

MORTALITY EFFECTS AND CHOICE ACROSS PRIVATE HEALTH INSURANCE PLANS*

JASON ABALUCK
MAURICIO CACERES BRAVO
PETER HULL
AMANDA STARC

Competition in health insurance markets may fail to improve health outcomes if consumers are not able to identify high-quality plans. We develop and apply a novel instrumental variables framework to quantify the variation in causal mortality effects across plans and measure how much consumers attend to this variation. We first document large differences in the observed mortality rates of Medicare Advantage plans in local markets. We then show that when plans with high mortality rates exit these markets, enrollees tend to switch to more typical plans and subsequently experience lower mortality. We derive and validate a novel “fallback condition” governing the subsequent choices of those affected by plan exits. When the fallback condition is satisfied, plan terminations can be used to estimate the relationship between observed plan mortality rates and causal mortality effects. Applying the framework, we find that mortality rates unbiasedly predict causal mortality effects. We then extend our framework to study other predictors of plan mortality effects and estimate consumer willingness to pay. Higher-spending plans tend to reduce enrollee mortality, but existing quality ratings are uncorrelated with plan mortality effects. Consumers place little weight on mortality effects when choosing plans. Good insurance plans dramatically reduce mortality, and redirecting consumers to such plans could improve beneficiary health. *JEL Codes: I11, C26, I13.*

*We thank Joe Altonji, Leemore Dafny, Michael Dickstein, Joe Doyle, David Dranove, Mark Duggan, Amy Finkelstein, Amit Gandhi, Craig Garthwaite, Marty Gaynor, Jon Gruber, Kate Ho, Larry Katz, Tim Layton, Victoria Marone, Tom McGuire, Sarah Miller, Matt Notowidigdo, Chris Ody, Molly Schnell, Fiona Scott Morton, Mark Shepard, Doug Staiger, Jeroen Swinkels, Bob Town, numerous seminar participants, and four anonymous referees for helpful comments. We thank Jerry Chang, Emily Crawford, Eilidh Geddes, Elise Parrish, Aaron Pollack, and Amy Tingle for outstanding research assistance. We gratefully acknowledge funding through a National Institute of Aging pilot grant (P01 AG005842-30). All remaining errors are assumed to be *iid* type-I extreme value.

© The Author(s) 2021. Published by Oxford University Press on behalf of the President and Fellows of Harvard College. All rights reserved. For Permissions, please email: journals.permissions@oup.com
The Quarterly Journal of Economics (2021), 1557–1610. doi:10.1093/qje/qjab017.
Advance Access publication on May 6, 2021.

I. INTRODUCTION

When product quality is difficult to observe, consumers and producers may make suboptimal choices and investments. This concern is heightened in healthcare markets, where the quality of providers or insurance plans can be especially hard to infer. If consumers cannot determine whether certain plans are more likely to improve their health, then competition is unlikely to incentivize insurers to invest in this dimension of quality. To better inform consumers, policy makers disseminate provider and plan quality measures. But there is little evidence for how well existing quality measures predict the causal effects of insurance plans on enrollee health, much less whether consumers attend to such differences in plan quality.

This article estimates the effects of different private health insurance plans on enrollee mortality, investigates why some plans are higher quality by this measure, and assesses whether consumer demand responds to plan mortality effects. Our setting is the Medicare Advantage (MA) market, in which beneficiaries choose from a broad array of private managed care plans that are subsidized by the government. The MA program is large and growing, covering more than one-third of Medicare beneficiaries (KFF 2019). Annual mortality in the elderly MA population is relatively high (4.7%).

Measuring plan mortality effects is fundamentally challenging. Differences in observed mortality rates may reflect nonrandom selection by consumers of different unobserved health, while quasi-experimental variation in plan choice is limited and likely underpowered to detect different mortality effects across individual plans. Quantifying the extent to which consumer demand responds to mortality effects is also difficult, since any set of effect estimates are noisy and potentially biased by nonrandom sorting. We develop tools to overcome these challenges by combining observational and quasi-experimental variation, following a small but growing literature on quality estimation in education and health (Chetty, Friedman, and Rockoff 2014; Angrist et al. 2017; Hull 2020). We add to this literature by showing that instrumental variables (IV) methods relating observational quality estimates to true causal effects require a previously overlooked condition governing individual choice. We build theoretical and empirical support for the condition in the MA setting and show how extensions of such IV regressions can

be combined with standard discrete-choice modeling to estimate consumer willingness to pay for plan quality.

We begin by documenting large differences in the one-year mortality rates of MA plans operating in the same county, after adjusting for observable differences in enrollee demographics and accounting for statistical noise. We refer to these adjusted mortality rates as “observational mortality,” which we calculate as a time-invariant plan characteristic measured over our sample period. If causal, our estimated variation in observational mortality would suggest that a one standard deviation higher-quality plan decreases beneficiary mortality by 1.1 percentage points—a 23% reduction in mortality from a baseline rate of 4.7%, comparable with the sizable variation in mortality effects across hospitals (Doyle et al. 2015; Doyle, Graves, and Gruber 2019; Hull 2020). Given conventional estimates of the value of a statistical life (VSL), such variation suggests consumers should value higher-quality MA plans at tens or even hundreds of thousands of dollars a year.

However, variation in our observational mortality measure may reflect unobserved sorting as well as causal plan health effects. We validate the measure by leveraging variation in MA choice sets arising from plan terminations. Intuitively, when plans with high or low observational mortality exit a market their enrollees tend to reenroll in plans that have more typical observational mortality. The enrollees of nonterminated plans, in contrast, tend to be inertial and so they tend to remain in high- or low-mortality plans. If the observational mortality variation reflects variation in true mortality effects, we would therefore expect the subsequent mortality of enrollees in high- (low-) mortality plans to decline (rise) when these plans exogenously exit the market, relative to the subsequent mortality of beneficiaries in similar plans that do not terminate. The magnitude of this relationship should further reveal the relationship between observational estimates and causal plan effects. All else equal, subsequent enrollee mortality should change one-for-one with observational predictions when plan-level selection bias is negligible or uncorrelated with observational mortality across plans.

We formalize this quasi-experimental approach to validating observational mortality with a novel IV framework. Our main parameter of interest is the mortality effect “forecast coefficient,” defined by the regression of unobserved plan mortality effects on observational mortality. Although not identifying mortality effects for individual plans, the forecast coefficient can be used

to evaluate many policies of interest. For example, it allows the prediction of average effects of policies (based on, e.g., information or incentives) that would redirect consumers to plans with different observational mortality levels. We show how a feasible beneficiary-level IV regression identifies the forecast coefficient under three assumptions. First, we assume that terminations affect the observational mortality of an enrollee's plan via subsequent plan enrollment. We verify that the first stage is strong in our setting. Second, we assume that any relationship between observational mortality and underlying beneficiary health is the same in terminated and nonterminated plans, conditional on observables. We build support for this assumption, which allows for direct termination effects, by showing that there are not economically meaningful differences in patient observables across terminated and nonterminated MA plans, and that past cohorts in these plans have similar mortality prior to termination. In some specifications, we isolate terminations arising from a nationwide change in reimbursement policy for a category of MA plans.

Our primary methodological contribution is to show that these two standard IV conditions are not generally enough to estimate the plan forecast coefficient. Instead, the IV exclusion restriction that identifies the forecast coefficient comprises a usual "balance condition" (which would be satisfied when terminations are as-good-as-randomly assigned) and a novel "fallback condition." In our setting, this second condition restricts the fallback (second choice) plans that enrollees choose after a plan termination. Fallback choices must be similar to those chosen initially in terms of the unforecastable component of plan mortality effects. We show how the fallback condition can be microfounded by a standard discrete-choice model, in which there is no persistent unobserved heterogeneity in choices, and how it can be relaxed under different model assumptions. We further show how the assumption can be investigated empirically by testing for observable differences in fallback plans following plan terminations.

Our IV framework shows that observational mortality is a strong predictor of true MA mortality effects. Across a variety of specifications, we find first-stage effects of terminations on enrolled plan observational mortality that closely match the associated reduced-form effects of terminations on enrollee mortality. Consequently, IV forecast coefficient estimates are close to and statistically indistinguishable from one. This finding does not rule out selection bias in individual plan mortality rates. Instead,

the finding shows that variation in observational mortality across plans accurately predicts variation in causal mortality effects, at least on average.

We extend our approach to answer a series of policy-relevant questions. We first generalize the three IV assumptions to estimate the relationship between plan mortality effects and plan characteristics other than observational mortality. We find that the most widely used measure of plan quality, CMS star ratings, is uncorrelated with plan mortality effects. Higher premium plans have better mortality effects, as do plans with more generous prescription drug coverage and higher medical loss ratios. Thus, in every way we measure, plans that spend more tend to reduce enrollee mortality. Overall, our estimates imply very large variation across plans. Future work should explore additional mechanisms—including networks of providers—that could help explain this variation.

We then extend the IV approach to measure the extent to which consumers value plan mortality effects. Plans with better mortality effects tend to have larger market shares conditional on premiums. We show how this finding can be used in our IV framework to estimate the implicit willingness to pay (WTP) for plan quality. Estimating WTP is challenging because we observe only noisy and biased measures of mortality effects. We show how this challenge can be overcome by using our IV framework to compute forecast coefficients that relate mortality effects to premium-adjusted mean utility for each plan. Under our three IV assumptions, these forecast coefficients can be used to compute an upper bound on consumer WTP for plan quality. We find a positive WTP, but it is several orders of magnitude smaller than standard VSL estimates. Thus, although we find consumers to be somewhat responsive to differences in plan quality, they underrespond relative to the large variation in mortality effects. In simple partial-equilibrium simulations, we find that redirecting consumers to higher-quality plans could produce large benefits.

Our analysis of MA plan quality adds to a growing literature estimating the impact of health insurance on health. [Miller, Johnson, and Wherry \(2021\)](#) and [Goldin, Lurie, and McCubbin \(2021\)](#), for example, show that gaining access to Medicaid leads to large mortality reductions. [Card, Dobkin, and Maestas \(2008\)](#) similarly document a discontinuous drop in mortality when beneficiaries age into Medicare. Less well studied is the question of whether different types of insurance plans in a market can

differentially affect health outcomes.¹ By connecting plan quality differences to consumer demand, we add to a long literature studying consumer attentiveness to plan heterogeneity (Abaluck and Gruber 2011, 2016; Ericson and Starc 2016; Handel 2013; Handel and Kolstad 2015). Our findings have general-equilibrium implications, to the extent consumer demand affects the characteristics of offered plans (Starc and Town 2020; Miller et al. 2019).²

Our analysis adds to a recent methodological literature combining observational and quasi-experimental variation to estimate heterogeneity in the quality of institutions, such as hospitals, doctors, nurses, teachers, schools, and regions (Kane and Staiger 2008; Fletcher, Horwitz, and Bradley 2014; Yakusheva, Lindrooth, and Weiss 2014; Chetty, Friedman, and Rockoff 2014; Angrist et al. 2016, 2017; Finkelstein et al. 2017; Doyle, Graves, and Gruber 2019; Hull 2020). This literature draws on “value-added” estimation methods originally developed in the field of education; we are the first to apply such methods to measure the health effects of individual health insurance plans. We extend this literature in two ways. First, we formalize and develop tests for a novel assumption (i.e., the fallback condition) under which IV can be used to measure the relationship between observational value-added estimates and causal effects in the presence of selection bias. Second, we show how conventional discrete-choice modeling can be integrated with such IV procedures to both microfound the key fallback condition and measure how sensitive consumer choice is to true value added (e.g., the implicit consumer WTP).

Broadly, our approach builds on many earlier studies using exogenous displacements from institutions or regions to estimate their causal effects. Examples include studies of industry wage differentials (Murphy and Topel 1987;

1. McGuire, Newhouse, and Sinaiko (2011), for example, note the lack of systematic analyses comparing health outcomes in MA to health outcomes in traditional Medicare. One exception is Duggan, Gruber, and Vabson (2018), who find that MA plan terminations in counties with a single MA plan lead to increased hospital utilization but no change in mortality. Even fewer studies compare the quality of Medicare Advantage plans. Geruso, Layton, and Wallace (2020), for example, study random assignment of low-income beneficiaries to alternative Medicaid managed care plans, finding large spending effects but lacking sufficient power to detect mortality differences.

2. Similarly, Gaynor, Moreno-Serra, and Propper (2013) find that hospitals improve care quality when they face demand pressure, with corresponding reductions in patient mortality.

Krueger and Summers 1988; Gibbons and Katz 1992) or firm wage premiums (Abowd, Kramarz, and Margolis 1999; Card et al. 2018) using job transitions, studies of neighborhood or place effects using natural disasters (Chetty and Hendren 2018; Deryugina and Molitor 2020) or housing demolitions (Jacob 2004; Chyn 2018), and studies of school or hospital effects using unanticipated closures (Angrist et al. 2016; Carroll 2019). We develop a new framework for using such displacements to evaluate the relationship between causal effects and observational proxies while allowing for the kinds of endogeneity in fallback choices that has been a concern in some of the earlier studies.

We organize the remainder of this article as follows. In Section II, we describe the institutional setting and data, document large variation in observational mortality across MA plans, and motivate our quasi-experimental validation approach. In Section III, we develop our econometric framework for IV estimation of forecast coefficients and related parameters. In Section IV, we present our main forecast coefficient estimates. In Section V, we study the correlates of mortality effects and estimate consumer WTP. We conclude in Section VI. Additional results and other material is given in an Online Appendix.

II. SETTING AND DATA

II.A. Medicare Advantage

The Medicare program was established in 1965 primarily to provide insurance coverage for Americans aged 65 and older. Parts A and B of the Medicare program are often referred to as “traditional Medicare” (TM). TM is centrally administered by the Centers for Medicare and Medicaid Services (CMS) and covers hospitalizations and physician services for most Medicare beneficiaries.³ In recent years, a large and growing share of beneficiaries have instead opted to receive coverage through a set of diverse private managed care plans (34% as of 2019; see KFF 2019). This parallel private program has gone by various names (see McGuire, Newhouse, and Sinaiko 2011 for a comprehensive history), but is currently known as Medicare Advantage (MA).

3. Some beneficiaries, known as dual-eligibles, receive insurance coverage from both Medicare and Medicaid. We include these beneficiaries in our analysis while controlling for dual-eligible status.

Medicare beneficiaries can choose between TM and typically many MA plans in their local market. Broadly, MA plans must provide all of the mandated insurance benefits of TM in exchange for a capitated monthly payment. Competitive plans may charge lower premiums or offer supplemental benefits to attract certain consumers. MA plans also tend to vary significantly in their insurance networks, with some restricting access to providers (similar to commercial HMOs) while offering more generous financial coverage or better cost sharing. Although there is significant geographic heterogeneity in MA enrollment, most markets offer a wide variety of MA plans to choose from. In 2011, for example, 19 MA plans operated in the average county (KFF 2021).

The MA program has historically had two broad and sometimes conflicting goals: to expand consumer choice and to reduce Medicare costs (Prospective Payment Assessment Commission 1998; Medicare Payment Advisory Commission 2001).⁴ Less discussed is the role of competition among MA plans in enhancing product quality, although policy makers recognize the need for beneficiaries to make informed decisions in the MA market. Consequently, some form of public plan quality ratings has existed since 1999, with current quality rankings (known as star ratings) provided since 2008. These ratings score plans on multiple dimensions, including quality of care and customer service. Star ratings have also begun to play a role in policy making, with the 2009 Affordable Care Act giving bonus payments to high-ranked MA plans. Unlike with other programs, such as Value-Based Purchasing for hospitals, MA plans are not currently ranked or rewarded for achieving low enrollee mortality rates.

Multiple insurers may enter or exit a local market in any given year and change MA consumer choice sets. Broadly, insurers consider the cost of maintaining a given network, the potential revenue from different groups of beneficiaries, and policies affecting federal reimbursement when deciding what plans to offer. Duggan, Gruber, and Vabson (2018) argue that the factors that drive plan exit are unlikely to relate to outcomes through any

4. The MA program has always been controversial. "Cherry-picking" of healthy beneficiaries by MA plans could lead to overpayment by the federal government or skew benefit design to attract favorable risks (Brown et al. 2014). Despite potential efficiency gains, a substantial portion of the private (financial) gains from the MA program likely accrue to insurers (Cabral, Geruso, and Mahoney 2018; Duggan, Starc, and Vabson 2016).

other channel. For example, a policy change in 2008 increased the fixed costs of certain MA plans, known as private-fee-for-service (PFFS). [Pelech \(2018\)](#) documents significant plan terminations in the year following the policy, with the market share of PFFS plans falling by two-thirds between 2008 and 2011. We leverage this specific policy variation in some analyses below.

II.B. Data and Summary Statistics

We use data on the universe of Medicare beneficiaries aged 65 or older in the 50 U.S. states or the District of Columbia from 2006 to 2011. For each beneficiary in each year, we observe the identity of their selected plan (both MA and TM), their local market (county), standard beneficiary demographics (age, sex, race, and dual-eligible status), and their end-of-year mortality status. For traditional Medicare enrollees, we further observe inpatient claims. We supplement these data with characteristics of plans such as annual premiums, star ratings, and medical loss ratios.

Our Medicare data consists of 186,603,694 beneficiary-years with nonmissing enrollment, demographics, and mortality information. We use the full sample to construct our observational mortality measure, as discussed below. For our IV analysis, we restrict attention to the subset of beneficiaries in 2008–2011 who ended the previous year in a MA plan. Because of changes to Medicare reimbursement policy ([Pelech 2018](#)), the vast majority of plan terminations we observe take place during these years. The restrictions yield an analysis sample of 11,442,053 enrollees in 34,559 plans, where we treat plans in different counties as different products. [Online Appendix B](#) describes the construction of these samples in detail.

[Table I](#) summarizes our analysis samples. Column (1) shows average demographics, outcomes, and plan characteristics for the universe of Medicare beneficiaries in 2008–2011. The average Medicare beneficiary is 77.5 years old; 85.5% are white, 41.9% are male, and 15.9% are low-income and eligible for Medicaid in addition to Medicare (dual-eligibles). In any given year of our sample, 10.0% of Medicare beneficiaries change plans and 5.6% die. Among all Medicare beneficiaries, 12.5% are enrolled in a health maintenance organization (HMO), 2.0% are enrolled in a preferred provider organization (PPO), and 2.4% are enrolled in a PFFS plan. Within a county-year, we find about 25 plans in the median beneficiary choice set (including both TM and MA plans).

TABLE I
SUMMARY STATISTICS

	All	IV sample		
	Medicare plans (1)	All MA plans (2)	Non-terminated (3)	Terminated (4)
Beneficiary age	77.5	77.3	77.3	77.0
% White	85.5	87.4	87.3	90.5
% Male	41.9	41.1	41.0	43.3
% Dual-eligible	15.9	8.2	8.3	6.2
% Switched plans	10.0	14.1	11.6	100.0
% Died	5.6	4.7	4.7	4.5
% HMO	12.5	73.3	74.8	21.9
% PPO	2.0	9.9	9.9	9.1
% PFFS	2.4	10.8	9.1	67.1
Median <i>N</i> plans in choice set	25	17	17	12
Total plans	226,459	34,559	25,140	9,419
<i>N</i> beneficiary-years	118,184,127	11,442,053	11,119,125	322,928

Notes. This table summarizes the analysis samples in 2008–2011. Column (1) reports average enrollee demographics, annual plan switching rates, annual mortality, and plan type for the full Medicare population. Column (2) restricts the sample to beneficiary-years who ended the previous year in a MA plan. Columns (3) and (4) present the sample divided into beneficiary-years previously enrolled in MA plans that did and did not terminate. The total number of plans in column (3) subtracts the number of plans that ever terminate in column (4) from the number of MA plans in column (2). Choice sets are defined as county-years; plans operating in different counties are treated as different plans. We round the % switched plans in the final column to 100% from 99.99%.

Column (2) of Table I summarizes the subpopulation of beneficiary-years who ended the previous year in any MA plan (our IV sample). MA enrollees are less likely to be dual-eligible than Medicare beneficiaries as a whole but are otherwise demographically similar. A higher rate of MA beneficiaries switch plans in a given year (14.1%) and their annual mortality rate is somewhat lower than in the full sample (4.7%). The vast majority of MA enrollees are in HMOs (73.7%), PPOs (9.9%), and PFFS plans (10.8%).⁵

Columns (3) and (4) of Table I summarize the subpopulations of enrollees of MA plans that did and did not terminate in the previous year. These two groups appear broadly similar, though beneficiaries in terminated plans are slightly less likely to be

5. The remainder are in Medicare cost and demonstration plans or plans specifically designed for dual-eligibles.

dual-eligible and are located in somewhat smaller markets.⁶ The largest difference in these samples is the annual plan-switching rate: while all beneficiaries previously enrolled in a terminated plan are forced to change to a new MA plan, only 11.6% of beneficiaries in nonterminated plans switch.⁷ The majority of terminated plans (67.1%, when weighted by beneficiaries) are PFFS, reflecting the 2008 policy change.

II.C. Observational Mortality

We begin our analysis by computing observational differences in one-year mortality rates among Medicare plans operating in the same county, adjusting for observable differences in plan enrollees and statistical noise. These observational mortality estimates come from OLS regressions, of the form

$$(1) \quad Y_{it} = \sum_j \mu_j D_{ijt} + X'_{it} \omega + v_{it},$$

where Y_{it} is an indicator for beneficiary i dying in year t and D_{ijt} indicates her enrollment in a given plan j at the start of this period. The control vector X_{it} contains observable characteristics of enrollees (age, sex, race, and dual-eligibility status) as well as a full set of county and year fixed effects. We allow the coefficient vector ω to vary flexibly by plan size (see [Online Appendix C.1](#) for details). Given the fixed effects and controls, variation in the observational mortality coefficients μ_j thus reflects within-county differences in one-year plan mortality rates among observably similar enrollees. We estimate this model across all plans (both MA and TM), treating plans operating in different counties as different plans.

6. [Online Appendix](#) Figure A.I shows that the majority of counties have at least one termination during our sample period. [Online Appendix](#) Table A.I shows that counties with and without terminations have similar demographics, though counties without terminations are somewhat smaller and more sparsely populated than counties with terminations.

7. [Online Appendix](#) Table A.II describes switching behavior in more detail. In the full sample, 85.9% of enrollees do not switch plans in any given year. Among those consumers, 2.7% enroll in a different plan offered by the same insurer and 11.4% enroll in a plan offered by a different insurer. Consumers in terminated plans switch by definition; 18.6% enroll in a different plan offered by the same insurer and 81.3% enroll in a plan offered by a different insurer. Thus the vast majority of termination-induced switches are to new insurers within a market. Separately, 17% of termination-induced switches are to PFFS plans and 20% are to TM.

We account for statistical noise in the observational mortality estimates by applying a conventional empirical Bayes correction (Morris 1983). This correction, detailed in [Online Appendix C.1](#), “shrinks” the estimated μ_j toward their county- and plan size-level mean, in proportion to their expected degree of estimation error. The shrinkage is larger for smaller plans but minimal for the larger plans that make up the majority of our sample; as discussed in the [Online Appendix](#), our shrinkage procedure further allows for correlation of observational mortality rates within an insurer’s offerings. In practice the shrinkage procedure plays a minimal role for the typical plan, which enrolls several thousand beneficiary-years. The average effective shrinkage coefficient is very close to 1, with 90% of plans having a coefficient greater than 0.92.⁸

Estimates of [equation \(1\)](#) reveal substantial within-county variation in MA plan mortality rates among observably similar beneficiaries. The estimated beneficiary-weighted standard deviation of μ_j across MA plans, after correcting for estimation error, is 1.1 percentage points or 23% of the average one-year mortality rate of 4.7%. [Figure I](#) plots the full distribution of shrunk observational mortality rates across MA plans. The solid line shows this distribution for our baseline specification of [equation \(1\)](#), with all observable controls included in X_{it} , and the dashed line shows the corresponding distribution for a simpler specification that omits the beneficiary demographic controls. We normalize average observational mortality in both models by the average in the complete model that includes TM. The model without controls has a slightly lower mean (implying that MA plans have observably healthier beneficiaries than TM plans, on average) and a 45% larger standard deviation of 1.6 percentage points.

The fact that the mean and standard deviation of observational mortality changes when beneficiary demographic controls are included suggests some degree of nonrandom selection. In other words, the variation in observational mortality from the simpler specification appears to be in part driven by observable differences in the health of plan enrollees and not the true mortality effects of plans. This selection appears to be primarily on two dimensions of our observable characteristics: age and dual-eligibility. Conditional on these characteristics, further controlling for beneficiary sex and race has little effect

8. [Online Appendix](#) Figure A.II shows the distribution of effective shrinkage coefficients. See [Online Appendix C.1](#) for details.

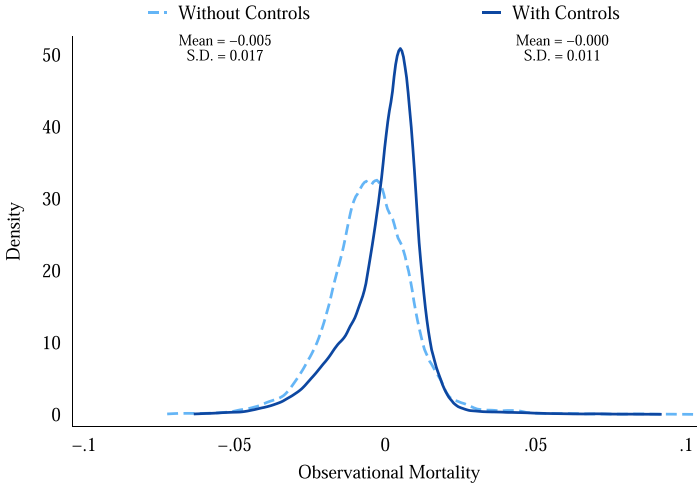


FIGURE I

Distribution of Plan Observational Mortality

This figure summarizes the enrollment-weighted distribution of observational mortality across MA plans. The solid dark line shows this distribution when observational mortality is estimated from [equation \(1\)](#), with all demographic controls, and the light dashed line shows the corresponding distribution for a simpler specification that omits age, race, sex, and dual-eligible status. Average observational mortality across all plans (TM and MA) is normalized to the average of the full model. Estimates are shrunk via the empirical Bayes procedure in [Online Appendix C.1](#). Estimated means and standard deviations of μ_j for MA plans are for the prior distribution, computed as described in [Online Appendix C.1](#), and shown for each estimation procedure.

on the estimated distribution of observational mortality (e.g., the noise-adjusted standard deviation of μ_j remains at 1.1 percentage points). Without further observables, we are unable to directly test for remaining selection bias in our benchmark specification.⁹ Instead, we derive an indirect validation based on termination-induced variation in MA choice sets.

II.D. Plan Terminations

To build intuition for our quasi-experimental approach to validating observational mortality, consider a set of beneficiaries who end a year enrolled in an MA plan with a high observational

9. Conventional value-added models in the education literature include, for example, measures of lagged test score outcomes to account for possible selection biases. For mortality, there is of course no analogous lagged outcome.

mortality rate μ_j . Since Medicare plan choice is highly inertial (only 14.1% of MA beneficiaries change plans in a given year, per Table I), most of these enrollees will remain in their high-mortality plan throughout the following year. Suppose, however, that at the end of the year the high-mortality plan terminates for a plausibly idiosyncratic reason (such as a federal change in reimbursement policy). This termination would force the plan's enrollees to make an active enrollment choice, and under standard regression to the mean, they will tend to switch to a new MA plan that is more typical in terms of μ_j .¹⁰ If the observational mortality rates were causal, then all else equal we would expect the mortality of this enrollee cohort to fall commensurate to the decline in μ_j . Identical logic holds for beneficiaries enrolled in exogenously terminated plans with low observational mortality rates: subsequent plan choice is likely to be more typical in terms of μ_j , relative to enrollees in nonterminated low observational mortality plans. If observational mortality variation reflects causal effects, then mortality should rise. Combining these two termination quasi-experiments may reveal the predictive content of our observational mortality rate estimates while allowing for direct termination effects that are common to the high- and low-mortality terminations.

Figure II illustrates the relationship between plan mortality rates and termination status for high- and low-mortality plans in our IV sample. The solid lines indicate regression-adjusted trends in observational mortality for beneficiaries before and after a plan termination, separately for beneficiaries previously enrolled in plans with above-median (orange/gray; color version online) and

10. Several examples are instructive. In 2008, a Florida-based HealthMarkets PFFS plan terminated, causing all beneficiaries to exit the contract the following year. Since HealthMarkets offered no other MA plans, all of the terminated beneficiaries switched to other insurers: one quarter switched to TM, and three-quarters switched to another MA plan. We also observe non-PFFS terminations at the plan level. For example, in 2009 a Blue Cross Blue Shield plan terminated in a number of Washington counties. Twenty percent of the enrollees then switched to another plan within the contract, 20% switched to TM, and 60% switched to other MA plans. Similar switching rates occurred for a 2010 termination of a CIGNA non-PFFS plan across several Virginia counties, with 80% of terminated enrollees switching to another MA plan and 20% switching to TM. Per Online Appendix Table A.II, these rates are broadly representative of the average switching rates in our analysis sample.

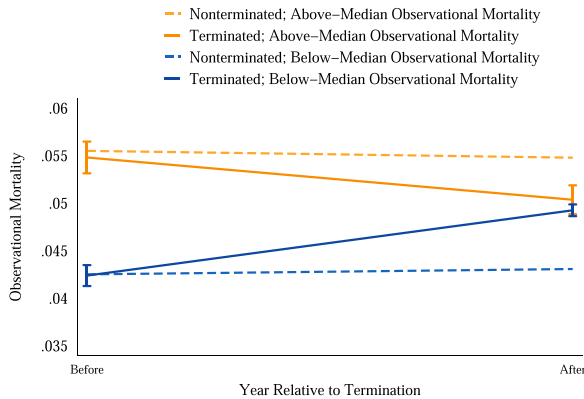


FIGURE II

Trends in Observational Mortality, Before and After Plan Terminations

This figure shows regression-adjusted trends in the observational mortality for enrollees in nonterminated and terminated MA plans, separately for plans with above- and below-median observational mortality. The median is defined over the entire IV sample. Data are plotted in the last year prior to termination for terminated plans and the following year. Termination effects are estimated in each year and median group by a separate regression that controls for county-by-year fixed effects; flexible interactions of lagged plan type and market shares; and beneficiary demographics (age in five-year bands, sex, race, and dual-eligibility status). County-clustered 95% confidence intervals for the termination effects are shown in brackets.

below-median (blue/black) mortality.¹¹ The dashed lines indicate comparable trends in observational mortality for beneficiaries in the same counties and years whose plans did not terminate, again separately for beneficiaries enrolled in above- and below-median mortality plans. The solid lines indicate a regression to the mean in plan choice following termination: those previously enrolled in high- and low-mortality plans tend to switch to more typical plans on average. At the same time, the dotted lines indicate inertia in plan choice absent termination: beneficiaries previously enrolled in high- and low-mortality plans tend to stay in these different plans provided they remain available. Bracketed 95% confidence intervals show that the posttermination difference in observational mortality is statistically significant

11. Specifically, we adjust the trends for county-by-year fixed effects, flexible interactions of lagged plan type and lagged market shares, and beneficiary demographics (age in five-year bands, sex, race, and dual-eligibility status).

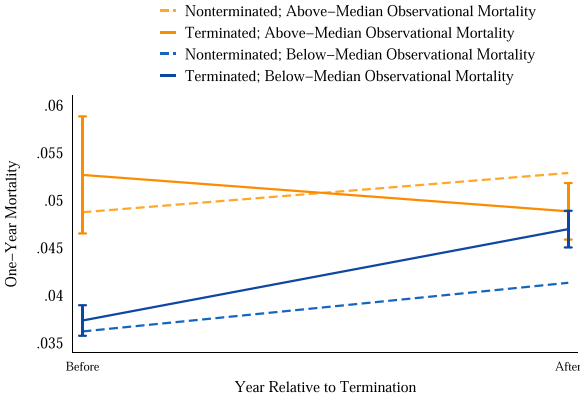


FIGURE III

Trends in One-Year Beneficiary Mortality, Before and After Plan Terminations

This figure shows regression-adjusted trends in the one-year mortality of enrollees of nonterminated and terminated MA plans, separately for plans with above- and below-median observational mortality. The median is defined over the entire IV sample. Data are plotted in the last year prior to termination for terminated plans and the following year. Termination effects are estimated in each year and median group by a separate regression that controls for county-by-year fixed effects; flexible interactions of lagged plan type and lagged market shares; and beneficiary demographics (age in five-year bands, sex, race, and dual-eligibility status). County-clustered 95% confidence intervals for these effects are shown in brackets.

for both high- and low-mortality plans, despite terminated and nonterminated plans having statistically indistinguishable observational mortality prior to termination.

Figure III illustrates the corresponding relationship between realized beneficiary mortality and plans' termination status for beneficiaries enrolled in high- and low-mortality plans. Here the solid and dashed lines correspond to the one-year mortality rates of the same groups of beneficiaries summarized in Figure II. Unlike the (time-invariant) observational mortality in Figure II, true mortality risk increases with age, such that the beneficiaries in nonterminated plans (dashed lines) exhibit an increasing trend in realized mortality. However, the solid blue (black) line (indicating the realized mortality of beneficiaries enrolled in a low-mortality plan prior to termination) exhibits a steeper trend, while the solid orange (gray) line (indicating the realized mortality of beneficiaries enrolled in a high-mortality rate plan prior to termination) exhibits a decreasing trend. Again the bracketed 95% confidence intervals show a significant termination effect

for both high- and low-mortality plans, with no statistically significant difference in average mortality prior to termination.

Together, the differential trends in [Figures II](#) and [III](#) suggest that a termination-induced move to MA plans with more typical observational mortality μ_j has a differential causal effect on actual mortality Y_{it} . This result suggests that the sizable variation in observational mortality we find in [Figure I](#) is not driven entirely by selection bias. At least some of the variation in observational mortality appears to be attributed to causal variation in MA plan mortality effects. Next we develop an econometric framework to formalize this logic and measure the predictive validity of observational mortality for such causal effects.

III. ECONOMETRIC FRAMEWORK

We use an IV framework, leveraging plan terminations, to measure the validity of observational mortality differences in predicting differences in causal plan mortality effects. While not identifying mortality effects for individual plans, this approach is sufficient to estimate the expected mortality impact of reallocating beneficiaries across observably different plans. We first outline the econometric setting and parameter of interest before providing three conditions under which this parameter is identified by an IV regression. We devote special attention to the third condition, what we call the fallback condition, which is novel to this article.

III.A. Plan Health Effects

We use a simple model to define causal plan effects and the IV parameter of interest. Let Y_{ijt} denote the potential mortality outcome of individual i in year t if she were to enroll in a plan j in her market. For the moment, we assume an additively separable model of $Y_{ijt} = \beta_j + u_{it}$; we extend our framework to account for unobserved treatment effect heterogeneity in [Section III.D](#). By normalizing the beneficiary-weighted average β_j in each market to zero, we can interpret each β_j as the average mortality effect from moving a random beneficiary to plan j , with u_{it} capturing latent differences in beneficiary health. Projecting u_{it} on a vector of observable characteristics X_{it} (which includes a constant) yields

$$(2) \quad Y_{ijt} = \beta_j + X'_{it}\gamma + \varepsilon_{it},$$

where $E[X_{it}\varepsilon_{it}] = 0$ by definition of the projection coefficient γ .

Consumers choose among the set of available plans in their market, with $D_{ijt} = 1$ indicating that consumer i enrolls in plan j in year t . Observed consumer mortality is then given by $Y_{it} = \sum_j Y_{ijt} D_{ijt}$. Substituting in the previous expression for Y_{ijt} yields

$$(3) \quad Y_{it} = \sum_j \beta_j D_{ijt} + X'_{it} \gamma + \varepsilon_{it}.$$

In contrast to the regression model (1), equation (3) is a causal model linking beneficiary plan choice D_{ijt} to subsequent mortality Y_{it} via the causal plan effects β_j .

Nonrandom plan selection creates fundamental econometric challenges in estimating plan mortality effects. To the extent that any given plan attracts consumers of poor (good) unobserved health, its observed mortality rate will be an upward-(downward-)biased estimate of β_j . For this reason, variation in the regression parameters μ_j that we estimate in equation (1) need not coincide with variation in the causal parameters β_j in equation (3): formally, average unobserved health ε_{it} need not be uncorrelated with the D_{ijt} choice indicators.

In principle, quasi-experimental variation in plan choice could be used to address such selection bias and estimate the full set of plan effects. This IV approach would require a set of exogenous variables Z_{ijt} to instrument for the plan choice indicators in equation (3). In practice, any available quasi-experimental variation in plan choice is unlikely to generate enough instruments for such a procedure (given the large number of MA plans in each market) nor have sufficient power to detect small differences in mortality effects (since mortality is relatively rare). We discuss our approach to quantifying variation in the plan mortality effects in light of these challenges.¹²

III.B. The Forecast Coefficient

Our first goal is to measure the relationship between observational mortality μ_j and true MA mortality effects β_j . Formally, we seek to estimate the MA forecast coefficient λ , defined by the projection of causal mortality effects β_j on observational

12. Estimating quality would also generally require structural assumptions (such as constant effects) that our approach does not impose. See Geweke, Gowrisankaran, and Town (2003) and Hull (2020) for applications of such models to estimate individual hospital quality.

mortality μ_j . Normalizing the means of both parameters to zero, this projection can be written as

$$(4) \quad \beta_j = \lambda\mu_j + \eta_j,$$

where η_j is mean-zero and uncorrelated with μ_j by definition. This regression is infeasible in the sense that the dependent variable β_j is neither observed nor estimated, despite measurement of the independent variable μ_j . The forecast coefficient nevertheless captures the predictive validity of the observational mortality measures. For example, μ_j is an average unbiased predictor of causal mortality effects when $\lambda = 1$, whereas observational mortality has little association with true causal effects when λ is small.¹³ We emphasize that [equation \(4\)](#) reflects an equilibrium statistical relationship, given by existing patterns of selection, and that λ is not a structural parameter.

Along with the forecast coefficient, [equation \(4\)](#) defines a forecast residual, η_j . This residual reflects the fact that for a given level of observational mortality μ_j , some plans may increase mortality by more or less than expected due to selection bias (even when $\lambda = 1$). Only when both $\lambda = 1$ and $\eta_j = 0$ for all j is observational mortality unbiased for individual MA plans (i.e., $\mu_j = \beta_j$).¹⁴ Since $Cov(\eta_j, \mu_j) = 0$, knowledge of the forecast coefficient is enough to place a lower bound on the variance in true causal effects, even in the presence of selection bias, by ignoring the contribution of η_j . Namely, $Var(\beta_j) \geq \lambda^2 Var(\mu_j)$.

Although it is not feasible to estimate [equation \(4\)](#) directly, we can relate it to observed enrollee mortality via the causal model [\(3\)](#). Substituting the former equation into the latter, we obtain

$$(5) \quad Y_{it} = \lambda\mu_{it} + X'_{it}\gamma + \varepsilon_{it} + \eta_{it},$$

where $\mu_{it} = \sum_j \mu_j D_{ijt}$ denotes the observational mortality of beneficiary i given her plan choice D_{ijt} and $\eta_{it} = \sum_j \eta_j D_{ijt}$ is the corresponding forecast residual of her selected plan.

13. This definition of the forecast coefficient aligns $1 - \lambda$ with the notion of “forecast bias” in the education value-added literature ([Kane and Staiger 2008](#); [Chetty, Friedman, and Rockoff 2014](#); [Angrist et al. 2017](#)).

14. [Chetty, Friedman, and Rockoff \(2014\)](#) refer to the analogue of $\mu_j \neq \beta_j$ as “teacher-level bias,” to contrast it with the weaker condition of $\lambda = 1$ (see also [Rothstein 2009](#)). [Angrist et al. \(2016, 2017\)](#) discuss IV-based tests of the joint null hypothesis that $\mu_j = \beta_j$ and $\lambda = 1$.

Equation (5) is again a causal model, linking observational mortality μ_{it} to realized mortality Y_{it} via the forecast coefficient λ . As with equation (3), OLS estimation of equation (5) will be biased when consumers of different unobserved health sort nonrandomly into plans.¹⁵ To estimate the forecast coefficient, we use an IV approach that follows the logic of Figures II and III. This approach uses an instrument for the observational mortality of an enrollee's plan that combines quasi-experimental choice set variation from plan terminations and the lagged observational mortality of an enrollee's plan. In contrast to the initial causal model, a single valid instrument is enough to identify λ in equation (5). However, there is a cost to simplifying equation (3), captured by the additional residual term η_{it} . We next discuss this cost in formalizing our IV approach.

III.C. Identification

1. Intuition and Related Literature. To see the basic logic of our IV approach, consider a market with three plans of equal market shares. Two of the plans, *A* and *B*, have an observational mortality of 0.05, and the third plan *C* has an observational mortality of 0.03. Suppose plan *C* exogenously terminates and subsequently all of its enrollees move to plan *A* or *B*. In either case, enrollees in plan *C* move to a plan where observational mortality is 2 percentage points higher. All else equal, the forecast regression (4) should then predict the resulting change in beneficiary mortality. If $\lambda = 1$, we expect mortality for the plan *C* cohort to rise by $5 - 3 = 2$ percentage points. If instead $\lambda = \frac{1}{2}$, we expect this cohort's mortality to rise by $\frac{1}{2}(5 - 3) = 1$ percentage point, as the 2 percentage point difference in observational mortality between plan *C* and either *A* or *B* would partly reflect selection bias and not causal effects. Such intuition mirrors the motivation for quasi-experimental evaluations of observational quality measures in other settings (Kane and Staiger 2008; Chetty, Friedman, and Rockoff 2014; Angrist et al. 2016; Doyle, Graves, and Gruber 2019).

15. In fact, when the control vector X_{it} is the same in these two models, OLS estimation of equation (5) (which uses the first-step estimates of μ_j from equation (3)) will mechanically give a λ estimate of 1, even when observational mortality is a badly biased predictor of true mortality effects. This result follows by standard projection algebra and highlights the importance of external quasi-experimental variation to identify λ .

A subtle but key ingredient to this intuition is “all else equal.” In the three-plan example, there is an implicit assumption that not only are terminations as good as randomly assigned to plan *C*, in the sense of being unrelated to unobserved beneficiary health ε_i , but that the plans chosen before and after its termination are representative in terms of η_j , the error term in [equation \(4\)](#). In fact, the presence of η_j may confound quasi-experimental inferences on λ , even when terminations are completely randomly assigned and thus independent of beneficiary health.

To see how the forecast residual can yield misleading quasi-experimental estimates of the forecast coefficient, suppose that while observational mortality is unbiased on average ($\lambda = 1$), there is still bias at the level of individual plans ($\eta_j \neq 0$). Concretely, suppose in the three-plan example that $\beta_A = \beta_C = 0.03$ and $\beta_B = 0.07$. In this case the exact mixture of fallback plans *A* and *B* determines how mortality responds to the termination. If all enrollees move to plan *B* following plan *C*'s termination, then mortality will rise by 4 percentage points. Given the observational mortality difference of 2 percentage points, a naive estimate of the forecast coefficient will be inflated by a factor of 2 (i.e., $\frac{\beta_B - \beta_C}{\mu_B - \mu_C} = 2\lambda$). Conversely, if all of *C*'s enrollees switch to plan *A*, one might falsely conclude that observational mortality has no relationship with true causal effects (i.e., $\frac{\beta_A - \beta_C}{\mu_A - \mu_C} = 0$). Only in the case where beneficiaries sort evenly into plans *A* and *B* following *C*'s termination, maintaining the equal market shares of the original plan choice distribution, will the comparison of actual mortality effects to observational mortality effects yield the correct estimate of $\lambda = 1$.

This potential challenge with quasi-experimental estimation of parameters like λ is quite general. For example, [Doyle et al. \(2015\)](#) uses ambulance referral patterns to measure returns to hospital spending, implicitly relating a hospital's average spending to its quality β_j . As they discuss, this approach requires more than random assignment of patients to ambulances. For their IV estimates to be unbiased, ambulance companies cannot systematically bring patients to higher-quality hospitals conditional on spending. Similarly, [Chetty, Friedman, and Rockoff \(2014\)](#) consider the case of teachers quasi-randomly moving across schools. To recover a forecast coefficient for grade-level teacher value-added parameters β_j , they require an additional assumption. New schools cannot be systematically good conditional on observational value-added. Within schools, grade assignments

further cannot track variation in value-added not captured by the observational measure.¹⁶

We formalize a novel solution to this general issue. The formal challenge in such settings is that the usual instrument exclusion restriction comprises two distinct assumptions: a familiar balance condition (satisfied when the instrument is as-good-as-randomly assigned) and a novel condition restricting the fallback choices of individuals subjected to a quasi-experimental shock (such as plan terminations, ambulance company assignment, or teacher moves). Intuitively, in our setting, the choices following a plan termination cannot systematically differ across terminated plans in ways that are correlated with the forecast residual η_j .

This solution can be compared and contrasted with other strategies using shocks to institutional choices in order to estimate causal effects. [Abowd, Kramarz, and Margolis \(1999\)](#) famously estimate worker and firm premiums using a two-way fixed effects model. Their approach identifies firm-specific β_j under a parallel trends restriction, akin to that of difference-in-differences, which assumes workers moving between different firms j would have seen similar wage changes absent a move ([Card et al. 2018](#); [Hull 2018](#)). The two-way fixed effects approach differs from our IV approach—as well as the examples above—in which an external shock to movement decisions is used to relate an observed characteristic of j to the β_j 's, without restricting outcome trends. [Gibbons and Katz \(1992\)](#) use an approach more similar to ours in the wage premium literature. They use plausibly exogenous plant closings to relate the observed differences in wages of industries j to industry premiums β_j . Here a worker's fallback industry following a plant closing need not be exogenous, and the authors test for this possibility. We provide a formal foundation for such an approach and propose new tests.

In our stylized example above, the fallback condition required that beneficiaries sort evenly into plans, which might suggest that this condition is generally quite strong. In fact, when pooling termination-induced choice set variation across many markets, the solution becomes weaker and more natural. We show below that the fallback condition holds in a wide range of discrete-choice models (including those typically estimated in

16. The education value-added literature typically considers quasi-experimental tests for selection bias, which can be thought to be conducted under the null hypothesis of $\eta_j = 0$ (see [Angrist et al. 2016, 2017](#)).

the industrial organization literature) and can be empirically investigated. Before presenting the general condition and its microfoundation, we first discuss the more standard first-stage and balance assumptions required by our IV approach.

2. The First-Stage and Balance Assumptions. Our approach to estimating the forecast coefficient uses an instrument which, as in [Figures II](#) and [III](#), leverages the interaction of past plan choice and plan terminations. Consider, for a beneficiary i observed in year t , the instrument

$$(6) \quad Z_{it} = \mu_{i,t-1} \times T_{i,t-1},$$

where $\mu_{i,t-1}$ denotes the observational mortality of the beneficiary's plan in the previous year, and $T_{i,t-1}$ is an indicator for whether that year was the plan's last (prior to termination). We derive conditions for this instrument to identify λ in a simplified setting where observational mortality is known without estimation error, there is no unobserved treatment effect heterogeneity, and we control only for characteristics of a beneficiary's plan in the previous year (including $\mu_{i,t-1}$ and $T_{i,t-1}$). We discuss how we relax each of these simplifying assumptions in [Section III.D](#).

An IV regression of beneficiary mortality Y_{it} on observational mortality μ_{it} , which instruments with Z_{it} and controls for X_{it} , identifies the forecast coefficient λ under three conditions, per [equation \(5\)](#). First, we require that the residualized instrument \tilde{Z}_{it} (that is, Z_{it} after partialing out X_{it} in the population) is correlated with observational mortality:

ASSUMPTION 1. (First Stage): $Cov(\tilde{Z}_{it}, \mu_{it}) \neq 0$.

The first-stage condition is highly intuitive in our setting. We expect most beneficiaries to remain in their previous year's plan because of inertia, unless the plan is terminated. Beneficiaries forced into an active choice by a termination, however, will tend to switch to more typical plans. This combination of inertia and regression to the mean implies that lagged terminations are likely to predict the observational mortality of year t choices differentially depending on lagged observational mortality, so that \tilde{Z}_{it} and μ_{it} are negatively correlated. Such negative correlation is shown in [Figure II](#), where terminated enrollees in below-median (above-median) observational mortality plans saw an increased

(decreased) observational mortality of their enrolled plan in the following year.

The second condition is a standard balance assumption: that Z_{it} is conditionally uncorrelated with unobserved beneficiary health ε_{it} .

ASSUMPTION 2. (Balance): $Cov(\tilde{Z}_{it}, \varepsilon_{it}) = 0$.

As-good-as-random assignment of plan terminations is sufficient but not necessary for this condition to hold. Because Z_{it} is given by the interaction of terminations and lagged observational mortality, and because both Z_{it} and X_{it} only vary at the lagged plan level, a minimal assumption is that any relationship between observational mortality and the average unobserved health of a plan's beneficiaries is the same for terminated and nonterminated plans. Formally, we can evaluate Assumption 2 in terms of the infeasible plan-level difference-in-differences regression,

$$(7) \quad \bar{\varepsilon}_{jt} = \phi_Z (\mu_j \times T_{j,t-1}) + X'_{j,t-1} \phi_X + e_{jt},$$

where $\bar{\varepsilon}_{jt} = E[\varepsilon_{it} | D_{ij,t-1} = 1]$ denotes the average unobserved health among beneficiaries previously enrolled in plan j and $X_{j,t-1}$ includes the lagged plan characteristics in X_{it} (including the μ_j and $T_{j,t-1}$ main effects). [Online Appendix C.2](#) shows that $Cov(\tilde{Z}_{it}, \varepsilon_{it}) = 0$ if and only if $\phi_Z = 0$ in the version of this regression that weights by lagged market shares. Since $T_{j,t-1}$ is included in $X_{j,t-1}$, this formulation of Assumption 2 makes clear that we allow both for terminated and nonterminated plans to enroll beneficiaries of systematically different unobserved health, and for plan terminations to have direct disruption effects. We only require that this imbalance or effect is not systematically related to the observational mortality measure.¹⁷ The similarity of the preperiod mortality in [Figure III](#) supports the stronger version of Assumption 2 in our setting; we develop and apply additional falsification tests of the sufficient balance assumption in [Section IV.A](#).

17. To see when this condition might fail, suppose that terminations among low observational mortality plans occur because population health appears to be systematically worsening but terminations among high observational mortality plans occur because of exogenous financial shocks. In this case, we might wrongly conclude that a relative decline in health among cohorts in terminated, low-mortality plans was due to those beneficiaries being reassigned to medium-mortality plans, and not because health was worsening among that population. The tests discussed below suggest such a story is unlikely in our setting.

3. *The Fallback Condition.* The third identification condition we formalize is novel, and follows the above intuition regarding fallback plans. Even when terminations are as good as randomly assigned (satisfying Assumption 2), consumers are not randomly assigned to fallback plans after terminations. Imbalance in the forecast residual η_j must thus be ruled out for Z_{it} to identify λ :

ASSUMPTION 3. (Fallback): $Cov(\tilde{Z}_{it}, \eta_{it}) = 0$.

Recall that $\eta_{it} = \sum_j D_{ijt} \eta_j$ is the forecast residual of the plan that consumer i selects in period t , potentially following a termination in time $t - 1$. For the instrument to be relevant, \tilde{Z}_{it} must be correlated with subsequent plan choice D_{ijt} ; thus, the as-good-as-random assignment with respect to η_j does not guarantee that \tilde{Z}_{it} is uncorrelated with η_{it} . Assumption 3 rules out this correlation, requiring fallback choices to be “typical” in a particular sense.

Interpreting Assumption 3 can be challenging because η_{it} is not structural. It instead arises from the statistical equation (4) and the potentially complex realizations of consumer choices and health that give rise to μ_j . We take two approaches to better understand the fallback condition. First, we give a plan-level interpretation analogous to equation (7). Second, we microfound the condition by asking what restrictions on consumer plan choices would cause it to hold.¹⁸

The fallback condition can be viewed (like Assumption 2) as restricting the relationship between observational mortality and a particular plan-level unobservable to be similar across terminated and nonterminated plans. Specifically, Assumption 3 restricts a plan-level difference-in-differences regression which replaces $\bar{\epsilon}_{jt}$ in equation (7) with $\bar{\eta}_{jt} = E[\eta_{it} \mid D_{ij,t-1} = 1]$. For the fallback condition to hold, the interaction of observational mortality μ_j and lagged plan termination $T_{j,t-1}$ must not predict $\bar{\eta}_{jt}$ conditional on the controls. This, in turn, says that the conditional relationship between μ_j and the average η_j of beneficiaries previously enrolled in terminated and nonterminated plans must be the same. This plan-level interpretation gives some intuition for the behavioral restrictions that might be sufficient. The

18. In Online Appendix C.3 we derive an alternative “monotonicity” condition that permits interpretation of the forecast IV regression coefficient as a weighted regression of β_j on μ_j with some convex (and estimable) weights. This condition is less plausible in our setting and identifies a less policy-relevant parameter than the unweighted λ .

fallback condition requires the first- and second-choice plans of consumers (i.e., the choices made before and after termination) to be similar, in terms of the relationship between the predictable dimension of plan quality μ_j and the unpredictable dimension η_j . Since the first-choice μ_j and η_j are uncorrelated by definition, the fallback condition requires that this lack of correlation remains as consumers switch from their first-choice plan to their second-choice plan. The fallback condition holds if consumers, after terminations, make similar choices from the remaining plans as new consumers in the market.

Microfounding the fallback condition requires behavioral restrictions on underlying consumer choice, since Assumption 3 is not ensured by as-good-as-random assignment of plan terminations. [Online Appendix C.4](#) presents a discrete-choice model that yields such restrictions. The simplest version of the model assumes that consumers in nonterminated plans are fully inertial, while consumers in terminated plans make an unrestricted choice that maximizes their latent utility U_{ijt} . We show that the fallback condition holds provided the IV control vector X_{it} includes any lagged characteristics of plans that lead to persistent unobserved heterogeneity in choice (along with $\mu_{i,t-1}$ and $T_{i,t-1}$). For example, suppose that consumer utility has the form

$$(8) \quad U_{ijt} = \alpha'_{it} W_j + \xi_j + u_{ijt},$$

where α_{it} captures potentially heterogeneous preferences over observed plan characteristics W_j , ξ_j denotes a fixed plan unobservable, and u_{ijt} captures unobserved idiosyncratic time-varying plan-specific preferences. We show in [Online Appendix C.4](#) that the fallback condition holds in this model (absent any functional-form assumptions) when α_{it} is either fixed across consumers or idiosyncratic over time. For general α_{it} , we show that the fallback condition holds provided flexible transformations of the lagged plan characteristics W_j are controlled for: namely, when one conditions on the characteristics of plans over which consumers exhibit heterogeneous and persistent preferences. Similar logic can be extended outside the utility model of [equation \(8\)](#): in [Online Appendix C.4](#) we discuss how any controls sufficient to capture persistent heterogeneity in plan choice probabilities can be included to satisfy Assumption 3 more generally. We further show that consumers in nonterminated plans need not be fully

inertial; the same logic can hold in models with partial inertia, such as that of [Ho, Hogan, and Scott Morton \(2017\)](#).¹⁹

The microfoundation suggests the novel fallback condition is likely to hold in discrete-choice specifications that are commonly estimated in both canonical and recent papers in the industrial organization literature. For example, [equation \(8\)](#) is the classic random-coefficient model of demand for differentiated products used in [Berry, Levinsohn, and Pakes \(1995\)](#). More recently, [Allende \(2019\)](#) employs a model in this class when estimating school value added. That said, there exist choice specifications that would violate the fallback condition. [Assumption 3](#) could fail if, for example, termination-induced changes in preferences cause consumers to select plans differently.²⁰

The microfoundation of the fallback condition has two implications for our IV approach. First, when estimating the MA forecast coefficient, it may be important to control for lagged plan characteristics over which consumers may have persistent heterogeneous preferences. We include such controls in our baseline specification, as discussed below. Second, as with the conventional balance assumption, the fallback condition may be investigated empirically. [Assumption 3](#) asserts that the forecast error of a beneficiary's plan, η_{it} , is conditionally uncorrelated with the instrument \tilde{Z}_{it} . We do not observe this residual directly, just as we do not observe the beneficiary residual ε_{it} that enters [Assumption 2](#). However, just as standard IV falsification tests can investigate whether the instrument is correlated with observable proxies of ε_{it} , we can construct and test for instrument balance on an observable proxy for η_{it} . Intuitively, we would check whether the observable characteristics of a beneficiary's fallback plans have a differential relationship with the observational mortality of her previous plan, across those previously enrolled in terminated and nonterminated plans. We conduct this test in the MA setting below.

19. We emphasize that our microfoundation does not assume consumers “select on μ_j ” but not on the unobservable η_j . Although intuitively sufficient for [Assumption 3](#), such selection would be generally difficult to microfound as both μ_j and η_j are statistical (not structural) objects.

20. Suppose, for example, that consumers in terminated high observational mortality plans learn to better identify plans with low β_j when forced to make an active choice. These consumers might choose plans with systematically smaller η_j following terminations; consequently, we may overstate the forecast coefficient by attributing a consumer's change in mortality to μ_j instead of η_j . The tests we discuss below suggest such a story is unlikely in our setting.

III.D. Extensions

We consider four extensions to our basic econometric framework before bringing it to the data. First, we note that although we have derived the first-stage, balance, and fallback conditions for an IV regression involving μ_j , in practice the observational mortality of each plan is not known and must be estimated. We show in [Online Appendix C.5](#) how each of these conditions extend to the case where μ_j is replaced with an empirical Bayes posterior mean of observational mortality μ_j^* . The untestable balance assumption is unchanged in this case, while the feasible IV regression fallback condition is satisfied under the same microfoundation we considered already. Importantly, we continue to estimate the same forecast coefficient λ with the feasible IV regression as we would if observational mortality were known, although increased estimation error in μ_j^* is likely to reduce power. In practice the issue of estimating μ_j should be of little empirical consequence in our setting, since the typical plan in our sample has thousands of enrollees and the typical shrinkage coefficient is correspondingly close to 1 (see [Online Appendix Figure A.II](#)).

Second, we note that we simplified the exposition by only considering an IV regression with lagged plan-level controls, of the form $X_{it} = \sum_j X_{j,t-1} D_{ij,t-1}$. This restriction also allows for controls at a level higher than plan, such as county-by-year fixed effects. In practice we further include controls that vary at the beneficiary level (such as demographics) in some IV specifications. When not necessary for identification, we expect such controls to absorb residual variation in beneficiary mortality and potentially yield precision gains.

Third, in [Online Appendix C.6](#) we show how our framework can accommodate unobservable selection on heterogeneous treatment effects. Our core argument proceeds similarly, although we require a further condition on unobserved selection on treatment effects. The new condition requires that any relationship between the degree of such Roy selection and observational mortality is again the same among consumers in terminated and non-terminated plans. Below we probe the role of treatment effect heterogeneity by allowing plan effects to vary by observables.

Finally, we note that although we have derived first-stage, balance, and fallback conditions for the purposes of estimating the forecast coefficient λ , analogous conditions can be imposed to estimate the coefficient from regressing plan effects β_j on any plan

observable W_j . The first stage for an instrument of the form $Z_{it} = W_{i,t-1} \times T_{i,t-1}$ (where $W_{i,t-1} = \sum_j W_j D_{ij,t-1}$) continues to derive power from a combination of plan choice inertia and termination-induced regression-to-the-mean; the balance assumption is analogous to Assumption 2, and the appropriate fallback condition continues to hold under our choice model microfoundation. We use this extension in Section V to study the observable correlates of plan quality, such as premiums and star ratings. We also show how our IV framework can be used to bound the implicit willingness to pay for plan quality using the association between plan mortality effects and premium-adjusted market shares.

IV. RESULTS

IV.A. Tests of Assumptions

We first investigate Assumption 1 by showing that termination-induced changes to consumers' choice sets lead to predictable changes in the observational mortality of the plan in which they subsequently enroll. We show this by estimating an OLS first-stage regression of

$$(9) \quad \mu_{it} = \pi_Z Z_{it} + X'_{it} \pi_X + v_{it},$$

where again μ_{it} denotes the plan observational mortality for beneficiary i at time t and $Z_{it} = \mu_{i,t-1} \times T_{i,t-1}$ is the interaction of observational mortality of the lagged plan and an indicator for lagged plan termination. To explore robustness, we sometimes replace the linear interaction with more flexible alternatives, such as interactions of percentiles of lagged observational mortality and lagged plan terminations. The baseline control vector X_{it} includes county-by-year fixed effects (such that we only exploit variation within choice sets), year- and county-specific termination main effects (to allow for flexible direct effects) and flexible interactions of lagged plan type, lagged observational mortality, and lagged plan size and market shares (to allow for a weakened fallback condition).²¹ In some specifications we include controls for beneficiary demographics (age in five-year bands, sex, race, and dual-eligibility status). We cluster standard errors at the

21. Plan type distinguishes traditional Medicare from several private alternatives: HMOs, local and regional PPOs, PFFS plans, and demonstration plans.

TABLE II
TESTS OF ASSUMPTIONS

	(1)	(2)
Panel A: First stage, dep. var.: observational mortality		
Instrument	-0.724 (0.015)	-0.0189 (0.0014)
F-Statistic	2,358.1	173.9
Panel B: Balance, dep. var.: predicted mortality		
Instrument	-0.020 (0.013)	-0.0011 (0.0006)
Panel C: Fallback, dep. var.: predicted forecast residual		
Instrument	0.010 (0.001)	-0.0000 (0.0001)
Specification	Linear	Median
Demographic controls	No	No
N beneficiary-years	11,441,205	

Notes. Panel A is based on estimation of equation (9) and presents the OLS coefficient in a first-stage regression of observational mortality on the instrument. Panel B replaces observational mortality as the dependent variable with a prediction of one-year mortality based on beneficiary demographics. Panel C uses as the dependent variable a prediction of the forecast residual based on plan characteristics. In column (1) the instrument is the interaction of observational mortality of the lagged plan and a lagged plan termination indicator. In column (2) the instrument is the interaction of an indicator for above-median observational mortality of the lagged plan and a lagged plan termination indicator. In all specifications, we control for the observational mortality of the lagged plan and termination main effects, county-by-year fixed effects, year- and county-specific termination effects, and interactions of lagged plan characteristics (as described in the text). Standard errors are clustered by county and reported in parentheses.

county level, allowing for arbitrary correlation in the regression residual across different beneficiaries, plans, and years.²²

First-stage coefficient estimates are reported in Table II, Panel A. The finding of $\pi_Z < 0$ is consistent with a combination of inertia and regression-to-the-mean in MA plan choice, first documented in Figure II. Beneficiaries enrolled in high- or low-mortality plans that are terminated in year $t - 1$ tend to choose plans in year t that are more typical in terms of observational mortality, relative to the mostly inertial beneficiaries in nonterminated plans; consequently, \tilde{Z}_{it} and μ_{it} are negatively correlated. In column (1), we estimate a termination-induced regression-to-the-mean of -0.72 , implying that a consumer in a 1 percentage point higher observational mortality plan in the previous period

22. The asymptotic standard errors do not account for finite-sample estimation error in the first-step estimates of μ_j . We have confirmed in bootstrap simulations that the contribution of such error is small, reflecting the fact that the typical plan enrolls several thousand beneficiaries with correspondingly precise μ_j estimates.

switches to a plan with 0.72 percentage point lower observational mortality in the period following termination, relative to a consumer in a similarly high-mortality plan that does not terminate. Column (2), corresponding more directly to [Figure II](#), shows that the termination of an above-median observational mortality plan in year $t - 1$ induces a differential reduction in the observational mortality of year t plans of 0.02 percentage points, relative to a termination of a below-median observational mortality plan. Both specifications yield high first-stage F -statistics, confirming the relevance of our instrument (Assumption 1).

[Figure IV](#), Panel A illustrates the first-stage relationship by replacing the linear instrument in [equation \(9\)](#) with one based on deciles of lagged observational mortality (controlling for decile main effects). We use this specification to plot the estimated contemporaneous plan observational mortality for enrollees who, in the previous year, were enrolled in plans of different deciles of observational mortality that did and did not terminate. The figure shows that although observational mortality of the lagged plan predicts current plan observational mortality among the nonterminated group, the relationship is essentially flat for terminated plans. The flattening again reflects the combination of inertia and regression-to-the-mean in plan choice that yields negative first-stage coefficients in [Table II](#), Panel A.²³

Next we build support for the balance condition (Assumption 2) by testing whether the instrument predicts observable differences in beneficiary health. We replace the observational mortality outcome in [equation \(9\)](#) with a prediction of one-year beneficiary mortality, obtained from a regression of one-year mortality on dummies for five-year age bands, sex, race, and dual-eligibility fixed effects (see [Online Appendix Table A.III](#) for model estimates). The results are in [Table II](#), Panel B. In contrast to the large and significant first-stage effects in Panel A, we cannot reject the null of instrument balance on predicted beneficiary mortality. With the baseline linear specification we obtain an insignificant coefficient of -0.020 , while in the median specification we obtain an insignificant coefficient of -0.0011 . Both estimates are more than an order of magnitude smaller than the corresponding first-

23. We normalize the height of each set of terminated and nonterminated points, at each decile of lagged observational mortality, by the regression model's prediction after removing average termination effects across all deciles. We emphasize that the overall trend in the lines is an intuitive normalization, with only the difference in the slopes used for IV identification.

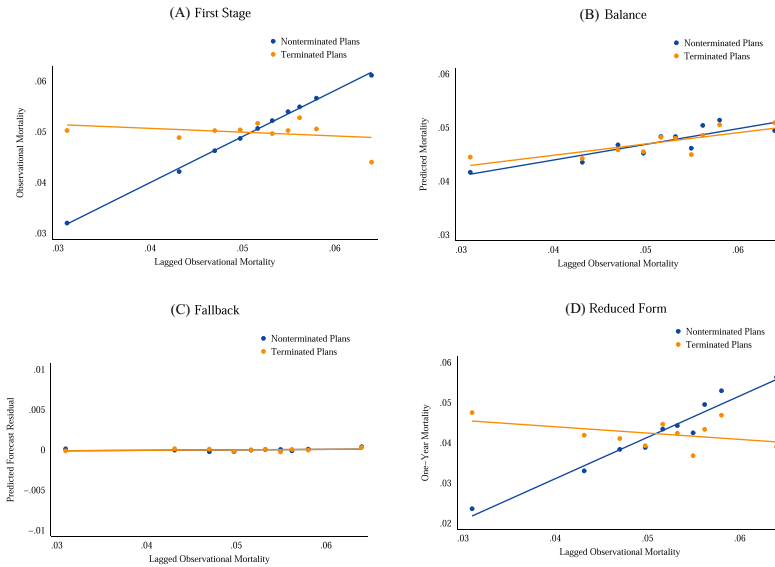


FIGURE IV

Graphical Tests of Assumptions and the Reduced Form

This figure illustrates the three assumptions in our IV approach, as well as the IV reduced form. Panel A shows average observational mortality by deciles of lagged observational mortality among nonterminated and terminated plans, controlling for county-by-year fixed effects and other observables in our baseline specification. Panel B shows the corresponding averages of predicted one-year mortality given omitted beneficiary demographics (age, sex, race, and dual-eligible status). Panel C shows the corresponding averages of a predicted forecast residual given omitted plan characteristics (star ratings, premiums, medical loss ratios, and an indicator for donut hole coverage). Panel D shows the corresponding averages of one-year mortality. Points are the average of each left-hand side variable in deciles of lagged plan observational mortality, predicted by the lagged observational mortality in the regression model, combined with the decile-specific termination effects estimated from specifications of the form of [equation \(9\)](#). The controls are as in [Table II](#), including decile main effects. Coefficients are normalized to remove termination main effects.

stage estimates. Finding balance for our instrument on predicted mortality is not surprising in light of the motivating [Figure III](#).

[Figure IV](#), Panel B illustrates the predicted mortality regressions by replacing the observational mortality measure in Panel A. We plot the average predicted mortality among terminated and nonterminated plans at different deciles of lagged observational mortality. In contrast to the clear first-stage effect, there is no differential trend in predicted mortality for terminated versus nonterminated plans. Any differential trend in the actual

mortality of beneficiaries in terminated and nonterminated plans is therefore unlikely to be due to preexisting differences in their health. [Online Appendix](#) Figure A.III similarly shows that our instrument appears visually balanced on age and average CMS risk scores, which attempt to predict enrollee costs based on demographics and diagnoses. Additional balance tests are given in [Online Appendix](#) Table A.IV.²⁴

Finally, we build support for the novel fallback condition (Assumption 3) by testing whether our instrument predicts an observable proxy for the forecast residual η_i . We construct the proxy by first regressing observational mortality on a set of observable plan characteristics (plan star ratings, premiums, medical loss ratios, and an indicator for donut hole coverage). We then take the residual from projecting the fitted values from this regression (as an observable proxy of β_j) on μ_j . This residual yields an observable proxy for η_j , and thus of $\eta_i = \sum_j \eta_j D_{ij}$ given a beneficiary's plan. [Table II](#), Panel C reports the resulting instrument coefficients from replacing the outcome in [equation \(9\)](#) with this proxy. In this case, we find a coefficient of 0.01 in the linear specification and a coefficient of effectively 0 in the median specification. Although statistically significant, the linear imbalance is quantitatively negligible—almost two orders of magnitude smaller than the associated first-stage effect. In [Online Appendix C.7](#), we show how the frameworks of [Altonji, Elder, and Taber \(2005\)](#) and [Oster \(2019\)](#) can be adopted to quantify the importance of imbalances on beneficiary and plan-level observables. The statistical imbalances are too small to substantially alter our forecast coefficient estimates, even under conservative assumptions.

[Figure IV](#), Panel C illustrates these predicted forecast residual regressions by replacing the predicted mortality measure in Panel B. As before, we see no systematic relationship between terminations and the predicted enrollee unobservable at any decile of lagged observational mortality. This result builds confidence in our third and final identification condition, suggesting

24. The appendix table shows that although it is visually small, the imbalance on plan risk scores and beneficiary age are statistically significant at conventional levels. At the same time, the other determinants of predicted mortality (sex, race, and dual-eligibility status) are not significantly imbalanced. The imbalances we find are furthermore quantitatively small: for example, in the linear specification in [Online Appendix](#) Table A.IV we find that a 1 percentage point higher observational mortality plan differentially enrolls individuals younger by 0.04 years in terminated versus nonterminated plans. We further explore the quantitative importance of these imbalances in [Online Appendix C.7](#), as discussed below.

TABLE III
FORECAST COEFFICIENT ESTIMATES

	(1)	(2)	(3)	(4)
Panel A: First stage, dep. var.: observational mortality				
Instrument	-0.724 (0.015)	-0.0189 (0.0014)	-0.724 (0.015)	-0.0189 (0.0014)
F-statistic	2,358.1	173.9	2,358.7	173.6
Panel B: Reduced form, dep. var.: one-year mortality				
Instrument	-0.764 (0.069)	-0.0214 (0.0025)	-0.745 (0.069)	-0.0203 (0.0023)
Panel C: Second stage (forecast coefficient), dep. var.: one-year mortality				
Observational mortality	1.056 (0.098)	1.130 (0.117)	1.029 (0.098)	1.073 (0.106)
Specification	Linear	Median	Linear	Median
Demographic controls	No	No	Yes	Yes
N beneficiary-years	11,441,205			

Notes. Panels A and C report first- and second-stage coefficient estimates from equations (9) and (10). Panel B reports the corresponding reduced-form coefficients. The dependent variable is observational mortality in Panel A and realized mortality in Panels B and C. In columns (1) and (3), the instrument is the interaction of observational mortality of the lagged plan and a lagged plan termination indicator. In columns (2) and (4), the instrument is the interaction of an indicator for above-median observational mortality of the lagged plan and a lagged plan termination indicator. In all specifications, we control for lagged observational mortality and termination main effects, county-by-year fixed effects, year- and county-specific termination effects, and interactions of lagged plan characteristics (as described in the text). Columns (3) and (4) additionally control for beneficiary demographics. Standard errors are clustered by county and reported in parentheses.

that termination-induced changes in observational mortality can be related to termination-induced changes in actual mortality to estimate the MA forecast coefficient. We present these IV estimates next.

IV.B. Forecast Coefficient Estimates

Table III reports first-stage, reduced-form, and second-stage estimates for our main IV specification. The second-stage estimates come from a regression of

$$(10) \quad Y_{it} = \lambda \mu_{it} + X'_{it} \gamma_X + \varepsilon_{it} + \eta_{it},$$

with the first stage given by equation (9). The second-stage coefficient λ estimates the observational mortality forecast coefficient under Assumptions 1–3. The reduced-form regression replaces the observational mortality outcome in equation (9) with the actual mortality outcome in equation (10). As before, we use this linear specification and an alternative specification that replaces the instrument with one constructed from an above-median

lagged observational mortality indicator. We also report two specifications for the control vector X_{it} ; one that mirrors the tests of our assumptions, and a second that adds beneficiary demographics (age, sex, race, and dual-eligible status). Given the balance of our instrument on these beneficiary observables, via the predicted mortality measure, we do not expect the inclusion of these controls to meaningfully affect the IV estimates (though it may increase their precision).

Table III, Panel A replicates the first-stage results reported in Table II, Panel A and confirms that these change little when we add the demographic controls. Panel B shows the corresponding reduced-form estimates from the same specifications. We find reduced-form coefficients of -0.76 and -0.75 for the linear specification (without and with demographic controls) and -0.0214 and -0.0203 for the median specification. These estimates are quite similar to the corresponding first-stage coefficients, reflecting the pattern first shown in Figures II and III: terminations tend to shift observational mortality and realized mortality by similar amounts.

Table III, Panel C shows that the similarity of first-stage and reduced-form effects yields high forecast coefficient estimates, in the range of 1.029 – 1.130 , with standard errors in the range of 0.098 – 0.117 . The point estimates are again similar with and without demographic controls, which reduce standard errors slightly. The median specification yields a somewhat higher forecast coefficient, though the estimates are not statistically distinguishable. Together, these IV estimates suggest the variation in observational mortality unbiasedly predicts variation in true mortality effects (i.e., that $\lambda \approx 1$).

Figure IV, Panel D illustrates this finding by plotting reduced-form variation in one-year mortality rates for beneficiaries in terminated and nonterminated plans by deciles of lagged observational mortality. The resulting differential trend (obtained by replacing observational mortality in equation (9) with actual one-year mortality) strongly mirrors that of the first stage in Panel A, consistent with the finding of a forecast coefficient that is close to 1. Lagged observational mortality strongly predicts the subsequent mortality of beneficiaries previously enrolled in nonterminated plans, but this relationship is effectively flat for beneficiaries previously enrolled in terminated plans (who switch to more typical plans). This finding is striking in contrast to Figure IV, Panel B which shows no such relationship for predicted one-year mortality. Beneficiaries in high- and low-mortality terminated plans appear

similar to those in corresponding nonterminated plans until they are induced by terminations to choose more average plans.²⁵

IV.C. Robustness Checks

We verify the robustness of our forecast coefficient estimates in a number of exercises summarized in [Online Appendix Table A.VI](#). First, we show that the estimates in [Table III](#) are unaffected by removing counties that do not see a plan termination during our sample period. The first row of [Online Appendix Table A.VI](#) shows we obtain similar forecast coefficient estimates of around 1.00–1.05 in this specification, with comparable standard errors. This finding is consistent with the fact that the vast majority of counties see MA plan terminations (see [Online Appendix Figure A.I](#)) and that counties with and without terminations are broadly similar (see [Online Appendix Table A.I](#)).

Second, we verify that similar results are obtained when we drop the minority of beneficiaries who switch from a MA plan to a TM plan (our baseline specification includes comparisons between the majority of MA plans and a single TM plan in each county). Although this specification may be biased by selecting on an endogenous variable, we nevertheless obtain similar forecast coefficients in the second row of [Online Appendix Table A.VI](#).

Third, we show that we obtain similar but less precise estimates when we limit attention to terminations of PFFS plans. [Pelech \(2018\)](#) links such terminations to a 2008 policy change that increased PFFS operating costs. Although these plan terminations are perhaps more plausibly exogenous, there may also be less variation across PFFS plans, which typically do not establish restrictive networks. The third row of [Online Appendix Table A.VI](#) shows that these plan terminations yield a similar

25. The similarity in forecast coefficient estimates for the linear and median specifications of [Table III](#) reflects a homogeneity in the first-stage and reduced-form relationship that is apparent across the support of lagged observational mortality. [Online Appendix Figure A.IV](#) shows this homogeneity more directly, by plotting the implied first-stage and reduced-form effects from [Figure IV](#), Panels A and D. [Online Appendix Figure A.IV](#), Panel A shows that these effects track each other closely across all 10 deciles. Panel B visualizes the implied forecast IV estimate by plotting the reduced-form estimates against the first-stage estimates and fitting a line through the origin. The slope of this line (which estimates λ) is similar to our linear and median coefficient estimates at 1.16, while its high R^2 (at 0.97) shows that the estimate is not driven by the exit of some select subset of plans with particular observational mortality.

forecast coefficient estimate of 1.08, with a standard error of 0.11. The corresponding median specification gives a slightly larger but similar estimate, with a similar standard error. The fourth row of [Online Appendix Table A.VI](#) reports the results of excluding PFFS terminations. Forecast coefficient estimates from this specification are more imprecise but qualitatively similar to (and not statistically distinguishable from) our baseline estimates.

We next investigate the role of treatment effect heterogeneity. The fifth row of [Online Appendix Table A.VI](#) shows that we obtain similar estimates, of around 1.05, when we exclude dual-eligible beneficiaries from both the IV sample and the sample used to construct the observational mortality measure. The sixth row further shows that our results are similar when we allow observational mortality to vary by beneficiary age, estimating [equation \(1\)](#) separately by five-year age bins. This specification yields forecast coefficients of around 1.03–1.07, with similar or slightly smaller standard errors. This robustness is especially striking as age and dual-eligible status appear to drive the majority of selection bias in the most naive observational mortality estimates, as discussed in [Section II.C](#). The findings suggest either that treatment effect heterogeneity is not first order in this setting, or that the extension of our framework in [Online Appendix C.6](#) (that accommodates such heterogeneity) is likely to hold.

We conclude this section by summarizing a number of additional robustness checks in [Online Appendix Table A.VI](#). We find similar results in the seventh row when including lagged plan risk scores as a control, consistent with the visual balance in [Online Appendix Figure A.III](#). We also find in the eighth row qualitatively similar (but slightly smaller) results when not shrinking the observational mortality estimates, consistent with the fact that the typical estimate is precise (per [Online Appendix Figure A.II](#)) and the discussion in [Online Appendix C.5](#). We further confirm in the ninth row that similar estimates of λ are obtained when dropping the smallest plans from our sample, which are likely to be the most imprecisely estimated. Finally, in the tenth row we report similar forecast coefficient estimates when using two-year mortality to both construct and validate observational mortality. This suggests that observed differences in plan mortality rates can be highly predictive of true plan causal effects at different horizons of beneficiary mortality.

IV.D. Interpretation

Taken together, our forecast coefficient estimates suggest that a large proportion of the sizable variation in observational mortality across MA plans reflects the causal impact of plan enrollment. It is worth emphasizing that this finding does not rule out selection bias in observational mortality, in the sense of $\mu_j \neq \beta_j$. Instead, our findings imply that variation in μ_j unbiasedly predicts variation in β_j , on average, despite any such selection bias. One might, for example, expect unobservably sicker beneficiaries to systematically prefer certain plans with more coverage. Our results and framework allow for this possibility: in the microfoundation of our fallback condition (discussed in [Online Appendix C.4](#)), we allow beneficiary preferences to correlate with their health in observed and unobserved ways, nesting common discrete-choice models of plan choice. A forecast coefficient near one can arise in such models even with systematic unobserved selection if the selection bias is negatively correlated with true causal effects (i.e., better plans attract unobservably sicker beneficiaries).²⁶ In this case (with $\eta_j \neq 0$), our forecast coefficient estimates give a lower bound on the variability of true causal effects: with $\lambda \approx 1$, the standard deviation of β_j is at least as large as the 1.1 percentage point standard deviation of μ_j found in [Section II.C](#).

Although an effect size this large may seem surprising, it is broadly consistent with a growing literature that shows large impacts of insurance status on health outcomes. Medicare as a whole has been found to have large mortality effects. [Card, Dobkin, and Maestas \(2008\)](#), for example, estimate a 20% mortality reduction in Medicare beneficiaries who are admitted to emergency departments.²⁷ This literature on place-based mortality effects estimates similarly large variation within Medicare

26. Formally, note that the forecast coefficient can be written $\lambda = \frac{\text{Cov}(\beta_j, \mu_j)}{\text{Var}(\mu_j)} = \frac{\text{Var}(\beta_j) + \text{Cov}(\beta_j, b_j)}{\text{Var}(\beta_j) + \text{Var}(b_j) + 2\text{Cov}(\beta_j, b_j)}$, where $b_j = \mu_j - \beta_j$ denotes selection bias for plan j . A forecast coefficient of $\lambda \approx 1$ can arise with nonzero bias when $\text{Cov}(\beta_j, b_j) \approx -\text{Var}(b_j)$, or when bias is sufficiently negatively correlated with the causal effect β_j . [Hull \(2020\)](#) finds such negative correlation between quality and selection in emergency hospital markets.

27. A growing literature also shows that insurance lowers mortality in the Medicaid program ([Goldin, Lurie, and McCubbin 2021](#); [Miller, Johnson, and Wherry 2021](#)). A 19% reduction in mortality in the MA program is thus within the range of the estimated extensive-margin effect of gaining health insurance more broadly ([Sommers, Gawande, and Baicker 2017](#)).

across all elderly beneficiaries, although these may capture both the joint impact of changing health systems and other demand-side factors.²⁸ Next we further argue that evidence on provider effects is consistent with the magnitudes we document.

V. CORRELATES OF PLAN EFFECTS

When combined with our observational mortality estimates, a forecast coefficient close to one implies large differences in causal mortality effects across plans. In this section, we investigate how these differences relate to observed plan attributes. We first ask whether plan characteristics predict observational mortality, μ_j . We then extend our basic IV framework to see whether these characteristics predict true mortality effects β_j . We consider different characteristics that may serve as proxies of plan quality, capture financial generosity and potential mechanisms, or measure consumer WTP for plan health effects.

V.A. Proxies of Plan Quality

We start by considering whether existing plan quality measures (star ratings) or prices (premiums) proxy for observational mortality and true plan effects. To help beneficiaries select plans, CMS produces star ratings on a 1–5 scale, with 5 stars indicating the highest quality. Star ratings depend on consumer satisfaction surveys and measures of clinical quality, but they explicitly do not condition on outcome data like mortality. In addition to making these ratings available to consumers, the government now pays “bonuses” to highly rated (4- and 5-star) plans.²⁹

Surprisingly, we find that CMS star ratings are positively correlated with our observational mortality measure, suggesting that higher-ranked plans have higher mortality rates.³⁰ The first column of [Table IV](#), Panel A, shows that a 1-star increase in a

28. [Finkelstein, Gentzkow, and Williams \(forthcoming\)](#) find that moving from a 10th percentile geographic region of health outcomes to a 90th percentile place reduces mortality by over 30%. [Deryugina and Molitor \(2020\)](#) also find evidence of large place effects.

29. See [Darden and McCarthy \(2015\)](#) for measures of demand responsiveness to star ratings and [Decarolis and Guglielmo \(2017\)](#) for an analysis of strategic incentives under the bonus program.

30. We study cross-sectional correlations with plan observables. Star ratings, for example, are averaged for each plan across all observed years (weighting by enrollment). We similarly average premiums and medical loss ratios.

TABLE IV
PLAN CHARACTERISTICS REGRESSIONS

	(1)	(2)	(3)	(4)	(5)
Panel A: OLS (observational mortality)					
Star rating	0.0008 (0.0003)				
Premium		0.0043 (0.0003)			0.0053 (0.0004)
Has donut hole coverage			-0.0022 (0.0003)		-0.0031 (0.0003)
Medical loss ratio				0.0019 (0.0023)	0.0002 (0.0024)
Panel B: IV (plan mortality effect)					
Star rating	-0.0011 (0.0012)				
Premium		-0.0058 (0.0019)			-0.0053 (0.0024)
Has donut hole coverage			-0.0046 (0.0016)		-0.0004 (0.0022)
Medical loss ratio				-0.0136 (0.0058)	-0.0127 (0.0058)
First-stage F -statistic	5,054.7	5,711.3	1,785.5	164.2	587.0
Maximum forecast R^2	0.0013	0.0267	0.0242	0.0048	0.0302
N beneficiary-years	11,441,205				

Notes. This table reports OLS and IV estimates of the regression of observational mortality and plan mortality effects, respectively, on plan characteristics. The dependent variable is observational mortality in Panel A and one-year mortality in Panel B. All specifications include the baseline controls in columns (3) and (4) of Table III. The IV specifications instrument by the interaction of lagged plan characteristics and terminations, controlling for main effects. Premiums are monthly and measured in hundreds of dollars. Missing plan characteristics are replaced by the average nonmissing value across plans. Standard errors are clustered by county and reported in parentheses. The maximum forecast R^2 is computed using the lower bound of $Var(\beta_j)$ implied by the observational mortality forecast coefficient in column (3) of Table III.

plan's ratings is associated with a 0.08 percentage point increase in observational mortality, controlling for county-by-year fixed effects and other baseline controls. This is a small but statistically significant positive correlation. Of course, this correlation could arise either because higher-ranked plans have worse mortality effects β_j or because sicker beneficiaries sort into higher star rating plans (causing selection bias $\mu_j - \beta_j$ to be positively correlated with star ratings).

To address selection bias, we recover the relationship between true mortality effects, β_j , and star ratings by an extension

of our IV approach. We estimate the analog of [equation \(10\)](#),

$$(11) \quad Y_{it} = \theta W_{it} + X'_{it}\rho + \varepsilon_{it} + \eta_{W,it},$$

which replaces the observational mortality treatment μ_{it} with a measure $W_{it} = \sum_j W_j D_{ijt}$ of a different enrolled plan characteristic W_j (here, star ratings), instruments with $Z_{it} = W_{i,t-1} \times T_{i,t-1}$, and replaces lagged observational mortality in X_{it} with the lagged plan characteristic $W_{i,t-1}$. For star ratings, the IV coefficient θ intuitively captures the extent to which termination-induced switches from low-rated plans to high-rated plans correlate with increased mortality Y_{it} . Formally, we can interpret θ as the plan-level regression analogous to [equation \(4\)](#) (which here projects plan effects β_j on star ratings W_j , instead of observational mortality μ_j) by natural extensions of our first stage, balance assumption, and fallback conditions to this setting.

IV estimates of θ show no relationship between star ratings and mortality effects. The first column of [Table IV](#), Panel B shows that a one-unit increase in star ratings is associated with a small and statistically insignificant 0.11 percentage point decrease in plan effects, with a standard error of 0.12 percentage points. This result suggests that the most commonly used measure of plan quality does not predict which plans systematically reduce beneficiary mortality on average.

We next investigate the correlation of observational mortality and plan effects with plan premiums. Premiums may also proxy for plan quality if quality investments are costly to insurers or if consumers demand higher-quality plans (we investigate the latter in more depth in [Section V.C](#)). In the second column of [Table IV](#), Panel A we find a positive and highly significant relationship between premiums and observational mortality, suggesting that a \$100 increase in monthly premiums is associated with a 0.43 percentage point increase in μ_j . Of course, as with star ratings, this correlation may be due to selection bias: plans may charge high premiums precisely because they enroll sicker-than-average beneficiaries.

IV estimates of the premium forecast coefficient are negative, suggesting that more expensive plans are of higher quality. The second column of [Table IV](#), Panel B suggests that a \$100 increase in monthly premiums (\$1,200 a year) is associated with a 0.58 percentage point decrease in β_j . In combination with the OLS estimate, this finding suggests that higher-premium plans are favored

by sicker consumers (consistent with the findings of [Starc 2015](#)). It also suggests that consumers may be leaving money on the table when it comes to the effective price of mortality reductions, a point we return to later. Even with conservative assumptions on the value of a statistical life, the dollar-equivalent mortality benefits of higher-premium plans appears to exceed the added cost.³¹

Although premiums (in contrast with star ratings) significantly predict plan mortality effects, they still explain a small share of quality variation. Because we can use the observational mortality variance and forecast coefficient to place a lower bound on the variance of β_j , we can use the star rating and premium forecast coefficients to place an upper bound on the R^2 from regressing plan effects on either of these plan characteristics.³² We find a maximum R^2 of 0.001 for star ratings and 0.027 for premiums, suggesting that only a small share of within-market quality variation can be explained by either observable.

We emphasize that these IV results are causal in a limited sense. They do not imply that, for example, a plan that raises premiums will improve its quality. This stronger claim (that we have recovered the causal impact of plan characteristics on β_j) only follows under stronger assumptions. Namely, it would require that there are no omitted plan characteristics that are correlated with premiums and also affect mortality—such that the regression of β_j on plan characteristics is itself causal. However, our results do suggest that higher-premium plans are of systematically higher quality and are more predictive of quality differences than CMS star ratings. To further explore potential mechanisms for plan quality differences, we turn to other plan characteristics.

V.B. Mechanisms

We investigate three mechanisms through which plans may affect beneficiary health: cost-sharing, direct control of beneficiary utilization, and provider networks.

31. At a conservative \$1 million VSL, a 0.58 percentage point reduction in mortality is worth \$5,800.

32. Formally, $\frac{\text{Var}(W_j^i \hat{\theta})}{\text{Var}(\beta_j)} \leq \frac{\text{Var}(W_j^i \hat{\theta})}{\text{Var}(\lambda_j \mu_j)}$ since $\text{Var}(\beta_j) \geq \text{Var}(\lambda_j \mu_j)$. To estimate the maximum R^2 in [Table IV](#) we compute beneficiary-weighted variances of $W_j^i \hat{\theta}$ and divide by beneficiary-weighted variances of $\hat{\lambda}_j \mu_j$ where $\hat{\lambda}$ comes from column (3) of [Table III](#).

We first study the potential role of cost-sharing, as proxied by whether a plan offers coverage in the Medicare Part D “donut hole”: a range of prescription drug expenditures at which some plans stop cost-sharing. In [Table IV](#), Panels A and B we find that plans offering donut hole coverage tend to both have lower observational mortality (0.2 percentage points) and significantly more negative plan effects (0.5 percentage points), on average. This contrast is consistent with earlier findings that sicker beneficiaries tend to select into plans with donut hole coverage (e.g., [Polyakova 2016](#)).³³ The finding of large plan effect differences among plans that offer donut hole coverage suggests that lower cost-sharing may be more broadly beneficial.³⁴

MA plans may also affect utilization through other means, such as prior authorization requirements or physician reimbursement ([Dillender 2018](#)). These supply-side controls could affect both utilization and quality. We study whether mortality effects correlate with overall expenditures, as measured by medical loss ratios (MLRs): the percentage of premiums that are paid out in claims.³⁵ In [Table IV](#), Panels A and B we find that plans with higher MLRs tend to have higher observational mortality, but significantly lower plan effects. A 10 percentage point increase in MLR is associated with a 0.14 percentage point reduction in the plan mortality effect, and we estimate a comparable coefficient if we condition on premiums. This finding suggests that expenditure levels predict plan quality, echoing a similar correlation found between hospital expenditure and mortality effects (e.g., [Doyle et al. 2015](#)), but sicker beneficiaries tend to be found in plans with higher loss ratios.

Finally, we relate our findings to estimates of provider heterogeneity. The existing literature documents large variation in hospital mortality effects ([Geweke, Gowrisankaran, and Town 2003](#); [Doyle et al. 2015](#); [Hull 2020](#)), with [Hull \(2020\)](#) and [Doyle et al. \(2015\)](#) finding evidence that such variation is reliably captured by observational models. Correspondingly, we find that a hospital ob-

33. Despite this, [Yang, Gilleskie, and Norton \(2009\)](#) argue that plans with prescription drug coverage increase survival.

34. At a \$1 million VSL, the social value of a 0.5 percentage point reduction in mortality from more generous drug coverage is \$5,000.

35. Due to data availability, we use 2011 MLRs data rather than averaging MLRs over years as with the other plan characteristics. MLRs also differ in being determined at the insurer level, see [Online Appendix B](#) for details.

servational mortality model estimated across all Medicare beneficiaries (with the same demographic controls) suggests a one standard deviation better hospital decreases one-year mortality by roughly 20%. Given the significant variation in provider networks across plans (e.g., [Chernew et al. 2004](#)) this variation suggests a plausible mechanism for the equally large variation that we find in plan-level mortality effects. However an IV analysis of this potential mechanism is infeasible, given limited data on MA networks.³⁶

Overall, this analysis of mechanisms paints a clear and consistent picture. More expensive and higher-spending plans tend to reduce beneficiary mortality while tending to attract sicker beneficiaries. Still, much of the variation in plan quality remains unexplained as shown by the relatively low maximum R^2 of 0.0302 in column (5) of [Table IV](#), which includes all financial measures.³⁷ The large residual variation leaves ample room for alternative but harder-to-measure channels, such as physician and hospital networks, to play an important role.

V.C. Demand for Plan Quality

We now estimate the extent to which higher-quality plans tend to attract a greater market share. This analysis follows a further extension of our IV framework, which allows us to estimate the implicit weight consumers place on plan mortality effects and estimate the implicit WTP for plan quality. Intuitively, we can estimate latent demand from a plan's market share after accounting for differences in prices. Our IV framework then allows us to relate demand to unobserved plan quality and recover the WTP from this relationship.

To formalize our approach, first consider how WTP might be computed if plan quality β_j were directly observed. A standard discrete-choice approach specifies consumers as selecting plans

36. Hospital network data is available from state inpatient databases, but consistent information on MA discharges is available only for three states (California, Maryland, and Massachusetts). While market shares and hospital observational mortality estimates can be combined to create a measure of hospital network quality, the fact that these data cover a relatively small number of markets makes it challenging to draw inferences.

37. We do not simultaneously include all five characteristics in [Table IV](#) because star ratings and premiums are highly correlated. This correlation makes the OLS regression in Panel A difficult to interpret and weakens the first stage in Panel B, below the point where the IV coefficients can be easily interpreted.

to maximize their latent utility U_{ij} , given by

$$(12) \quad U_{ij} = \alpha p_j + \xi_j + u_{ij},$$

where p_j denotes the observed premium of plan j , ξ_j collects all other relevant characteristics of plans (observed or unobserved by the econometrician), and u_{ij} is a set of unobserved taste shocks for consumer i . We follow the usual assumption that u_{ij} follows a type-I extreme value distribution but make no other parametric assumptions and allow premiums to be endogenous in the sense of being correlated with ξ_j . Projecting ξ_j on β_j across plans, we obtain a decomposition of $\xi_j = \tau\beta_j + \psi_j$ with ψ_j uncorrelated with β_j . We expect both α and τ to be negative, as both higher premiums and larger mortality effects (worse quality) will tend to decrease demand. The ratio $\frac{\tau}{100 \times \alpha}$ captures the WTP for plan quality: the decrease in premiums sufficient to offset a 1 percentage point increase in mortality effects β_j , on average across other characteristics ψ_j .

When p_j and β_j are both observed, standard discrete-choice methods (e.g., [Berry 1994](#)) may be used to estimate the WTP parameter, perhaps using instruments to account for the possible endogeneity of premiums with respect to β_j and ψ_j . In practice β_j is not known; we instead observe the unbiased prediction $\lambda\mu_j^*$, where λ is again the observational mortality forecast coefficient (approximately one, in this setting) and μ_j^* is posterior observational mortality. Naively using this proxy in discrete choice-estimation of WTP is likely to generate bias for at least two reasons. First, estimation error in μ_j^* (due to finite samples) is likely to bias estimates of τ and α , potentially in the direction of attenuating the WTP estimate. Second, even when $\lambda = 1$, there may be unobserved differences in quality (i.e., nonzero η_j) that may add further bias of ambiguous sign.³⁸

We employ an alternative WTP estimation procedure that combines the discrete-choice formulation with our IV framework for estimating plan forecast coefficients. [Equation \(12\)](#) implies that variation in log plan market shares recovers the normalized

38. Alternative revealed-preference approaches may be used to overcome some of these identification challenges and bound WTP under certain conditions. See [Pakes et al. \(2015\)](#) for a discussion.

systematic component of consumer utility, which we denote δ_j :

$$(13) \quad \ln(s_j) - \ln(s_0) = \delta_j \equiv \alpha p_j + \tau \beta_j + \psi_j,$$

where we have without loss normalized the plan characteristics as relative to an outside option with market share s_0 . Given an estimate or calibrated value of the premium coefficient α , we may back out from this expression $\xi_j = \delta_j - \alpha p_j$. We can then use our IV approach to implicitly regress β_j on this ξ_j , identifying a forecast coefficient of

$$(14) \quad \kappa \equiv \frac{\text{Cov}(\beta_j, \xi_j)}{\text{Var}(\xi_j)} = \tau \frac{\text{Var}(\beta_j)}{\text{Var}(\xi_j)},$$

using the fact that $\text{Cov}(\beta_j, \psi_j) = 0$ by construction. Given [equation \(13\)](#), $\text{Var}(\xi_j) = \text{Var}(\delta_j - \alpha p_j)$ is identified by market shares and the premium coefficient α . Our observational mortality forecast coefficient further identifies a lower bound on $\text{Var}(\beta_j) \geq \lambda^2 \text{Var}(\mu_j)$. The forecast coefficient κ thus identifies a lower bound on $\tau = \kappa \frac{\text{Var}(\xi_j)}{\text{Var}(\beta_j)} \geq \kappa \frac{\text{Var}(\xi_j)}{\lambda^2 \text{Var}(\mu_j)}$ (recalling that $\tau < 0$, and thus $\kappa < 0$, when consumers value plan quality). The estimated or calibrated value of $\alpha < 0$ then yields an upper bound on consumer WTP, $\frac{\tau}{100 \times \alpha}$.

We show this calculation in [Table V](#) for a range of possible premium elasticities given in the first column.³⁹ In column (2), we translate these elasticities to a value for α by dividing by the beneficiary-weighted average premium. In column (3), we report corresponding estimates of κ , obtained from an IV regression of one-year mortality on the implied mean utility δ_j of a beneficiary's plan with our usual specification of the instrument and controls. These estimates are again valid under natural analogs of our Assumptions 1–3, as in [Sections V.A](#) and [V.B](#). For each premium elasticity we obtain a negative coefficient estimate, suggesting that β_j is negatively correlated with δ_j or that higher-quality plans tend to have higher premium-adjusted market shares (consistent with a similar finding for hospitals in [Chandra et al. 2015](#)).⁴⁰

39. [Curto et al. \(2021\)](#) estimate an elasticity of -7 in this setting. Elasticities less than 1 in magnitude are implausible, because they are inconsistent with insurer profit maximization; nevertheless we include an elasticity of -0.5 .

40. [Online Appendix Table A.VII](#) shows that the star rating IV results are not sensitive to functional form. It further shows that PFFS plans have lower mortality effects. This is suggestive, as PFFS typically do not establish restrictive networks. In unreported regressions, we find that the largest insurers (Humana, United, and

TABLE V
WILLINGNESS TO PAY BOUNDS

Premium elasticity (1)	Premium coefficient (α) (2)	Forecast coefficient (κ) (3)	Minimum quality coefficient (τ) (4)	Maximum WTP: $\frac{\tau}{100 \times \alpha}$ (5)
-10	-0.0225	-0.0003 (0.0001)	-449.27 (141.77)	199.73 (63.03)
-7	-0.0157	-0.0004 (0.0001)	-322.65 (100.33)	204.91 (63.72)
-3.5	-0.0079	-0.0009 (0.0003)	-175.73 (52.69)	223.21 (66.93)
-1	-0.0022	-0.0028 (0.0009)	-67.12 (22.50)	298.37 (100.02)
-0.5	-0.0011	-0.0024 (0.0013)	-33.19 (17.19)	295.14 (152.86)

Notes. Column (5) reports estimates of the upper bound on quality WTP described in the text, for different values of the premium elasticity given in column (1). WTP is expressed in dollars per percentage point reduction in one-year mortality. The forecast estimates in column (3) are obtained by an IV regression of one-year mortality on the adjusted mean utility (ξ_j) of a beneficiary's plan, instrumented by the interaction of lagged adjusted mean utility interacted with lag terminations and controlling for lag adjusted mean utility and lag termination main effects along with the baseline controls in Table III (including demographics). Mean utility is adjusted by the premium utility coefficient (in column (2)) implied by the elasticity in column (1). The estimation sample is as in Table III. Column (4) translates the forecast coefficient estimate to an estimate of the quality coefficient bound described in the text. Standard errors are clustered by county and reported in parentheses.

Table V, column (4) uses these estimates to compute our upper bound on τ , and column (5) reports our corresponding estimates of the WTP for a 1 percentage point increase in plan quality.

For the wide range of possible premium elasticities, we estimate an upper bound on WTP of around \$200–\$295, implying that consumers are willing to pay no more than this amount to offset a 1 percentage point increase in one-year mortality.⁴¹ The largest of these estimates is around half of the average yearly premium in the sample (roughly \$600) and extremely small relative to conventional estimates of the value of a statistical life (around \$10 million for the average American and 20% of that, or

Blue plans) appear to supply higher-quality plans. These are interesting avenues for future research.

41. Naive WTP estimates based on $\lambda\mu_j^*$ tend to be lower in magnitude and negative. For example, a premium elasticity of -1 yields an implied WTP of -82.39 with a standard error of 15.57—a finding that would imply consumers are willing to pay for increases in mortality risk. This reflects the fact that observational mortality is increasing in plan size and decreasing in premiums, even as we find an opposite-signed relationship for true plan effects.

\$2 million by age 80; see [Kniesner and Viscusi 2019](#) and [Murphy and Topel 2006](#)). In [Online Appendix Table A.IX](#), we compute a range of VSL estimates for our marginal enrollees given assumptions about discounting, the value of a statistical-life year, mortality probabilities, and quality of life. These assumptions imply VSL estimates for marginal enrollees ranging from \$0.65 million to \$2.65 million. With these values, a 1 percentage point reduction in mortality would be worth between \$6,500 and \$26,500. Although our WTP bounds increase and become more imprecise as we use a lower premium elasticity, our largest estimate is an order of magnitude lower.

The finding that consumers are relatively insensitive to plan mortality effects is broadly consistent with a literature demonstrating that consumers overweight easily observable features, such as premiums, when choosing between health insurance plans ([Abaluck and Gruber 2011](#)). Many institutional features may explain the finding of low WTP for mortality effects in this setting. First, consumers may not have access to adequate information about quality. Although disclosure of plan quality has long been mandatory, CMS star ratings have only been publicly available since 2008, and we find them to be uncorrelated with the mortality effects above. Second, even when information is available, consumers may not be aware of it or may be unsure how to map it into outcomes they care about ([Dafny and Dranove 2008](#); [Darden and McCarthy 2015](#)).

Our forecast coefficient estimates in [Section IV.B](#) suggest that MA plan mortality effects are enormously variable in a market and can be predicted by observational mortality differences. At the same time, the WTP estimates suggest that consumers place little weight on this dimension of plan quality when making enrollment decisions. In [Online Appendix D](#), we conduct several policy simulations to quantify the partial-equilibrium health benefits of reassigning beneficiaries to alternative plans based on observed mortality effects. Policies that shift people between plans have potentially large benefits: people reassigned from plans at the 75th percentile of observed $\hat{\mu}_j$ to plans at the 25th percentile would see mortality fall by 0.6 percentage points per year. Given our VSL estimates, this implies that the potential benefits from reduced mortality risk in such scenarios could be substantial.⁴²

42. We note that the simulations are partial equilibrium and do not consider other important plan attributes, including financial characteristics. We discuss additional limitations in [Online Appendix D](#).

VI. CONCLUSIONS

We find large within-market differences in mortality rates across MA plans after adjusting for observable differences in enrollee characteristics and statistical noise. We show that this variation unbiasedly predicts true plan mortality effects with a novel instrumental variables framework. Publicly available quality measures are uncorrelated with true mortality effects. Perhaps as a result, consumer demand is underresponsive to this dimension of plan quality. Our results suggest broad scope for policy interventions based on these measures.

We make two main contributions to the broader literature on health insurance plan choice. First, we show that mortality effects are critical for assessing consumer choices. Papers that study only the financial consequences of plan choice miss an important dimension of plan quality. Second, our findings suggest large returns to understanding the market and plan-level determinants of plans' mortality effects. We find that plans with higher premiums, more generous drug coverage, and higher spending tend to reduce consumer mortality. Richer data are needed to investigate the role of plan networks and better understand the determinants of plan quality overall.

Methodologically, this article adds to a recent literature combining quasi-experimental and observational variation to estimate heterogeneous quality of institutions (such as schools and hospitals). We derive a novel condition for quasi-experimental variation in institutional choice to recover forecast coefficients in the presence of selection bias. We show how these forecast coefficients can be used to quantify the benefits of policies that assign individuals to different alternatives. We further show how our approach can be used to recover the sensitivity of consumer choices to unobserved causal effects and to estimate the WTP for these attributes. These methods may prove useful in many settings where consumers select institutions of differing quality and price.

From a policy perspective, our results suggest there may be large benefits from directing consumers to lower observational mortality plans. While the government does not currently release risk-adjusted mortality information, such information might be incredibly important. Our results also imply that insurers face weak incentives to invest in improving consumer health, which could be strengthened by new contractual or organizational forms

(e.g., integrating conventional health insurers with life insurance, as in [Kojien and Van Nieuwerburgh 2020](#)).

These conclusions come with important caveats. Publishing observational mortality rates might induce plans to invest in selecting healthier beneficiaries rather than improving health.⁴³ Furthermore, our partial-equilibrium simulations do not incorporate capacity constraints or allow premiums or quality to adjust with demand. Such effects could offset our implied gains, although the health effects are large enough that they may be first order.⁴⁴ The long-term consequences of better quality information are more difficult to gauge, but no less important. Making consumers more attentive to differences in plan health effects could accelerate the adoption of technologies that provide higher-quality care at lower cost.

YALE UNIVERSITY SCHOOL OF MANAGEMENT AND NATIONAL BUREAU OF ECONOMIC RESEARCH, UNITED STATES

BROWN UNIVERSITY, UNITED STATES

BROWN UNIVERSITY AND NATIONAL BUREAU OF ECONOMIC RESEARCH, UNITED STATES, E-MAIL: PETER.HULL@BROWN.EDU

KELLOGG SCHOOL OF MANAGEMENT, NORTHWESTERN UNIVERSITY, AND NATIONAL BUREAU OF ECONOMIC RESEARCH, UNITED STATES

SUPPLEMENTARY MATERIAL

An [Online Appendix](#) for this article can be found at *The Quarterly Journal of Economics* online.

DATA AVAILABILITY

Code replicating the tables and figures in this article can be found in [Abaluck et al. \(2021\)](#) in the Harvard Dataverse, <https://doi.org/10.7910/DVN/HTDG7I>.

43. Existing programs subsidize plans that score better on measures like star ratings, which we find to be uncorrelated with causal mortality effects. Although such programs might be improved by targeting risk-adjusted mortality, this could lead to insurer gaming. See [Decarolis and Guglielmo \(2017\)](#) for an analysis of the MA Quality Bonus Payment Demonstration program.

44. Nevertheless, the methods we develop here could help in quantifying these additional effects: for example, with quasi-experimental variation in the number of enrollees per plan, one could in principle investigate whether plans which experience enrollment shocks become less effective at promoting health.

REFERENCES

- Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc, "Replication Data for: Mortality Effects and Choice across Private Health Insurance Plans," (2021), Harvard Dataverse, <https://doi.org/10.7910/DVN/HTDG71>.
- Abaluck, Jason, and Jonathan Gruber, "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 101 (2011), 1180–1210.
- , "Evolving Choice Inconsistencies in Choice of Prescription Drug Insurance," *American Economic Review*, 106 (2016), 2145–2184.
- Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 67 (1999), 251–333.
- Allende, Claudia, "Competition under Social Interactions and the Design of Education Policies," Stanford University, Working Paper, 2019.
- Altonji, J., T. Elder, and C. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 113 (2005), 151–184.
- Angrist, Joshua, Peter Hull, Parag Pathak, and Christopher Walters, "Interpreting Tests of School VAM Validity," *American Economic Review: Papers & Proceedings*, 106 (2016), 388–392.
- , "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 132 (2017), 871–919.
- Berry, Steven T., "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 25 (1994), 242–262.
- Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (1995), 841–890.
- Brown, J., M. Duggan, I. Kuziemko, and W. Woolston, "How Does Risk-Selection Respond to Risk-Adjustment: Evidence from the Medicare Advantage Program," *American Economic Review*, 104 (2014), 3335–3364.
- Cabral, Marika, Michael Geruso, and Neale Mahoney, "Do Larger Health Insurance Subsidies Benefit Patients or Producers? Evidence from Medicare Advantage," *American Economic Review*, 108 (2018), 2048–2087.
- Card, D., C. Dobkin, and N. Maestas, "The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare," *American Economic Review*, 98 (2008), 2242–2258.
- Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline, "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36 (2018), S13–S70.
- Carroll, Caitlin, "Impeding Access or Promoting Efficiency? Effects of Rural Hospital Closure on the Cost and Quality of Care," University of Minnesota, Working Paper, 2019.
- Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson, "Healthcare Exceptionalism? Performance and Allocation in the U.S. Healthcare Sector," *American Economic Review*, 106 (2015), 2110–2144.
- Chernew, Michael E., Walter P. Wodchis, Dennis P. Scanlon, and Catherine G. McLaughlin, "Overlap in HMO Physician Networks," *Health Affairs*, 23 (2004), 91–101.
- Chetty, R., J. N. Friedman, and J. E. Rockoff, "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates," *American Economic Review*, 104 (2014), 2593–2632.
- Chetty, Raj, and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects," *Quarterly Journal of Economics*, 133 (2018), 1107–1162.
- Chyn, Eric, "Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children," *American Economic Review*, 108 (2018), 3028–3056.
- Curto, Vilsa, Liran Einav, Jonathan Levin, and Jay Bhattacharya, "Can Health Insurance Competition Work? Evidence from Medicare Advantage," *Journal of Political Economy*, 129 (2021), 570–606.

- Dafny, L., and D. Dranove, "Do Report Cards Tell Consumers Anything They Don't Already Know? The Case of Medicare HMOs," *RAND Journal of Economics*, 39 (2008), 790–821.
- Darden, M., and I. M. McCarthy, "The Star Treatment: Estimating the Impact of Star Ratings on Medicare Advantage Enrollments," *Journal of Human Resources*, 50 (2015), 980–1008.
- Decarolis, F., and A. Guglielmo, "Insurers' Response to Selection Risk: Evidence from Medicare Enrollment Reforms," *Journal of Health Economics*, 56 (2017), 383–396.
- Deryugina, Tatyana, and David Molitor, "Does When You Die Depend on Where You Live? Evidence from Hurricane Katrina," *American Economic Review*, 110 (2020), 3602–3633.
- Dillender, Marcus, "What Happens When the Insurer Can Say No? Assessing Prior Authorization as a Tool to Prevent High-Risk Prescriptions and to Lower Costs," *Journal of Public Economics*, 165 (2018), 170–200.
- Doyle, Joseph, John Graves, and Jonathan Gruber, "Evaluating Measures of Hospital Quality: Evidence from Ambulance Referral Patterns," *Review of Economics and Statistics*, 101 (2019), 841–852.
- Doyle, Joseph, John Graves, Jonathan Gruber, and Samuel Kleiner, "Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns," *Journal of Political Economy*, 123 (2015), 170–214.
- Duggan, Mark, Amanda Starc, and Boris Vabson, "Who Benefits When Government Pays More? Pass-through in the Medicare Advantage Program," *Journal of Public Economics*, 141 (2016), 50–67.
- Duggan, Mark, Jonathan Gruber, and Boris Vabson, "The Consequences of Health Care Privatization: Evidence from Medicare Advantage Exits," *American Economic Journal: Economic Policy*, 10 (2018), 153–186.
- Ericson, Keith, and Amanda Starc, "How Product Standardization Affects Choice: Evidence from the Massachusetts Health Insurance Exchange," *Journal of Health Economics*, 50 (2016), 71–85.
- Finkelstein, Amy, Matthew Gentzkow, Peter Hull, and Heidi Williams, "Adjusting Risk Adjustment—Accounting for Variation in Diagnostic Intensity," *New England Journal of Medicine*, 376 (2017), 608–610.
- Finkelstein, Amy, Matthew Gentzkow, and Heidi L. Williams, "Place-Based Drivers of Mortality: Evidence from Migration," *American Economic Review* (forthcoming).
- Fletcher, J., L. Horwitz, and E. Bradley, "Estimating the Value Added of Attending Physicians on Patient Outcomes," NBER Working Paper no. 20534, 2014.
- Gaynor, M., R. Moreno-Serra, and C. Propper, "Death by Market Power: Reform, Competition, and Patient Outcomes in the National Health Service," *American Economic Journal: Economic Policy*, 5 (2013), 134–166.
- Geruso, Michael, Timothy J. Layton, and Jacob Wallace, "Are All Managed Care Plans Created Equal? Evidence from Random Plan Assignment in Medicaid," NBER Working Paper no. 27762, 2020.
- Geweke, John, Gautam Gowrisankaran, and Robert J. Town, "Bayesian Inference for Hospital Quality in a Selection Model," *Econometrica*, 71 (2003), 1215–1238.
- Gibbons, Robert, and Lawrence Katz, "Does Unmeasured Ability Explain Inter-Industry Wage Differentials?," *Review of Economic Studies*, 59 (1992), 515–535.
- Goldin, Jacob, Ithai Z. Lurie, and Janet McCubbin, "Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach," *Quarterly Journal of Economics*, 136 (2021), 1–49.
- Handel, B. R., and J. T. Kolstad, "Getting the Most from Marketplaces: Smart Policies on Health Insurance Choice," Brookings Hamilton Project Discussion Paper 2015-08, 2015.
- Handel, Benjamin, "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 103 (2013), 2643–2682.

- Ho, Kate, Joseph Hogan, and Fiona Scott Morton, "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program," *RAND Journal of Economics*, 48 (2017), 877–905.
- Hull, Peter, "Estimating Treatment Effects in Mover Designs." Brown University, Working paper, 2018.
- , "Estimating Hospital Quality with Quasi-Experimental Data," Brown University, Working paper, 2020.
- Jacob, Brian A., "Public Housing, Housing Vouchers, and Student Achievement: Evidence from Public Housing Demolitions in Chicago," *American Economic Review*, 94 (2004), 233–258.
- Kane, Thomas J., and Douglas O. Staiger, "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation," NBER Working paper no. 14607, 2008.
- KFF, "A Dozen Facts about Medicare Advantage in 2019," KFF, 2019, <https://www.kff.org/medicare/issue-brief/a-dozen-facts-about-medicare-advantage-in-2019/>.
- , "Medicare Advantage 2021 Spotlight: First Look," KFF, 2021, <https://www.kff.org/medicare/issue-brief/medicare-advantage-2021-spotlight-first-look/>.
- Kniesner, Thomas J., and W. Kip Viscusi, "The Value of A Statistical Life," in *Oxford Research Encyclopedia of Economics and Finance* (2019), 19–15.
- Koijen, Ralph S. J., and Stijn Van Nieuwerburgh, "Combining Life and Health Insurance," *Quarterly Journal of Economics*, 135 (2020), 913–958.
- Krueger, Alan B., and Lawrence H. Summers, "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, 56 (1988), 259–293.
- McGuire, T., J. Newhouse, and A. Sinaiko, "An Economic History of Medicare Part C," *Milbank Quarterly*, 89 (2011), 289–332.
- Medicare Payment Advisory Commission, "Report to the Congress: Medicare in Rural America," Medicare Payment Advisory Commission Technical Report, 2001.
- Miller, Keaton S., Amil Petrin, Robert Town, and Michael Chernew, "Optimal Managed Competition Subsidies," NBER Working Paper no. 25616, 2019.
- Miller, Sarah, Norman Johnson, and Laura R. Wherry, "Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data," *Quarterly Journal of Economics*, 136 (2021), 1783–1829.
- Morris, Carl, "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78 (1983), 47–55.
- Murphy, Kevin M., and Robert H. Topel, "Unemployment, Risk, and Earnings: Testing for Equalizing Wage Differences in the Labor Market," in *Unemployment and the Structure of Labor Markets*, Kevin Lang and Jonathan S. Leonard, eds. (Oxford: Basil Blackwell, 1987), 103–140.
- , "The Value of Health and Longevity," *Journal of Political Economy*, 114 (2006), 871–904.
- Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business and Economic Statistics*, 37 (2019), 187–204.
- Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii, "Moment Inequalities and Their Application," *Econometrica*, 83 (2015), 315–334.
- Pelech, D., "Paying More for Less? Insurer Competition and Health Plan Generosity in the Medicare Advantage Program," *Journal of Health Economics*, 61 (2018), 77–92.
- Polyakova, M., "Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D," *American Economic Journal: Applied Economics*, 8 (2016), 165–195.
- Prospective Payment Assessment Commission, "Medicare Prospective Payment and the American Health Care System Report to the Congress," Prospective Payment Assessment Commission Technical Report, 1998.
- Rothstein, Jesse, "Student Sorting and Bias in Value-Added Estimation: Selection on Observables and Unobservables," *Education Finance and Policy*, 4 (2009), 537–571.

- Sommers, Benjamin D., Atul A. Gawande, and Katherine Baicker, "Health Insurance Coverage and Health—What the Recent Evidence Tells Us," *New England Journal of Medicine*, 377 (2017), 586–593.
- Starc, A., "Insurer Pricing and Consumer Welfare: Evidence from Medigap," *RAND Journal of Economics*, 45 (2015), 198–220.
- Starc, Amanda, and Robert J. Town, "Externalities and Benefit Design in Health Insurance," *Review of Economic Studies*, 87 (2020), 2827–2858.
- Yakusheva, O., R. Lindrooth, and M. Weiss, "Nurse Mortality-Added and Patient Outcomes in Acute Care," *Health Services Research*, 49 (2014), 1767–1786.
- Yang, Zhou, Donna B. Gilleskie, and Edward C. Norton, "Health Insurance, Medical Care, and Health Outcomes a Model of Elderly Health Dynamics," *Journal of Human Resources*, 44 (2009), 47–114.



CALL FOR NOMINATIONS

\$300,000 Nemmers Prize in Economics

Northwestern University invites nominations for the Erwin Plein Nemmers Prize in Economics, to be awarded during the 2026–27 academic year. The prize pays the recipient \$300,000. Recipients of the Nemmers Prize present lectures, participate in department seminars, and engage with Northwestern faculty and students in other scholarly activities.

Details about the prize and the nomination process can be found at nemmers.northwestern.edu. Candidacy for the Nemmers Prize is open to those with careers of outstanding achievement in their disciplines as demonstrated by major contributions to new knowledge or the development of significant new modes of analysis. Individuals of all nationalities and institutional affiliations are eligible except current or recent members of the Northwestern University faculty and past recipients of the Nemmers or Nobel Prize.

Nominations will be accepted until January 14, 2026.

The Nemmers prizes are made possible by a generous gift to Northwestern University by the late Erwin Esser Nemmers and the late Frederic Esser Nemmers.