCE: RAND analysis of CMS and other data.

Table G.18. Estimated Association Between VBID Participation and Medicare Advantage Part D Costs (enrollment-weighted)

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
MAPD Cost						
Year						
VBID General						
2020	–8.24	11.16	–30.33	13.14	0.482	215
2021	20.47	11.86	–2.77	43.00	0.089	409
2022	28.24	11.76	5.40	50.39	0.014	659
Hospice						
2021	–9.33	19.41	–47.01	29.68	0.699	41
2022	–25.89	19.85	–63.87	13.14	0.230	140
MA Cost						
Year						
VBID General						
2020	–8.44	11.22	–30.36	13.56	0.464	217
2021	12.22	10.08	–8.34	31.47	0.240	408
2022	19.93	9.59	1.24	38.65	0.038	669
Hospice						
2021	–7.86	18.31	–42.65	28.68	0.729	41
2022	–24.78	15.59	–54.57	5.67	0.124	137
Part D Cost						
Year						
VBID General						
2020	–2.91	3.47	–10.20	3.43	0.403	215
2021	6.63	3.45	0.05	13.41	0.049	418
2022	4.05	4.11	–3.82	12.02	0.329	666
Hospice						
2021	–1.56	2.95	–7.12	4.35	0.637	41
2022	–0.71	6.59	–12.91	12.75	0.930	135

SOURCE: RAND analysis of CMS and other data.

Table G.19. Estimated Association Between VBID Participation and Medicare Advantage Part D Premium (enrollment-weighted)

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
MAPD Premium						
Year						
VBID General						
2020	–0.50	0.98	–2.59	1.19	0.661	215
2021	2.43	1.34	0.13	5.42	0.037	408
2022	0.52	1.45	–2.00	3.62	0.731	688
2023	0.49	1.23	–1.96	2.96	0.664	945
Hospice						
2021	–3.73	1.84	–8.07	–0.93	0.003	41

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
2022	–0.06	1.37	–2.82	2.50	0.978	140
2023	0.06	2.38	–4.68	4.64	0.998	157
MA Premium						
Year						
VBID General						
2020	–0.87	1.11	–3.13	1.10	0.463	214
2021	–0.69	0.73	–2.12	0.69	0.336	398
2022	–0.04	0.51	–1.02	0.92	0.958	664
2023	1.12	0.71	–0.33	2.50	0.123	1,051
Hospice						
2021	0.52	1.49	–2.49	3.36	0.743	41
2022	0.55	0.84	–1.10	2.26	0.529	144
2023	1.17	1.93	–2.53	5.07	0.520	227
Part D Total Premium						
Year						
VBID General						
2020	0.75	0.91	–1.05	2.56	0.412	213
2021	2.51	0.76	1.04	3.98	<0.001	456
2022	0.18	0.83	–1.45	1.77	0.791	634
2023	–0.13	1.04	–2.13	1.98	0.958	960
Hospice						
2021	–4.41	1.76	–8.17	–1.38	0.002	41
2022	–0.78	1.46	-3.74	1.97	0.600	136
2023	–1.41	1.47	–4.30	1.48	0.313	153

SOURCE: RAND analysis of CMS and other data.

Table G.20. Estimated Association Between Participation in VBID General or Hospice Benefit Component and Number of Supplemental Benefits Offered (enrollment-weighted)

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
VBID General						
2020	–0.15	0.51	–1.11	0.88	0.708	212
2021	–1.09	0.36	–1.81	–0.41	<0.001	401
2022	–0.76	0.25	–1.26	–0.27	0.003	629
2023	–0.53	0.32	–1.18	0.05	0.077	927
Hospice						
2021	–0.47	0.63	–1.76	0.75	0.430	42
2022	0.86	0.58	–0.29	1.96	0.146	147
2023	0.78	0.58	–0.40	1.91	0.212	192

SOURCE: RAND analysis of CMS and other data.

Appendix H. Beneficiary-Level VBID General Outcomes

Tables H.1 reproduces the beneficiary-level findings described in the main text in tabular form. We provide these tables so that readers can easily access the point estimate, standard error, CI, p-value, and ESS.

Table H.1. Estimated Associations Between Participation in VBID General and Beneficiary-Level Outcomes, 2020 and 2021

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Targeted beneficiary risk scores						
2020	0.055	0.013	0.029	0.080	<0.001	318,205
2021	0.070	0.015	0.041	0.100	<0.001	3,683,191
Adherence to diabetes Medication						
2020	0.011	0.003	0.004	0.018	0.001	89,266
2021	0.004	0.003	–0.001	0.009	0.118	605,376
Adherence to cholesterol medication						
2020	0.012	0.004	0.004	0.021	0.004	219,061
2021	0.004	0.002	0.000	0.008	0.041	1,706,587
Adherence to hypertension medication						
2020	–0.001	0.004	–0.010	0.007	0.790	168,650
2021	0.002	0.002	–0.002	0.005	0.285	1,510,052
Adherence to breast cancer screening recommendations						
2020	0.026	0.012	0.003	0.050	0.027	43,287
2021	0.023	0.010	0.003	0.043	0.025	497,851
Inpatient utilization						
2020	0.035	0.009	0.018	0.052	<0.001	318,205
2021	0.036	0.009	0.018	0.053	<0.001	3,683,191
Beneficiary OOP drug costs ($)						
2020	1.11	4.25	–7.21	9.43	0.794	318,205
2021	–24.59	4.28	–32.99	–16.20	<0.001	3,683,191

SOURCE: RAND analysis of CMS and other data. Analyses correspond to those reported in the main text. Risk scores are reported in levels; adherence measures represent percentage point changes; inpatient utilization reflects percentage changes; beneficiary drug costs are reported in dollars. All estimates are from stratified, entropy-balanced DD regressions estimated separately for each year. Inpatient regressions were fitted using Poisson models.

Table H.2 shows the association between VBID General and ED visits. We did not report this outcome in the main text because, in prior work, we found evidence that the parallel trends assumption does not hold for ED visits (Eibner et al., 2023). Specifically, the number of ED visits reported in the encounter data diverged from the number reported in plan bids in 2020 for VBID plans relative to non-VBID plans. Conceptually, reported ED utilization should be equivalent in encounter data and bid data. However, the GAO and others have raised concerns that the encounter data may be inaccurate (GAO, 2017; MedPAC, 2024). It is possible that VBID plans improved their data reporting for ED visits, or made other related changes, in a manner that corresponded with VBID implementation. This divergence did not occur for inpatient utilization, and our prior evaluation found less reason to be concerned about that outcome. While the analysis below suggests that VBID General was associated with 8.8% and 4.7% increases in ED visits in 2020 and 2021, we cannot conclude that VBID was associated with increased ED use given data challenges.

Table H.2. Estimated Association Between Participation in VBID General and Utilization of Emergency Department Visits

Effect	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Year						
2020	0.088	0.038	0.013	0.162	0.022	318,205
2021	0.047	0.019	0.010	0.085	0.014	3,683,191

SOURCE: RAND analysis of CMS and other data. Estimates represent transformed coefficient from a Poisson model. Analysis is inconclusive due to evidence that encounter data reporting of ED visits in VBID participating plans changed in 2020 relative to nonparticipants.

Appendix I. Contract-Level VBID General Outcomes

This appendix provides additional detail on the changes in Star Ratings methodologies and contract-level Star Ratings results from Chapter 7. This appendix does not include separate results for contracts participating in the hospice component. We conducted these analyses for our last report, but the findings were not significant, likely because there are not enough beneficiaries entering hospice every year to contribute to changes in Star Ratings at the contract level. Furthermore, the Star Ratings do not capture the hospice quality of care experiences.

Changes to Star Ratings Methodology

CMS often makes small changes the methodology for the Star Ratings, which include changes to individual measures, such as changing a measure's weight or removing or adding measures. However, the COVID-19 pandemic prompted CMS to make significant methodological changes to the Star Ratings for the 2019 and 2020 measurement years. To reduce the impact of the COVID-19 pandemic on plans for the 2020 Star Ratings measurement, CMS allowed contracts to use the "better of" methodology, whereby if a measure-level Star Rating was lower than in the previous year, the previous year's value would be used CMS, 2021). For these reasons, we removed these measurement years from our analysis.

For the 2021 measurement year, CMS implemented upper and lower limits ("guardrails") on changes in the cut-points for non-CAHPS measures to improve stability in the measures over time (CMS, 2023b). Cut-points create categories along a measure's distribution, and contracts falling into a category are assigned a specific star value. For example, scoring above 90 on a specific measure yields a 5-star rating. Guardrails will limit the change in the cut-point to 5 percentage points for measures on a 0–100 scale or 5% for measures with other scales.

Outlier removal, prior to the assessment of the cut-points, took effect in 2024 (affecting 2022 measure year Star Ratings) for measures using data other than CAHPS (CMS. It is likely that both of these adjustments limited the number of contracts receiving 5 stars, and thus lowered the overall average Star Rating (CMS, 2023a) for both VBID-participating and nonparticipating contracts. A lawsuit from several MA plans over the way these changes were applied prompted CMS to recalculate the Star Ratings for the 2022 measurement year (Coleman, Duran, and Lazio, 2024). We have used the recalculated Star Ratings for our analyses.

Finally, CMS increased weights on the measures derived from the CAHPS data (patient experience, complaints, and access measures) for the 2021 measurement year.

Contract-Level Analyses

The contract-level analysis for the Star Ratings outcomes follows a similar analysis to the plan-level analysis. Because of the difficulty in achieving balance for the contracts, we used a more limited set of balance characteristics, and we present the unweighted results only, though contract enrollment is included as a balancing characteristic.

Detailed Regression Results

Overall Star Rating

Table I.1 shows the associations between VBID and overall Star Rating for 2021 and 2022 that are also presented in the main report. A contract is considered to be a VBID contract if at least one plan in the contract implemented VBID General. We find evidence of increases in Star Ratings associated with VBID General participation.

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

Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
2021	0.20	0.08	0.04	0.35	0.016	167
2022	0.19	0.07	0.06	0.32	0.004	265

SOURCE: RAND analysis of CMS data.

Overall Star by Level of Contract Exposure to VBID

In Table I.2, we change 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, regardless of whether the beneficiary was targeted for VBID or received VBID benefits. While in general, there is an association between exposure to VBID General and increases in overall Star Rating, there is not a clear dose-response relationship whereby more exposure to VBID leads to greater increases in Star Ratings.

Table I.2. Estimated Association Between Participation in VBID General and Overall Star Rating, by Levels of VBID Exposure within the Participating Contracts

Effect and Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
25% VBID exposed						
2021	0.15	0.10	–0.05	0.33	0.115	120
2022	0.17	0.08	0.02	0.32	0.020	195
50% VBID exposed						
2021	0.15	0.10	–0.05	0.35	0.128	112
2022	0.19	0.09	0.02	0.36	0.039	151
75% VBID exposed						
2021	0.12	0.13	–0.15	0.35	0.333	74
2022	0.24	0.10	0.04	0.44	0.026	106

SOURCE: RAND analysis of CMS data.

Domain-Level Star Ratings

In Table I.3, we report associations between VBID and domain-level Star Ratings that are presented in the main text. We do not find evidence of a statistically significant association with VBID General participation on domain-level Star Ratings, except for managing chronic conditions in both 2021 and 2022, and improved customer service in 2021.

Table I.3. Estimated Association Between Participation in VBID General and Domain-Level Star Ratings

Effect and Year	Estimate	Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value	ESS
Staying Healthy						
2021	0.06	0.08	–0.11	0.21	0.477	170
2022	–0.04	0.10	–0.24	0.17	0.795	279
Managing Chronic Conditions						
2021	0.20	0.08	0.05	0.36	0.012	156
2022	0.21	0.08	0.06	0.37	0.011	284
Member Experience						
2021	0.22	0.14	–0.04	0.48	0.112	176.
2022	0.06	0.14	–0.21	0.34	0.638	270
Member Complaints						
2021	–0.05	0.12	–0.28	0.17	0.703	162
2022	–0.06	0.11	–0.26	0.15	0.630	254
Customer Service						
2021	0.38	0.14	0.10	0.65	0.010	154
2022	0.04	0.12	–0.21	0.29	0.734	278
Drug Safety						
2021	0.04	0.10	–0.15	0.22	0.742	164
2022	0.03	0.08	–0.14	0.19	0.766	268

SOURCE: RAND analysis of CMS data.

Appendix J. Descriptive and Beneficiary-Level Hospice Benefit Component Outcomes

In this appendix, we present information to support the analysis of Hospice Benefit component outcomes, discussed in Chapters 9 and 11 in the main report.

In- and Out-of-Network Hospice Characteristics

To better understand characteristics of the hospices in model-participating plans' service areas, in Chapter 9, we describe in-network and OON hospices, comparing in-network and OON hospices that served at least one VBID beneficiary. We conducted statistical significance testing via logistic, multinomial logistic, or linear regression models in which the outcomes were the characteristics of interest and the predictors included indicators for hospice network status.

Outcome Measure Definitions

As described in the 2023 Evaluation Report (Eibner et al., 2023b), we evaluated hospice care patterns (hospice enrollment, hospice length of stay (LOS), proportion of beneficiaries who discharged alive from hospice, and proportion of beneficiaries who received visits from professional hospice staff in the last three days of life). LOS in hospice is an important outcome for two reasons: A short LOS (operationalized here in two ways commonly used in the literature: less than three days and less than seven days; Forst et al., 2018; Teno et al., 2012) indicates insufficient time for patients and families to fully realize the benefits of hospice (Rickerson et al., 2005), whereas a very long LOS (operationalized here as more than 180 days) may be an indicator of inappropriate enrollment of patients into hospice based on the hospice eligibility criterion of a life expectancy of six months or less (Wachterman et al., 2011). Although being discharged from hospice alive can be a positive outcome for patients whose quality of life and prognosis improved such that they no longer need hospice services, high rates of live discharge can also indicate inappropriate overenrollment of beneficiaries in hospice. Having professional visits in at least two of the last three days of life is an established quality indicator for hospice care delivery (Teno et al., 2016). Caregiver-reported hospice care experiences, measured here by a weighted average of eight CAHPS Hospice Survey measures (Anhang Price et al., 2018), reflect the degree to which care is patient- and family-centered, a core aspect of hospice care quality.

Table J.1 lists all outcomes assessed in regression analyses and their corresponding denominators. Some outcome variables had additional eligibility requirements and restrictions. For both VBID and comparison beneficiaries, all outcome variables were obtained from FFS

hospice claims, with the exception of the CAHPS Hospice Survey summary measure score, which was obtained from CAHPS Hospice Survey responses. Because these responses do not contain beneficiary names or identification numbers, we linked CAHPS Hospice Survey responses to hospice claims data by matching based on available variables (hospice CMS Certification Number [CCN], beneficiary date of death, date of birth, hospice admission date, primary diagnosis, and sex). Match rates for the 2022 claims were similar to previous years (98.31% among eligible respondents with Medicare).

The CAHPS Hospice Survey is administered to family caregivers after the death of a hospice patient; CMS requires that hospices meeting eligibility criteria contract with a survey vendor to collect CAHPS Hospice Survey data as part of their participation in the Hospice Quality Reporting Program. CAHPS Hospice Survey measures are endorsed by CMS' Consensus-Based Entity, the Partnership for Quality Measurement, and assess aspects of care important to hospice patients and their families, including hospice team communication, timeliness of care, respectful treatment, help for pain and other symptoms, emotional and spiritual support, and training the family to care for hospice patients at home ("CAHPS® Hospice Survey," 2024). Summary CAHPS Hospice Survey measure scores were calculated for each beneficiary by averaging each beneficiary's score across CAHPS measures. In this calculation, the six composite measures assessing specific aspects of care experience received equal weight, whereas the two global assessment measures, overall rating and willingness to recommend, each received half weight, as both are overall assessments of care delivered by the hospice. For beneficiaries that were missing scores for a given measure, mean scores within year were imputed for the measure. Scores were adjusted for mode of survey administration, and weights for these outcomes additionally accounted for differences in case-mix using the following variables, in keeping with CMS guidance for adjustment of CAHPS Hospice Survey measure scores: decedent age, payer for hospice care (including payers in addition to Medicare listed in the hospice administrative record), primary diagnosis, and length of final episode of hospice care; respondent age, education, relationship to caregiver, language spoken at home, and survey language; and response percentile (the length of lag time between decedent death and survey response) (CMS, 2024).

Table J.1. Denominators for Outcome Variables

Outcome Measure	Denominator
Hospice enrollment in the year of death	All decedents
Length of final episode of hospice care (days)	Hospice decedents
Final LOS less than three days	Hospice decedents
Final LOS less than seven days	Hospice decedents
Final LOS more than 180 days	Hospice decedents
Professional visits in at least 2 of last 3 days of life	Hospice decedents
Summary CAHPS Hospice Survey score	Hospice decedents
Any live discharges from hospice in the given year	Hospice enrollees
Transfer from hospice in the given year	Hospice enrollees
Revocation in the given year	Hospice enrollees
Death within 30 days of a live discharge but before the end of the calendar year	Hospice enrollees who had a live discharge
Transfer to another hospice within 7 days of a live discharge but before the end of the calendar year 2023	Hospice enrollees who had a live discharge

Entropy Balancing and Covariate Balance

As discussed in Appendix C, to estimate the 2021 treatment effect, our DD analyses weight each of five groups (pre-period Hospice-participating, pre-period nonparticipating, first post-period nonparticipating, second post-period nonparticipating, and second post-period participating) beneficiaries to the 2021 post-period participating group. Similarly, the 2022 effect for plans that joined the Hospice Benefit component in 2022 weights 2019 and 2021 participating and nonparticipating groups to look like the 2022 participating group, and also weights the 2022 nonparticipating group to look like the target 2022 participating group. The 2022 treatment effect is then estimated as a combination of the 2022 estimated effect for beneficiaries in plans that joined the Hospice Benefit component in 2021, and the estimated effect for those that joined in 2022, where the weights are equal to the number of beneficiaries observed in each type of Hospice-participating plan in 2022. We use this balancing approach so that compositional changes in these groups over time are not interpreted as treatment effects. This section gives information on balance between the Hospice-participating and nonparticipating plans in our outcomes analyses.

In the interest of brevity, we do not present full balance tables for each outcome, but 10 sets of EB weights were fit for each primary outcome model presented in the Hospice section to account for any differences in population definitions and item nonresponse in the outcomes. Table J.2 summarizes weights to estimate the 2022 treatment effect for plans that joined the Hospice Benefit component in 2022 for the hospice enrollment outcome. As with VBID General beneficiary-level models, we use SMDs as our primary measure of balance. In our case, we calculate the mean of a covariate among the post–Hospice Benefit component-participating group, subtract the weighted mean of one of the other groups (for example, 2019 pre–Hospice Benefit component-nonparticipating), and divide that difference by the standard deviation of the

covariate among 2022 post–Hospice Benefit component participating individuals. Because the final DD estimate uses differences involving all combinations of pre- and post-VBID and participating- and nonparticipating, lack of balance for any of the SMDs for a given covariate can potentially lead to bias in the treatment effect estimate.

In general, it was difficult to achieve balance between the Hospice-participating and comparison groups. Two important considerations for these analyses were that (1) relatively few beneficiaries were in Hospice-participating plans (and relatively few plans were Hospice-participating, especially in 2021) and (2) the distribution of pretreatment characteristics was very different between the participating and nonparticipating groups. For example, in 2021, a majority of the Hospice-participating beneficiaries were in Puerto Rico, though there are similar challenges balancing the 2022 data as well. The relatively small number of participating beneficiaries reduces the statistical power relative to what would have been expected if the data had been better balanced between participating and nonparticipating plans, and the small number of participating plans makes balancing plan-level characteristics from the pre- to post-period difficult. Furthermore, because the of the strong differences in covariate distributions between the Hospice-participating and nonparticipating groups, it is difficult to achieve good balance between participating and nonparticipating plans.

In many cases, balance at the 0.1 or even 0.2 level of SMD was not achieved. For the hospice enrollment outcome in particular, the weighting did not reduce the SMDs as much as hoped and it also created year-to-year fluctuations in the pre-period and control data that called into question the parallel tends assumption. For that reason, we balanced on a subset of variables for that outcome, and controlled for the variables that were not brought into balance in the outcomes model. See Table J.2. The other outcomes balance the same characteristics except for the CAHPS outcome, which additionally includes CAHPS Hospice case mix adjustors as balancing characteristics (CMS, 2023a; CMS, 2023b).

Even given relatively wide allowances for imbalance and/or not balancing on some characteristics, the ESSs are drastically lower than the nominal sample sizes with some ESSs below 100 (Table J.3). The reductions in ESS result in relatively wide CIs (especially after accounting for plan-level fixed effects). In cases where we were simply unable to balance a characteristic between two groups, the uncertainty is not fully represented in the CIs. In a most extreme case where a characteristic is perfectly correlated with the treatment group indicator, we are unable to say whether that characteristic or the Hospice Benefit component participation is driving the difference. Balance summaries for other outcomes are available in Table J.4.

Table J.2. Descriptive Statistics of Select Balancing Variables for Beneficiaries in Newly-Participating Plans in 2022 for the Hospice Enrollment Outcome

Variable (frequency, % unless otherwise noted)	Comparison Group 2022 SMD	Hospice-Participating Group 2021 Pre-VBID SMD	Comparison Group 2021 SMD	Hospice Participating Group 2019 Pre-VBID SMD	Comparison Group 2019 SMD
Area-level income	–0.10	–0.20	–0.20	0.08	–0.15
MA penetration	0.20	0.20	0.04	0.20	0.19
Urbanicity	–0.12	0.06	0.10	–0.07	–0.12
HPSA	–0.20	–0.20	–0.13	–0.05	–0.19
Percentage over age 65	–0.20	–0.19	–0.18	0.11	0.20
Puerto Rico county	0.00	0.12	0.11	0.15	0.15
SDI	0.08	0.12	0.17	–0.12	0.17
Social Deprivation Score	0.08	0.12	0.17	–0.12	0.17
Percentage who did not work, ages 16–64	0.13	0.14	0.16	–0.03	0.20
Percentage disabled, ages 18–64, civil noninstitutionalized	0.20	0.20	0.20	0.10	0.10
Total (MAPD + PDP) LIS enrollees in 2021 as % of total Medicare enrollment	–0.20	–0.11	–0.10	–0.20	0.10
MAPD LIS enrollees in 2021 as percentage of total Medicare enrollment	0.19	0.17	0.04	0.14	0.20
BCBS affiliate	0.13	0.08	0.08	0.00	–0.19
For-profit status	–0.10	–0.03	–0.01	0.02	0.20
For profit beneficiary months*	–0.10	–0.03	–0.01	0.02	0.20
MA penetration rate	1.65	0.59	–0.02	1.63	0.56
Median income*	0.20	0.20	0.20	–0.20	–0.20
PO enrollment*	0.65	0.50	0.50	0.36	0.20
Star Rating (overall)*	0.78	0.61	–0.49	1.25	0.93
Enrollment	–1.10	–0.58	–0.58	0.45	0.14
Part C cost to CMS	–0.10	–0.05	–0.15	0.06	0.20
Bids – MA*	0.72	0.03	–0.19	0.49	0.06
MA premiums	–0.01	–0.01	0.00	–0.20	0.05
$0 premium plan	–0.20	–0.20	–0.10	–0.20	–0.15
Cost of MSB	0.20	0.20	0.14	0.20	0.18
Rebate dollars amount*	0.75	0.56	0.23	0.95	0.48
Administrative costs (bid data)	–0.18	–0.09	–0.05	–0.03	0.18
OOP maximum (Part C)	0.20	0.19	0.20	–0.10	–0.17
PDSS participant*	1.61	0.86	0.64	1.61	0.53
Type of plan	0.20	0.13	0.10	0.20	–0.16
SNP type (C, D, I)	–0.20	–0.20	–0.20	0.03	0.03
SNP type (D)*	–0.34	–0.16	–0.07	–0.43	–0.03
SNP type (I)	–0.20	–0.10	0.00	0.13	0.05
No-bonus county*	–0.23	–0.06	0.18	0.24	–0.57

Variable (frequency, % unless otherwise noted)	Comparison Group 2022 SMD	Hospice-Participating Group 2021 Pre-VBID SMD	Comparison Group 2021 SMD	Hospice Participating Group 2019 Pre-VBID SMD	Comparison Group 2019 SMD
Single-bonus county	0.20	0.20	0.20	0.20	0.20
Double-bonus county*	0.03	–0.08	–0.22	–0.26	0.23
Age	–0.05	0.06	–0.06	0.08	0.20
Male	–0.01	–0.01	–0.01	0.01	0.04
Dual	–0.2	–0.19	–0.07	–0.20	–0.06
LIS status, Level 1	0.06	0.02	0.06	–0.20	0.00
LIS, Level 2	–0.20	–0.20	–0.09	–0.11	–0.12
LIS, Level 3	–0.20	–0.19	–0.10	0.06	0.02
LIS, Level 4	0.04	0.03	0.00	0.04	0.05
Disabled	0.03	–0.04	0.04	–0.11	–0.07
ESRD	0.06	0.03	0.01	0.07	–0.05
Beneficiary risk score	–0.18	–0.03	–0.04	–0.14	0.14
HCC HIV/AIDS	–0.03	–0.07	–0.01	0.02	0.06
HCC acute leukemia	–0.20	0.20	–0.01	–0.20	0.20
HCC lung cancer	–0.09	0.15	–0.01	–0.09	0.19
HCC lymphoma	–0.01	0.14	–0.01	0.01	0.15
HCC colorectal cancer	0.01	0.09	–0.01	0.02	0.16
HCC breast cancer	0.06	0.20	0.00	0.07	0.22
HCC DM with acute complications	–0.05	0.09	0.00	–0.04	0.11
HCC DM with chronic complications*	0.04	0.53	0.01	0.02	0.64
HCC DM without complications	–0.03	0.17	–0.01	0.01	0.24
HCC end stage liver disease	–0.15	0.07	0.01	–0.14	0.15
HCC cirrhosis	–0.01	0.03	0.00	0.00	0.12
HCC chronic hepatitis	–0.01	–0.03	0.00	0.00	–0.03
HCC RA	0.07	0.20	0.01	0.06	0.20
HCC schizophrenia	–0.03	0.14	0.00	0.00	0.14
HCC depression/bipolar	0.08	0.20	0.00	0.11	0.20
HCC CHF*	–0.59	0.45	–0.02	–0.56	0.44
HCC MI	–0.20	0.20	0.00	–0.19	0.26
HCC ACS	–0.10	0.15	–0.03	–0.04	0.15
HCC angina pectoris*	0.09	0.28	0.03	0.07	0.28
HCC arrythmias	–0.16	0.20	0.00	–0.13	0.20
HCC ICH*	–0.25	–1.05	–0.01	–0.27	–0.36
HCC stroke	–0.07	0.09	0.01	–0.11	0.20
HCC vascular disease with complications	–0.13	0.20	–0.02	–0.06	0.22
HCC vascular disease*	0.02	0.71	–0.06	0.06	0.70
HCC COPD*	–0.08	0.71	0.00	–0.11	0.71
HCC acute renal failure	–0.47	0.39	–0.02	–0.47	0.33
HCC CKD Stage 5	0.00	0.04	–0.01	0.02	0.08
HCC CKD Stage 4	0.09	0.17	0.02	0.08	0.18

Variable (frequency, % unless otherwise noted)	Comparison Group 2022 SMD	Hospice-Participating Group 2021 Pre-VBID SMD	Comparison Group 2021 SMD	Hospice Participating Group 2019 Pre-VBID SMD	Comparison Group 2019 SMD
RxHCC dementia	–0.09	0.20	–0.02	–0.04	0.20
RxHCC high cholesterol	0.24	1.07	0.02	0.26	1.26
Any HCC	–0.31	0.92	–0.02	–0.28	1.03
Date of death (day of year)	–0.09	–0.12	–0.17	–0.08	–0.09
Number of months continuously enrolled in the plan (in the pre-12 months) that is anchored at death date (whether it's FFS or MA)	–0.04	–0.20	0.01	–0.05	–0.20

SOURCE: RAND analysis of CMS and other data.
NOTE: CKD = chronic kidney disease; DM = disease management; HIV/AIDS = human immunodeficiency virus/acquired immunodeficiency syndrome; ICH = intracerebral hemorrhage; MI = myocardial infarction; PDP = Part D plan; RA = rheumatoid arthritis. Balance measures are calculated after EB weights have been applied. Data elements denoted with an asterisk (*) were not included in the EB weights and were controlled for in the outcomes model.

Table J.3. Effective Sample Sizes Due to Application of Entropy Balancing Weights for Newly Participating Plans in 2022

Outcome Measure	Hospice Participating Group 2022 (nominal *N*)	Comparison Group 2019 ESS (nominal *N*)	Comparison Group 2021 ESS (nominal *N*)	Comparison Group 2022 (nominal *N*)	Hospice participating Group Pre Period 2019 (nominal *N*)	Hospice participating Group Pre Period 2021 (nominal *N*)
Hospice enrollment in the year of death	18,288	131,269 (549,269)	1,768 (788,265)	695,141 (787,406)	10,011 (18,288)	57.3 (16,493)
Length of final episode of hospice care (days)	7,787	677 (283,920)	46 (357,690)	241,372 (373,449)	90 (5,120)	17 (7,522)
Professional visits in at least 2 of last 3 days of life	5,901	427 (190,666)	41 (248,225)	152,171 (265,073)	155 (3,726)	49 (5,614)
Summary CAHPS Hospice Survey score	1,706	1,056 (70,461)	82 (77,044)	43,586 (80,289)	137 (1,342)	1,546 (1,736)
Any live discharges from hospice	12,079	739 (406,794)	53,705 (505,735)	331,318 (531,918)	74 (7734)	161 (10,862)

SOURCE: RAND analysis of CMS and other data.

Table J.4. Measures of Imbalance Following Application of Entropy Balancing Weights for Sensitivity Analysis That Includes All Covariates in Weighting

Outcome Measure	Mean ASMD	Max ASMD	Percentage of Covariates with ASMD Above 0.1	Percentage of Covariates with ASMD Above 0.2
Hospice enrollment in the year of death*	0.12	0.2	55	16
Length of final episode of hospice care (days)	0.16	1.0	52	30
Professional visits in at least 2 of last 3 days of life	0.16	1.0	50	30
Summary CAHPS Hospice Survey score	0.12	1.0	37	17
Any live discharges from hospice	0.14	0.5	54	23
Transfer from hospice in the given year	0.14	0.5	54	23
Revocation in the given year	0.14	0.5	53	23
Death within 30 days of a live discharge	0.15	0.6	55	26
Transfer to another hospice within 7 days of a live discharge	0.15	0.6	55	26

SOURCE: RAND analysis of CMS and other data.

Outcome Summaries

Table J.5 shows weighted and unweighted outcomes for 2019 and 2021 (pre–Hospice Benefit component) and 2022 (post–Hospice Benefit component) for plans that joined in 2022. (Because the 2022 Hospice Benefit component-participating group is our target, no weights are applied to that group). Especially prior to weighting, there are substantial differences between the Hospice-participating and comparison groups at baseline. For example, 41% of decedents in the Hospice-participating group were enrolled in hospice in the year of their death, compared with 53% in the nonparticipating comparison group in that same year. The differences are generally reduced after weights are applied, but substantial differences remain at baseline, meaning that we must rely on DD to resolve any lingering imbalances (whether because of observed or unobserved characteristics) between the participating and comparison groups.

Table J.5. Outcomes of Interest for Hospice Benefit Component Participants and Comparison POs Weighted to Newly Participating in 2022 Covariates, Mean (standard deviation)

Outcome	Weighting	Participating, 2022	Comparison, 2022	Participating, 2021	Comparison, 2021	Participating, 2019	Comparison, 2019
Hospice enrollment in the year of death	Weighted	N/A	0.49 (0.5)	0.35 (0.48)	0.32 (0.47)	0.62 (0.49)	0.55 (0.5)
Length of final episode of hospice care (days)	Weighted	N/A	64.76 (143.31)	62.72 (136.07)	40.44 (103.9)	65.85 (160.9)	62.68 (139.42)
Final LOS less than 3 days	Weighted	N/A	0.13 (0.33)	0.15 (0.36)	0.13 (0.34)	0.07 (0.26)	0.14 (0.34)
Final LOS less than 7 days	Weighted	N/A	0.33 (0.47)	0.28 (0.45)	0.38 (0.49)	0.24 (0.43)	0.36 (0.48)
Final LOS more than 180 days	Weighted	N/A	0.1 (0.3)	0.14 (0.35)	0.05 (0.22)	0.08 (0.27)	0.1 (0.29)
Professional visits in at least 2 of last 3 days of life	Weighted	N/A	0.66 (0.48)	0.72 (0.45)	0.54 (0.5)	0.61 (0.49)	0.59 (0.49)
Summary CAHPS Hospice Survey score	Weighted	N/A	81.8 (21.1)	81.8 (21.5)	83.4 (20.7)	80.5 (22.5)	82.1 (20.6)
Any live discharges from hospice	Weighted	N/A	0.11 (0.32)	0.19 (0.39)	0.12 (0.33)	0.11 (0.32)	0.11 (0.31)
Transfer	Weighted	N/A	0.02 (0.14)	0.02 (0.13)	0.02 (0.14)	0.02 (0.14)	0.02 (0.14)
Revocation	Weighted	N/A	0.05 (0.22)	0.08 (0.28)	0.05 (0.22)	0.05 (0.23)	0.02 (0.15)
Death within 30 days of a live discharge	Weighted	N/A	0.1 (0.3)	0.17 (0.38)	0.09 (0.29)	0.1 (0.31)	0.13 (0.34)
Transfer to another hospice within 7 days of a live discharge	Weighted	N/A	0 (0.03)	0 (0)	0 (0.02)	0 (0.03)	0 (0.02)
Hospice enrollment in the year of death	Unweighted	0.46 (0.5)	0.46 (0.5)	0.49 (0.5)	0.47 (0.5)	0.47 (0.5)	0.53 (0.5)
Length of final episode of hospice care (days)	Unweighted	70.33 (164.76)	70.33 (164.8)	65.16 (143.9)	68.79 (154.6)	63.8 (142.5)	70.9 (155.0)
Final LOS less than 3 days	Unweighted	0.13 (0.33)	0.13 (0.33)	0.13 (0.33)	0.13 (0.34)	0.14 (0.34)	0.11 (0.31)
Final LOS less than 7 days	Unweighted	0.33 (0.47)	0.33 (0.47)	0.33 (0.47)	0.33 (0.47)	0.34 (0.47)	0.3 (0.46)
Final LOS more than 180 days	Unweighted	0.1 (0.31)	0.1 (0.31)	0.1 (0.3)	0.1 (0.31)	0.1 (0.29)	0.11 (0.31)
Professional visits in at least 2 of last 3 days of life	Unweighted	0.68 (0.47)	0.68 (0.47)	0.65 (0.48)	0.67 (0.47)	0.64 (0.48)	0.69 (0.46)
Summary CAHPS Hospice Survey score	Unweighted	81.2 (21.1)	81.2 (21.1)	81.34 (21.3)	81.64 (21.6)	81.5 (21.3)	82.24 (19.8)

Outcome	Weighting	Participating, 2022	Comparison, 2022	Participating, 2021	Comparison, 2021	Participating, 2019	Comparison, 2019
Any live discharges from hospice	Unweighted	0.13 (0.34)	0.13 (0.34)	0.12 (0.33)	0.14 0.35)	0.13 (0.34)	0.15 (0.36)
Transfer	Unweighted	0.02 (0.14)	0.02 (0.14)	0.02 (0.14)	0.02 0.15)	0.02 (0.14)	0.02 (0.14)
Revocation	Unweighted	0.06 (0.24)	0.06 (0.24)	0.06 (0.23)	0.07 0.25)	0.06 (0.24)	0.07 (0.26)
Death within 30 days of a live discharge	Unweighted	0.09 (0.28)	0.09 (0.28)	0.09 (0.29)	0.09 0.29)	0.1 (0.29)	0.12 (0.33)
Transfer to another hospice within 7 days of a live discharge	Unweighted	0 (0)	0 (0)	0 (0.03)	0 (0)	0 (0.02)	0 (0.04)

SOURCE: RAND analysis of CMS and other data.
NOTE: All outcomes are at the beneficiary level. Comparison participants include beneficiaries from comparison plans and beneficiaries from VBID-participating POs that are not participating in the Hospice Benefit component.

Table J.6 reports our DD estimates of the association between changes in our outcomes of interest and the Hospice Benefit component. These results are discussed in Chapter 11. Because of the relatively low ESSs, many of the nonsignificant estimates have CIs that include values that may correspond to meaningful policy effects, so we are not able to rule out meaningful effects of Hospice Benefit component participation. For example, while the estimated association between VBID and the probability of entering hospice is a non–statistically significant 1.9 percentage point reduction, based on the 95% CI, we cannot rule out changes ranging from a 4.5 percentage point reduction to a 0.7 percentage point increase.

As described in Appendix C, the main Hospice Benefit component models reported in Table J.6 removed covariates that could not be balanced with ASMD < 0.2 from the balancing algorithm and, instead, controlled for them as regression covariates. As a sensitivity analysis, we retained all variables in the balancing algorithm regardless of ASMD. Table J.7, shown after Table J.6, reports the results of this sensitivity analysis. None of the marginally significant results from Table J.6 (for the CAHPS summary score in 2021, LOS in 2022, and live discharge in 2022) remain statistically significant even at the 90% confidence level in Table J.7. However, the negative association between the Hospice Benefit component and revocation remained statistically significant in 2022, and became statistically significant in 2021. In Table J.7, we also found a statistically significant, negative association between the Hospice Benefit component and hospice enrollment.

Table J.6. Outcomes of Entropy-Balanced DD Models for the Hospice Benefit Component

Outcome (frequency, % unless otherwise noted)	Year	Estimate	Bootstrap Standard Error	95% CI Lower Bound	95% CI Upper Bound	*p*-value
Hospice enrollment in the year of death	2021	–0.019	0.013	–0.045	0.007	0.16
Length of final episode of hospice care (days)	2021	–0.49	4.86	–10.0	9.0	0.92
Final LOS less than three days	2021	–0.009	0.007	–0.02	0.006	0.23
Final LOS less than seven days	2021	0.011	0.011	–0.01	0.03	0.34
Final LOS more than 180 days	2021	0.007	0.009	–0.01	0.03	0.44
Professional visits in at least 2 of last 3 days of life	2021	0.007	0.018	–0.03	0.04	0.68
Summary CAHPS Hospice Survey score	2021	1.9	1.10	–0.21	4.08	0.08
Any live discharges from hospice	2021	–0.001	0.008	–0.02	0.02	0.91
Transfer	2021	0.001	0.004	–0.007	0.009	0.78
Revocation	2021	–0.009	0.006	–0.02	0.003	0.14
Death within 30 days of a live discharge	2021	–0.001	0.02	–0.04	0.04	0.96
Transfer to another hospice within 7 days of a live discharge	2021	0.001	0.001	0	0.003	0.15
Hospice enrollment in the year of death	2022	–0.017	0.014	–0.044	0.010	0.21
Length of final episode of hospice care (days)	2022	0.19	2.508	–4.72	5.11	0.94
Final LOS less than three days	2022	0.005	0.006	–0.007	0.016	0.43
Final LOS less than seven days	2022	0.013	0.008	–0.002	0.029	0.09
Final LOS more than 180 days	2022	0.001	0.004	–0.008	0.01	0.78
Professional visits in at least 2 of last 3 days of life	2022	-0.028	0.01	–0.047	–0.009	0.004
Summary CAHPS Hospice Survey score	2022	–0.046	0.618	–1.257	1.17	0.94
Any live discharges from hospice	2022	–0.01	0.005	–0.02	0.001	0.08
Transfer	2022	0	0.002	–0.004	0.004	0.94
Revocation	2022	–0.012	0.004	–0.021	–0.004	0.003
Death within 30 days of a live discharge	2022	–0.032	0.022	–0.075	0.011	0.15
Transfer to another hospice within 7 days of a live discharge	2022	0.000	0.001	–0.001	0.001	0.87

SOURCE: RAND analysis of CMS and other data.
NOTE: In addition to balancing, models controlled for plan-level fixed effects.

Table J.7. Sensitivity Analysis of Outcomes of Entropy-Balanced DD Models for the Hospice Benefit Component

Outcome (frequency, % unless otherwise noted)	Year	Estimate	Bootstrap Standard Error	95% CI Lower Bound	95% CI Upper Bound	p-value
Hospice enrollment in the year of death	2021	–0.064	0.029	–0.121	–0.007	0.029
Length of final episode of hospice care (days)	2021	0.75	18.3	-35.1	36.6	0.97
Final LOS less than three days	2021	–0.017	0.011	–0.04	0.005	0.123
Final LOS less than seven days	2021	–0.011	0.019	–0.048	0.025	0.547
Final LOS more than 180 days	2021	0.015	0.021	–0.026	0.056	0.47
Professional visits in at least 2 of last 3 days of life	2021	–0.008	0.027	–0.060	0.044	0.76
Summary CAHPS Hospice Survey score	2021	1.45	1.44	–1.38	4.28	0.31
Any live discharges from hospice	2021	–0.014	0.008	–0.031	0.002	0.088
Transfer	2021	0.002	0.004	–0.006	0.009	0.67
Revocation	2021	–0.014	0.007	–0.028	–0.001	0.035
Death within 30 days of a live discharge	2021	0.015	0.021	–0.026	0.055	0.47
Transfer to another hospice within 7 days of a live discharge	2021	0.001	0.001	–0.001	0.002	0.47
Hospice enrollment in the year of death	2022	–0.078	0.031	–0.138	–0.018	0.011
Length of final episode of hospice care (days)	2022	18.1	21.6	–24.2	60.4	0.40
Final LOS less than three days	2022	–0.003	0.018	–0.039	0.033	0.87
Final LOS less than seven days	2022	–0.026	0.021	–0.067	0.015	0.25
Final LOS more than 180 days	2022	0.041	0.024	–0.007	0.089	0.093
Professional visits in at least 2 of last 3 days of life	2022	0.005	0.023	–0.04	0.05	0.82
Summary CAHPS Hospice Survey score	2022	–0.35	0.79	–1.90	1.21	0.66
Any live discharges from hospice	2022	–0.009	0.007	–0.023	0.005	0.20
Transfer	2022	0	0.002	–0.003	0.003	0.99
Revocation	2022	–0.014	0.005	–0.023	–0.005	0.002
Death within 30 days of a live discharge	2022	–0.011	0.015	–0.04	0.018	0.45
Transfer to another hospice within 7 days of a live discharge	2022	–0.008	0.007	–0.023	0.006	0.25

SOURCE: RAND analysis of CMS and other data.
NOTE: Rather than removing hard-to-balance covariates from the EB model as is the case with our primary results. These models also controlled for plan-level fixed effects.

References

Anhang Price, Rebecca, Brian Stucky, Layla Parast, Marc N. Elliott, Ann Haas, Melissa Bradley, and Joan M. Teno, "Development of Valid and Reliable Measures of Patient and Family Experiences of Hospice Care for Public Reporting," *Journal of Palliative Medicine*, Vol. 21, No. 7, 2018.

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, 2009.

Butler, Danielle C., Stephen Petterson, Robert L. Phillips, and Andrew W. Bazemore, "Measures of Social Deprivation That Predict Health Care Access and Need Within A Rational Area of Primary Care Service Delivery," *Heath Services Resources*, Vol. 48, No. 2, Part 1, April 2013.

CAHPS Hospice Survey Project Team, "CAHPS® Hospice Survey," webpage, 2024. As of September 11, 2024: https://www.hospicecahpssurvey.org/

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

Center for Medicare & Medicaid Innovation, *Value-Based Insurance Design Model Incorporation of the Medicare Hospice Benefit into Medicare Advantage: CY 2021 Request for Applications*, 2020.

Centers for Medicare & Medicaid Services, *CAHPS Hospice Survey CMA Q4 2023 Final*, spreadsheet, 2023a. As of September 11, 2024: https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fhospicecahpssurvey.org%2Fglobalassets%2Fhospice-cahps4%2Fpublic-reporting%2Fscoring-and-analysis%2Fcare-compare-current%2Fcahps-hospice-survey-cma-q4-2023-final.xlsx&wdOrigin=BROWSELINK

Centers for Medicare & Medicaid Services, "Case-Mix Adjustments for Publicly Reported CAHPS® Hospice Survey Results," 2023b. As of September 11, 2024: https://hospicecahpssurvey.org/globalassets/hospice-cahps4/public-reporting/scoring-and-analysis/care-compare-current/cma_public_document-for-website.pdf

Centers for Medicare & Medicaid Services, *Fact Sheet - 2022 Part C and D Star Ratings*, Centers for Medicare & Medicaid Services, 2021. As of September 11, 2024: https://www.cms.gov/files/document/2022-star-ratings-fact-sheet1082021.pdf

Centers for Medicare & Medicaid Services, "2024 Medicare Advantage and Part D Star Ratings," 2023a. As of September 11, 2024: https://www.cms.gov/newsroom/fact-sheets/2024-medicare-advantage-and-part-d-star-ratings

Centers for Medicare & Medicaid Services, *Medicare 2023 Part C & D Star Ratings Technical Notes*, 2023b. As of September 11, 2024: https://www.cms.gov/files/document/2023-star-ratings-technical-notes.pdf

Centers for Medicare & Medicaid Services, "CAHPS® Hospice Survey: Pricing and Analysis," webpage, 2024. As of September 11, 2024: https://hospicecahpssurvey.org/en/public-reporting/scoring-and-analysis/

CMS—*See* Centers for Medicare & Medicaid Services.

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

Coleman, Kathryn A., Vanessa S. Duran, and Jennifer Lazio, *Update to 2025 Quality Bonus Payment Determinations* Centers for Medicare & Medicaid Services, 2024. As of September 11, 2024: https://www.cms.gov/files/document/updateto2025qualitybonuspaymentdeterminations.pdf

de Chaisemartin, Clément, and Xavier D'Haultfœuille, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, Vol. 110, No. 9, 2020.

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, *Appendices: 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, Center for Medicare & Medicaid Innovation, 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 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, Center for Medicare & Medicaid Innovation, 2023b.

Forst, Deborah, Eric Adams, Ryan Nipp, Allison Martin, Areej El-Jawahri, Ayal Aizer, and Justin T. Jordan, "Hospice Utilization in Patients with Malignant Gliomas," *Journal of Neuro-Oncology*, Vol. 20, No. 4, March 27, 2018.

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

Guest, Greg, Kathleen M. MacQueen, and Emily E. Namey, "Comparing Thematic Data," in G. Guest, K. M. MacQueen, and E. E. Namey, eds., *Applied Thematic Analysis*, SAGE Publications, Inc., 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*, Centers for Medicare & Medicaid Services, Center for Medicare & Medicaid Innovation, 2022.

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

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

Rickerson, Elizabeth, Joan Harrold, Jennifer Kapo, Janet T. Carroll, and David Casarett, "Timing of Hospice Referral and Families' Perceptions of Services: Are Earlier Hospice Referrals Better?," *Journal of the American Geriatric Society*, Vol. 53, No. 5, May, 2005.

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

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, 2013.

Teno, Joan M., David Casarett, Carol Spence, and Stephen Connor, "It Is "Too Late" or Is It? Bereaved Family Member Perceptions of Hospice Referral when their Family Member was on Hospice for Seven Days or Less," *Journal of Pain and Symptom Management*, Vol. 43, No. 4, April 2012.

Teno, Joan M., Mike Plotzke, Thomas Christian, and Pedro Gozalo, "Examining Variation in Hospice Visits by Professional Staff in the Last 2 Days of Life," *JAMA Internal Medicine*, Vol. 176, No. 3, March 2016.

Wachterman, Melissa W., Edward R. Marcantonio, Roger B. Davis, and Ellen P. McCarthy, "Association of Hospice Agency Profit Status with Patient Diagnosis, Location of Care, and Length of Stay," *JAMA*, Vol. 305, No. 5, February 2, 2011.

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