

Mortality Effects and Choice Across Private Health Insurance Plans

Jason Abaluck, Mauricio Caceres Bravo,
Peter Hull, and Amanda Starc*

March 2021

Abstract

Competition in health insurance markets may fail to improve health outcomes if consumers are not able to identify high quality plans. We document large differences in the mortality rates of Medicare Advantage (MA) plans within local markets. We then show that when high (low) mortality plans exit these markets, enrollees tend to switch to more typical plans and subsequently experience lower (higher) mortality. We develop a framework that uses this variation to estimate the relationship between observed mortality rates and causal mortality effects; we find a tight link. We then extend the framework to study other predictors of 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.

*Abaluck: jason.abaluck@yale.edu, Yale University School of Management and NBER; Caceres Bravo: mauricio.caceres.bravo@gmail.com, Brown University; Hull: hull@uchicago.edu, The University of Chicago and NBER; Starc: amanda.starc@kellogg.northwestern.edu, The Kellogg School of Management, Northwestern University, and NBER. 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, 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, and numerous seminar participants for helpful comments. We thank Jerray 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.

1 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 healthcare 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 impacts of insurance plans on enrollee health, much less whether consumers attend to such differences in plan quality.

This paper estimates the effects of different private health insurance plans on enrollee mortality, investigate why some plans are higher quality by this measure, and assess 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 high, at 4.7%.

Measuring plan mortality effects is fundamentally challenging. Differences in observed mortality rates may reflect non-random selection by consumers of different unobserved health, while quasi-experimental variation in plan choice is both limited and likely under-powered 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 likely noisy and potentially biased by non-random 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 et al., 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 to the sizable variation in mortality effects

across hospitals (Doyle et al., 2015, 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 more per year.

However, variation in our observational mortality measure may reflect unobserved sorting as well as causal plan health effects. We next validate the measure by leveraging leverage 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 re-enroll in plans that have more typical observational mortality. The enrollees of non-terminated plans, in contrast, tend to be highly inertial and thus tend to remain in high- or low-mortality plans. If the observational mortality variation reflects variation in true mortality effects, we would 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 furthermore 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 selection bias is negligible.

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. While 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 impacts 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 impact the observational mortality of an enrollee’s plan via subsequent plan enrollment. We verify that the first stage is quite strong in our setting. Second, we assume that any relationship between observational mortality and underlying beneficiary health is the same in terminated and non-terminated plans, conditional on observables. We build support for this assumption, which allows for direct termination effects, by showing that enrollees in terminated MA plans are observably similar to those in non-terminated plans and that past cohorts in these plans have similar mortality rates prior to termination. In some specifications, we isolate terminations arising from a nationwide change in reimbursement policy for a category of Medicare Advantage 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 which 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 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 this third assumption can be microfounded theoretically by a standard discrete choice model. We further show how the assumption can be investigated empirically by testing for observable differences in fallback plans after exogenous terminations.

Our IV framework shows that observational mortality is a valid predictor of true MA mortality effects. Across a variety of specifications, we find first-stage effects of terminations on enrolled plan observational mortality which 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 on average.

We then 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. We further find suggestive evidence that plan networks could account for large differences in mortality effects by directing enrollees to high-quality providers, though we lack data to precisely measure the importance of this channel.

We next 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 one which is several orders of magnitude smaller than standard VSL estimates. While consumers have some ability to identify higher quality plans, we find that they underrespond to mortality effects. Consumers are only weakly sensitive to plan quality, and 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 et al. (2019) and Goldin et al. (2021), for example, show that gaining access to Medicaid leads to large mortality reductions. Card et al. (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, 2016a; Ericson and Starc, 2016; Handel, 2013; Handel and Kolstad, 2015). Our findings have general equilibrium implications, to the extent consumer demand impacts the characteristics of offered plans (Starc and Town, 2020; Miller et al., 2019).²

Our analysis also 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 (Hull, 2020; Fletcher et al., 2014; Yakusheva et al., 2014; Kane and Staiger, 2008; Chetty et al., 2014; Angrist et al., 2016, 2017; Doyle et al., 2019; Finkelstein et al., 2017). The 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 to 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 in order to estimate their causal effects. Examples include studies of industry wage differentials (e.g. Krueger and Summers, 1988; Murphy and Topel, 1987; Gibbons and Katz, 1992) or firm wage premiums (e.g. Abowd et al., 1999; Card et al., 2018) using job transitions, studies of neighborhood or place effects using natural disasters (e.g. Chetty and Hendren, 2018; Deryugina and Molitor, 2020) or housing demolitions (e.g. Jacob, 2004; Chyn, 2018), and studies of school or hospital effects using unanticipated closures (e.g. 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 the paper as follows. In Section 2, we describe the institutional

¹McGuire et al. (2011), for example, note the lack of systematic analysis comparing health outcomes in MA to health outcomes in traditional Medicare. One exception is Duggan et al. (2016) who find that MA plan terminations in counties with only 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 et al. (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.

²Similarly, Gaynor et al. (2013) find that hospitals improve care quality when they face demand pressure, with corresponding reductions in patient mortality.

setting and data, document large variation in observational mortality across MA plans, and motivate our quasi-experimental validation approach. In Section 3, we develop our formal framework for IV estimation of forecast coefficients and related parameters. In Section 4, we present our main forecast coefficient estimates. In Section 5, we study the correlates of mortality effects and estimate consumer WTP. We conclude in Section 6.

2 Setting and Data

2.1 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 typically 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 et al. (2011) for a comprehensive history), but is currently known as Medicare Advantage (MA).

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. While 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 reduce Medicare costs (Commission, 2001, 1998).⁴ Less discussed is the role of competition among MA plans in enhancing product quality, though policymakers 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 2007. These ratings score plans on multiple dimensions, including quality

³Some 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.

⁴The MA program has always been controversial. “Cherry-picking” of healthy beneficiaries by MA plans could lead to over-payment 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 et al., 2018; Duggan et al., 2016).

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 et al. (2016) argue that the factors that drive plan exit are unlikely to relate to outcomes through any 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.

2.2 Data and Summary Statistics

We use data on the universe of Medicare beneficiaries aged 65 or older in one of 50 US 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 non-missing 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. Appendix B describes the construction of these samples in detail.

Table 1 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

Table 1: 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 N Plans in Choice Set	25	17	17	12
Total Plans	226,459	34,559	25,140	9,419
N 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%.

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).

Columns 2-4 of Table 1 summarize 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 1 summarize the subpopulations of enrollees of MA plans that did and did not terminate in the previous year. Broadly, these two groups are observationally similar, though beneficiaries in terminated plans are slightly less likely to be dual-eligible and are in somewhat smaller markets.⁶ The largest difference in these samples is the annual plan-switching

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

⁶Appendix Figure A1 shows that the majority of counties have at least one termination during our sample period.

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 non-terminated plans switch.⁷ The majority of terminated plans (67.1%, when weighted by beneficiaries) are PFFS, reflecting the 2008 policy change.

2.3 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 accounting for statistical noise. These observational mortality estimates come from ordinary least squares (OLS) regressions, of the form

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

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 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.

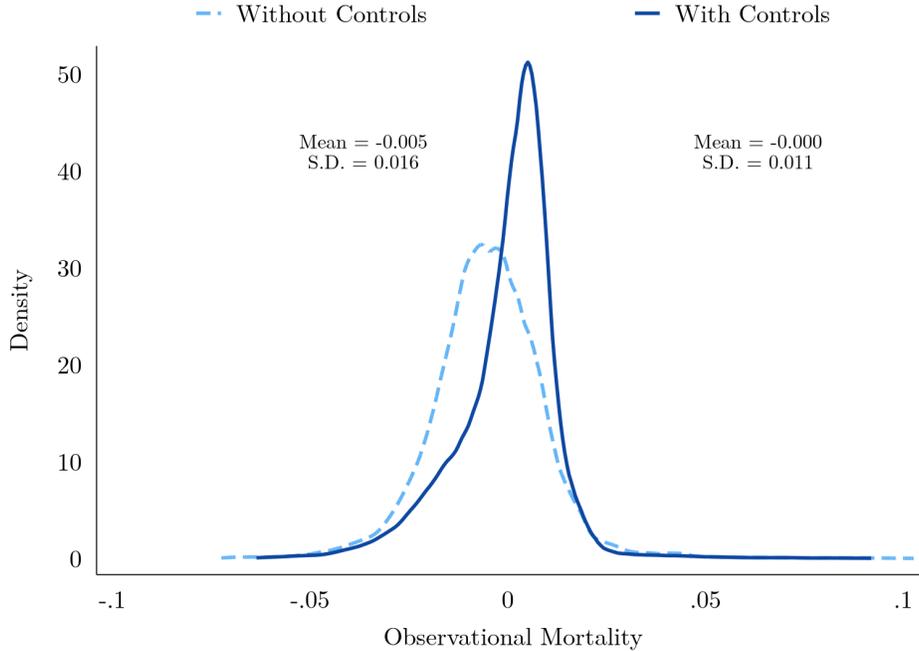
We account for statistical noise in the observational mortality estimates by applying a conventional empirical Bayes correction (Morris, 1983). This correction, detailed in Appendix C.1, “shrinks” the estimated μ_j towards their county- and plan size-level mean, in proportion to their expected degree 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 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 one, with 90% of plans having a coefficient greater than 0.92.⁸

Appendix Table A1 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.

⁷Appendix Table A2 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.

⁸Appendix Figure A2 shows the distribution of effective shrinkage coefficients. See Appendix C.1 for details.

Figure 1: Observational Mortality



Notes: 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, while 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 Appendix C.1. Estimated means and standard deviations of μ_j for MA plans are for the prior distribution, computed as described in Appendix C.1, and shown for each estimation procedure.

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 1 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} , while 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 non-random 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 on the estimated distribution of observational mortality (e.g. the noise-adjusted standard deviation of μ_j remains at 1.1 percentage points). Absent 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.

2.4 Plan Terminations

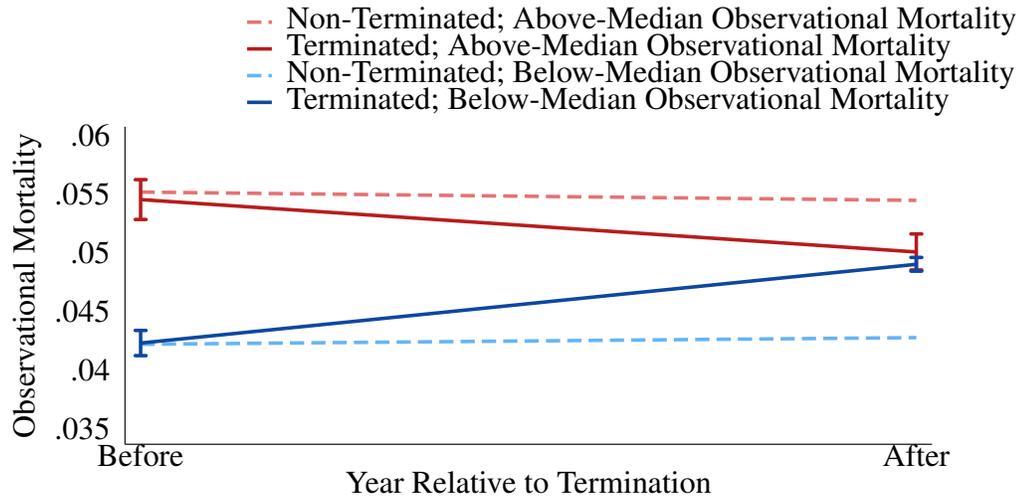
To build intuition for our quasi-experimental approach to validating observational mortality, consider a set of beneficiaries who start the year enrolled in a MA plan with a high observational mortality rate μ_j . Since Medicare plan choice is highly inertial (only 14.1% of MA beneficiaries change plans in a given year, per Table 1), most of these enrollees will remain in their high-mortality plan throughout the year. Suppose, however, that at the start 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 non-terminated 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 on mortality that are common to the high- and low-mortality terminations.

Figure 2 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

⁹Conventional 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.

¹⁰Several 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, with three quarters switching 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. 20% 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 Appendix Table A2, these rates are broadly representative of the average switching rates in our analysis sample.

Figure 2: Plan Terminations and Observational Mortality



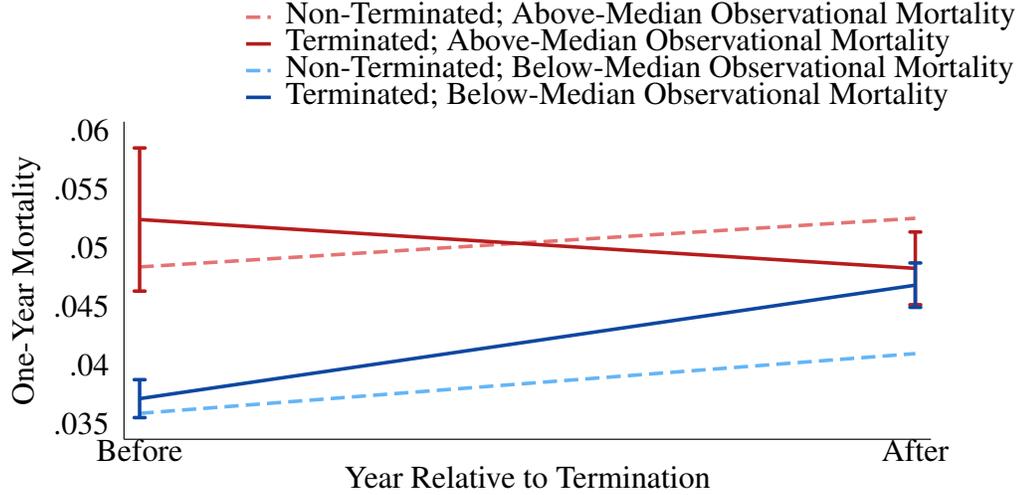
Notes: This figure shows regression-adjusted trends in the observational mortality for enrollees in non-terminated and terminated MA plans, separately for plans with above- and below-median observational mortality. The median is defined over the entire IV sample. Data is plotted in the last year prior to termination for terminated plans and the following year. Termination effects are estimated in each year and median group by a separate regression which controls for county-by-year fixed effects; flexible interactions of lagged plan type, and lagged market shares; and beneficiary demographics (age in 5-year bands, sex, race and dual-eligibility status). County-clustered 95% confidence intervals for the termination effects are shown in brackets.

observational mortality for beneficiaries before and after a plan termination, separately for beneficiaries previously enrolled in plans with above-median (blue) and below-median (red) 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 similar mortality 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 very different plans provided their plans remain available. Bracketed 95% confidence intervals show that the post-termination difference in observational mortality is statistically significant for both high- and low-mortality plans, despite terminated and non-terminated plans having statistically indistinguishable observational mortality prior to termination.

Figure 3 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

¹¹Specifically, following our IV specifications below, we adjust for county-by-year fixed effects; flexible interactions of lagged plan type, and lagged market shares and beneficiary demographics (age in 5-year bands, sex, race and dual-eligibility status).

Figure 3: Plan Terminations and Beneficiary Mortality



Notes: This figure shows regression-adjusted trends in the one-year mortality of enrollees of non-terminated and terminated MA plans, separately for plans with above- and below-median observational mortality. The median is defined over the entire IV sample. Data is 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 which controls for county-by-year fixed effects; flexible interactions of lagged plan type, and lagged market shares; and beneficiary demographics (age in 5-year bands, sex, race and dual-eligibility status). County-clustered 95% confidence intervals for these effects are shown in brackets.

summarized in Figure 2. Unlike with the average μ_j of enrolled plans, mortality risk increases with age, such that the beneficiaries in non-terminated plans (dashed blue and red lines) exhibit an increasing trend in realized mortality. However, the solid blue line (indicating the realized mortality of beneficiaries enrolled in a low-mortality plan prior to termination) exhibits a steeper trend while the solid red 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 2 and 3 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} .¹² This result suggests that the sizable variation in observational mortality we find in Figure 1 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. We next develop an econometric framework to formalize this logic and measure precisely the predictive validity of observational mortality for such causal effects.

¹²By comparing termination effects across above- and below-median plans in these figures, we allow for a direct “disruption” effect of terminations, provided it is common to both groups. We formalize this feature of our econometric approach in the following section.

3 Econometric Framework

We use an instrumental variables (IV) framework leveraging plan terminations to measure the predictive validity of observational mortality differences in predicting 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 term the fallback condition, which is novel to this paper.

3.1 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 3.4 below. 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

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

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

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

In contrast to the regression model (1) in the previous section, 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 next discuss our approach to quantifying variation in the plan mortality effects in light of these challenges.¹³

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

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

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 on average unbiased predictor of causal mortality effects when $\lambda = 1$, while observational mortality has little association with true causal effects when λ is small.¹⁴ 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. $\mu_j = \beta_j$).¹⁵ 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, $Var(\beta_j) \geq \lambda^2 Var(\mu_j)$.

¹³Estimating β_j would also generally require structural assumptions that our approach does not impose. See Geweke et al. (2003) and Hull (2020) for applications of such models to estimate hospital mortality effects.

¹⁴This 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 et al., 2014; Angrist et al., 2017).

¹⁵Chetty et al. (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 $\mu_j = \beta_j$ and $\lambda = 1$.

While 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

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

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.

Equation (5) is again a causal model, linking observational mortality μ_{it} to realized mortality Y_{it} via the forecast coefficient λ . As with the initial causal model (3), OLS estimation of Equation (5) will be biased when consumers of different unobserved health sort non-randomly into plans.¹⁶ To estimate the forecast coefficient, we instead use an IV approach that follows the logic of Figures 2 and 3. This approach leverages 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). There is, however, a cost to simplifying Equation (3), captured by the additional residual term η_{it} . We next discuss this cost in formalizing our IV approach.

3.3 Identification

Intuition and 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 that 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 $\lambda = 1/2$, 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 then partly reflect selection bias and not causal effects. Such intuition mirrors the motivation for quasi-experimental evaluations of observational quality measures in other settings (e.g. Kane and Staiger, 2008; Chetty et al., 2014; Angrist et al., 2016; Doyle et al., 2019).

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 ,

¹⁶In 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 .

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 residual, 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 naïve 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. (2019) 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 et al. (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.¹⁷

We next formalize such conditions in a novel and general way. 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 .

Our approach can be contrasted with other strategies using shocks to institutional choices in

¹⁷The education value-added literature typically considers quasi-experimental tests for selection bias, which can be thought to impose the null hypothesis of $\eta_j = 0$ (e.g. Angrist et al. (2016, 2017)).

order to estimate causal effects. Abowd et al. (1999) estimate worker and firm premiums using a two-way fixed effects model. This 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 a similar approach 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. Below 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 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.

The First-Stage and Balance Assumptions Our approach to estimating the forecast coefficient uses an instrument which, as in Figures 2 and 3, leverages the interaction of past plan choice and plan terminations. Consider, for a beneficiary i observed in year t , the instrument

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

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 first 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. We discuss how we relax each of these simplifying assumptions in Section 3.4 below.

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 partialling 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 due to 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 2, 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. Since Z_{it} is given by the interaction of terminations and lagged observational mortality, and since both Z_{it} and X_{it} only vary at the lagged plan level, we only require that any relationship between observational mortality and the average unobserved health of a plan’s beneficiaries is the same for terminated and non-terminated plans. Formally, we can evaluate Assumption 2 in terms of the infeasible plan-level difference-in-differences regression,

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

where $\bar{\varepsilon}_{jt} = E[\varepsilon_{it} \mid D_{ij,t-1} = 1]$ 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} . 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 non-terminated 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.¹⁸ The similarity of the pre-period mortality in Figure 3 supports the stronger version of Assumption 2 in our setting; we develop and apply additional falsification tests of the sufficient balance assumption in Section 4.1 below.

¹⁸To 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.

The Fallback Condition The third identification condition we consider 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 . For the instrument to be relevant, \tilde{Z}_{it} must be correlated with plan choice D_{ijt} so 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 this correlation out, 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 which 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.¹⁹

The fallback condition can be viewed (as with Assumption 2) as restricting the relationship between observational mortality and a particular plan-level unobservable to be similar across terminated and non-terminated 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 non-terminated plans must be the same. This plan-level interpretation gives some intuition for the behavioral restrictions that might be sufficient. The 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. Appendix C.4 presents a discrete choice model that yields such restrictions. The simplest version of the model assumes that consumers in non-terminated plans are fully inertial, while consumers in

¹⁹In Appendix C.3 we derive an alternative “monotonicity” condition which permits interpretation of the forecast IV regression coefficient as a weighted regression of β on μ with some convex (and estimable) weights. This condition is less plausible in our setting and identifies a less policy-relevant parameter if it holds than the unweighted λ .

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. Suppose, for example, that consumer utility has the form

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

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 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 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 non-terminated plans need not be fully inertial; the same logic holds in models with partial inertia such as that of (Ho et al., 2017).²⁰

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 et al. (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.²¹

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

²⁰We emphasize that our microfoundation does not assume consumers “select on μ_j ” but not on the unobservable η_j . While 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.

²¹Suppose, 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.

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} which enters Assumption 2. However, just as standard IV falsification tests can investigate whether the instrument is correlated with observable proxies of such 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 non-terminated plans. We conduct this test in the MA setting below.

3.4 Extensions

We consider four extensions to our basic econometric framework before bringing it to the data. First, we note that while 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 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 above. 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 one (see Appendix Figure A2).

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 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 while 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 5 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.

4 Results

4.1 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

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

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).²² In some specifications we also include controls for beneficiary demographics (age in 5-year bands, sex, race and dual-eligibility status). We cluster standard errors at the county level, allowing for arbitrary correlation in the regression residual across different beneficiaries, plans, and years.²³

First-stage coefficient estimates are reported in Panel A of Table 2. The finding of $\pi_Z < 0$ is consistent with a combination of inertia and regression-to-the-mean in MA plan choice, first doc-

²²Plan type distinguishes traditional Medicare from several private alternatives: health maintenance organizations, local and regional preferred provider organizations, private fee-for-service plans, and demonstration plans.

²³The 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.

Table 2: Tests of Assumptions

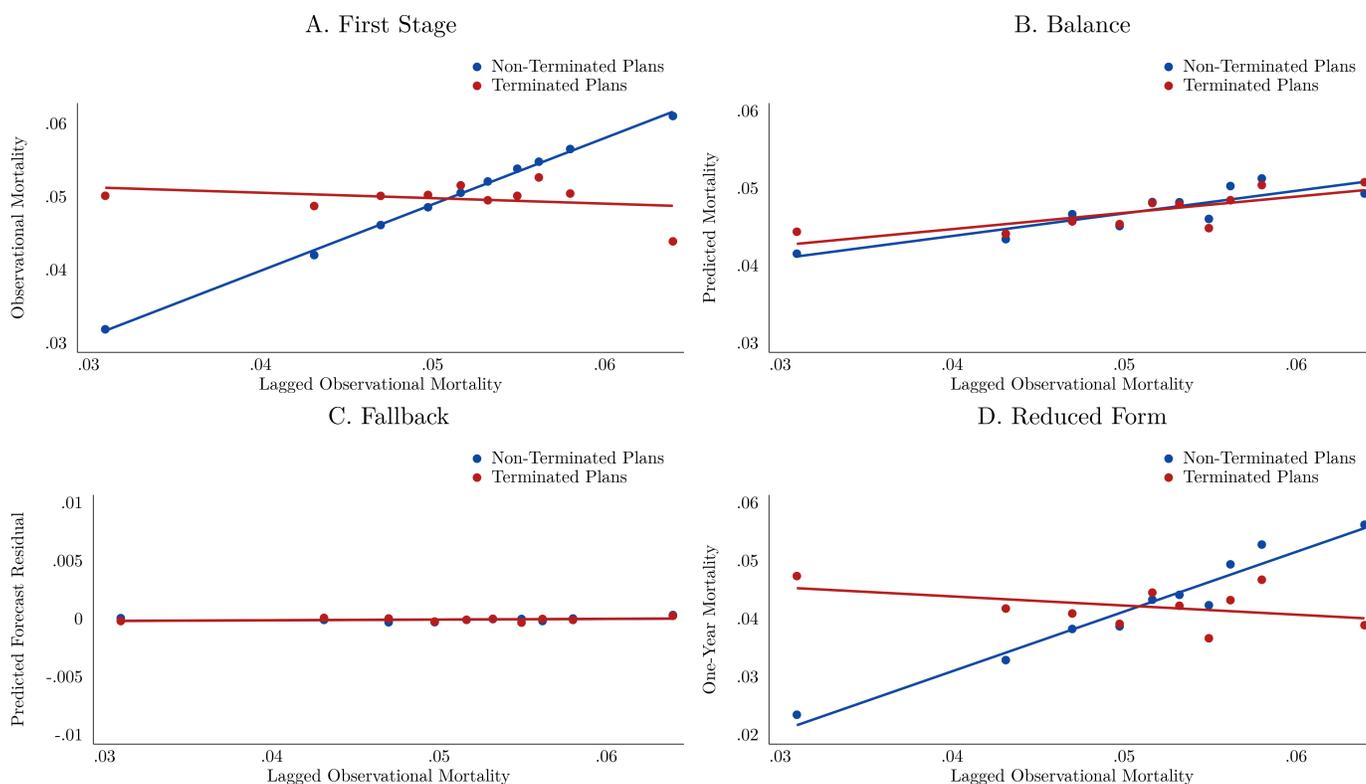
	(1)	(2)
<i>Dep. Var.: Observational Mortality</i>		
	A. First Stage	
Instrument	−0.724 (0.015)	−0.0189 (0.0014)
F Statistic	2,358.1	173.9
<i>Dep. Var.: Predicted Mortality</i>		
	B. Balance	
Instrument	−0.020 (0.013)	−0.0011 (0.0006)
<i>Dep. Var.: Predicted Forecast Residual</i>		
	C. Fallback	
Instrument	0.010 (0.001)	−0.0000 (0.0001)
Specification	Linear	Median
Demographic Controls	No	No
N Beneficiary-Years	11,441,205	

Notes: Panel A of this table 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.

umented in Figure 2. Beneficiaries enrolled in high- or low-mortality plans that are terminated in year $t - 1$ tend to choose plans in year t which are more typical in terms of observational mortality, relative to the mostly inertial beneficiaries in non-terminated 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 one percentage point higher observational mortality plan in the previous period switches to a plan with 0.72 percentage points 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 2, 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).

Panel A of Figure 4 illustrates the first-stage relationship by replacing the linear instrument in

Figure 4: Graphical Tests of Assumptions and the Reduced Form



Notes: This figure illustrates the three assumptions in our IV approach and the IV reduced form. Panel A shows average observational mortality by deciles of lagged observational mortality among non-terminated 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, MLRs, 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 lag 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 as in Table 2, including decile main effects. Coefficients are normalized to remove termination main effects.

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 while observational mortality of the lagged plan predicts current plan observational mortality among the non-terminated 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 Panel A of Table 2.²⁴

We next build support for the balance 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 5-year age bands, sex, race, and dual-eligibility fixed effects (see Appendix Table A3 for model estimates). The results are in Panel B of Table 2. 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 of these estimates are more than an order of magnitude smaller than the corresponding first-stage estimates. Finding balance for our instrument on predicted mortality is not surprising in light of the motivating Figure 3.

Panel B of Figure 4 illustrates the predicted mortality regressions by replacing the observational mortality measure in Panel A. We plot the average predicted mortality among terminated and non-terminated 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 non-terminated plans. Any differential trend in the actual mortality of beneficiaries in terminated and non-terminated plans is therefore unlikely to be due to pre-existing differences in their health. Appendix Figure A3 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 Appendix Table A4.²⁵

Finally, we build support for the novel fallback condition (Assumption 3) by testing whether

²⁴We normalize the height of each set of terminated and non-terminated 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.

²⁵The appendix table shows that while 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 small quantitatively: for example, in the linear specification in Table A4 we find that a one percentage point higher observational mortality plan differentially enrolls individuals younger by 0.04 years in terminated vs. non-terminated plans. We further explore the quantitative importance of these imbalances in Appendix C.7, as discussed below.

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. Panel C of Table 2 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 zero in the median specification. While statistically significant, the linear imbalance is quantitatively negligible—almost two orders of magnitude smaller than the associated first-stage effect. In Appendix C.7, we show how the frameworks of Altonji et al. (2005) and Oster (2019) can be adopted to quantify the importance of imbalances on both beneficiary and plan-level observables. The statistical imbalances are too small to substantially alter our forecast coefficient estimates, even under conservative assumptions.

Panel C of Figure 4 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 that termination-induced changes in observational mortality can be related to termination-induced changes in actual mortality to estimate the MA forecast coefficient. We next present these IV estimates.

4.2 Forecast Coefficient Estimates

Table 3 reports first-stage, reduced-form, and second-stage estimates for our main IV specification. The second-stage estimates come from a regression of

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

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 both this linear specification and an alternative specification which replaces the instrument with one constructed from an above-median lag observational mortality indicator. We also report two specifications for the control vector X_{it} ; one which mirrors the tests of our assumptions, and a second which 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 above, we do not expect the inclusion of these controls to meaningfully affect the IV estimates (though it may increase their precision).

Table 3: Forecast Coefficient Estimates

	(1)	(2)	(3)	(4)
<i>Dep. Var.: Observational Mortality</i>				
	A. First Stage			
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
<i>Dep. Var.: One-Year Mortality</i>				
	B. Reduced Form			
Instrument	-0.764 (0.069)	-0.0214 (0.0025)	-0.745 (0.069)	-0.0203 (0.0023)
<i>Dep. Var.: One-Year Mortality</i>				
	C. Second Stage (Forecast Coefficient)			
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 of this table 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.

Panel A of Table 3 replicates the first-stage results reported in Panel A of Table 2 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 of -0.0214 and -0.0203 for the median specification. Each of these estimates are quite similar to the corresponding first-stage coefficients, reflecting the pattern first shown in Figures 2 and 3: terminations tend to shift observational mortality and realized mortality by similar amounts.

Panel C of Table 3 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 baseline 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. $\lambda \approx 1$).

Panel D of Figure 4 illustrates this finding by plotting reduced-form variation in one-year mortality rates for beneficiaries in terminated and non-terminated 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 one. Lagged observational mortality strongly predicts the subsequent mortality of beneficiaries previously enrolled in non-terminated 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 Panel B of Figure 4, which shows no such relationship for predicted one-year mortality. Beneficiaries in high- and low-mortality terminated plans appear similar to those in corresponding non-terminated plans until they are induced by terminations to choose more average plans.²⁶

4.3 Robustness Checks

We verify the robustness of our forecast coefficient estimates in a number of exercises summarized in Appendix Table A6. First, we show that the estimates in Table 3 are unaffected by the removal of counties which do not see a plan termination during our sample period. The first row of Appendix Table A6 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 Appendix Figure A1) and that counties with and without terminations are broadly similar (see Appendix Table A1).

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). While this specification may be biased by selecting on an endogenous variable, we nevertheless obtain similar forecast coefficients in the second row of Appendix Table A6.

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 which

²⁶The similarity in forecast coefficient estimates for the linear and median specifications of Table 3 reflects a homogeneity in the first-stage and reduced-form relationship that is apparent across the support of lagged observational mortality. Appendix Figure A4 shows this homogeneity more directly, by plotting the implied first-stage and reduced-form effects from Panels A and D of Figure 4. Panel A of Appendix Figure A4 shows that these effects track each other closely across all ten 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.

increased PFFS operating costs. While 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 Appendix Table A6 shows that these plan terminations yield a similar 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 standard error of 0.12. The fourth row of Appendix Table A6 reports the results of excluding non-PFFS Terminations. Forecast coefficient estimates from this specification are more imprecise, but not statistically distinguishable from our baseline estimates.

We next investigate the role of treatment effect heterogeneity. The fifth row of Appendix Table A6 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 naïve observational mortality estimates, as discussed in Section 2.3. The findings suggest either that treatment effect heterogeneity is not first-order in this setting, or that the extension of our framework in Appendix C.6 (that accommodates such heterogeneity) is likely to hold.

We conclude this section by summarizing a number of additional robustness checks in Appendix Table A6. We find similar results in the seventh row when including lagged plan risk scores as a control, consistent with the visual balance in Appendix Figure A3. We also find in row eight qualitatively similar (but slightly smaller) results when not shrinking the observational mortality estimates, consistent with the fact that the typical estimate is precise (per Appendix Figure A2) and the discussion in Appendix C.5. We further confirm in row nine 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 observed differences in plan mortality rates can be highly predictive of true plan causal effects at different horizons of beneficiary mortality.

4.4 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 Appendix C.4), we allow beneficiary preferences to correlate with their health in both 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).²⁷ 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 2.3.

While 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 et al. (2008), for example, estimate a 20% mortality reduction in Medicare beneficiaries who are admitted to emergency departments.²⁸ The literature on place-based mortality effects estimates similarly large variation within Medicare across all elderly beneficiaries, though these may capture both the joint impact of changing health systems and other demand side factors.²⁹ Below, we further argue that evidence on provider effects is consistent with the magnitudes we document.

5 Correlates of Plan Effects

When combined with our observational mortality estimates, a forecast coefficient close to one implies large differences across plans in causal mortality effects. 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 willingness to pay for plan health effects.

²⁷Formally, note that the forecast coefficient can be written $\lambda = \frac{Cov(\beta_j, \mu_j)}{Var(\mu_j)} = \frac{Var(\beta_j) + Cov(\beta_j, b_j)}{Var(\beta_j) + Var(b_j) + 2Cov(\beta_j, b_j)}$, where $b_j = \mu_j - \beta_j$ denotes selection bias for plan j . A forecast coefficient of $\lambda \approx 1$ can arise with non-zero bias when $Cov(\beta_j, b_j) \approx -Var(b_j)$, 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.

²⁸A growing literature also shows that insurance lowers mortality in the Medicaid program (Miller et al., 2019; Goldin et al., 2021). A 19% reduction in mortality within the MA program is thus within the range of the estimated extensive-margin effect of gaining health insurance more broadly (Sommers et al., 2017).

²⁹Finkelstein et al. (2019) 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.

5.1 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.³⁰

Surprisingly, we find that CMS star ratings are positively correlated with our observational mortality measure, suggesting higher-ranked plans have higher mortality rates.³¹ The first column of Table 4, Panel A, shows that a one-star increase in a 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 eliminate selection bias, we next 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),

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

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 plan 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 Panel B in Table 4 shows that a one-unit increase in star ratings is associated with a smaller 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.

³⁰See 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.

³¹We 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 4: 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 3. 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 non-missing 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 3.

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 5.3 below). In the second column of Panel A in Table 4 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 Panel B in Table 4 suggests that a \$100 increase in monthly premiums (\$1,200 per 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 below. 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.³²

Although premiums (in contrast with star ratings) significantly predict plan mortality effects, they similarly explain a small share of quality variation. Since 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.³³ 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 which 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 impact 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 next turn to other plan characteristics.

³²At a conservative \$1 million VSL, a 0.5 percentage point reduction in mortality is worth \$5,000.

³³Formally, $Var(W_j'\theta)/Var(\beta_j) \leq Var(W_j'\theta)/Var(\lambda\mu_j)$ since $Var(\beta_j) \geq Var(\lambda\mu_j)$. To estimate the maximum R^2 in Table 4 we compute beneficiary-weighted variances of $W_j'\hat{\theta}$ and divide by beneficiary-weighted variances of $\hat{\lambda}\mu_j$ where $\hat{\lambda}$ comes from column 3 of Table 3.

5.2 Mechanisms

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

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 Panels A and B of Table 4 we find that plans which offer 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)).³⁴ The finding of large plan effect differences among plans which offer donut hole coverage suggests that lower cost-sharing may be more broadly beneficial.³⁵

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 next study whether mortality effects correlate with overall expenditures, as measured by medical loss ratios (MLRs): the percentage of premiums which are paid out in claims.³⁶ In Panels A and B of Table 4 we find that plans with higher MLRs tend to have higher observational mortality, but significantly lower plan effects. A ten 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 that 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 (Hull, 2020; Doyle et al., 2015; Geweke et al., 2003), 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 observational 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.³⁷

³⁴Despite this, Yang, Gilleskie, and Norton (2009) argue plans with prescription drug coverage increase survival.

³⁵At a \$1 million VSL, the social value of more generous drug coverage is almost \$5,000 per year.

³⁶Due 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 Appendix B for details.

³⁷Hospital network data is available from State Inpatient Databases, but consistent information on Medicare Ad-

Overall, this analysis of mechanisms paints a clear and consistent picture. More expensive and higher spending plans tend to reduce beneficiary mortality while also 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 4, which includes all financial measures.³⁸ 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.

5.3 Demand for Plan Quality

We next 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 willingness to pay (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 to maximize their latent utility U_{ij} , given by

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

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 $\tau/(100 \times \alpha)$ captures the WTP for plan quality: the decrease in premiums sufficient to offset a one 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

vantage 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.

³⁸We do not simultaneously include all five characteristics in Table 4 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.

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. Naïvely 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. non-zero η_j) that may add further bias.³⁹

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 systematic component of consumer utility, which we denote δ_j :

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

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

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

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 then 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, $\tau/(100 \times \alpha)$.

We show this calculation in Table 5 for a range of possible premium elasticities given in the first column.⁴⁰ In column 2, we translate these elasticities to a value for α , 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 5.1 and 5.2. For each premium elasticity we

³⁹Alternative 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.

⁴⁰Curto et al. (2015) estimate an elasticity of -7 in this setting. Elasticities less than one in magnitude are implausible, since they are inconsistent with insurer profit maximization; nevertheless we include an elasticity of -0.5.

Table 5: Willingness to Pay Bounds

Premium Elasticity	Premium Coefficient (α)	Forecast Coefficient (κ)	Minimum Quality Coefficient (τ)	Maximum WTP: $\tau/(100 \times \alpha)$
(1)	(2)	(3)	(4)	(5)
-10	-0.0225	-0.0003 (0.0001)	-618.76 (133.87)	275.08 (59.51)
-7	-0.0157	-0.0004 (0.0001)	-444.40 (94.82)	282.24 (60.22)
-3.5	-0.0079	-0.0009 (0.0003)	-242.31 (49.90)	307.78 (63.38)
-1	-0.0022	-0.0011 (0.0003)	-93.61 (21.49)	416.14 (95.53)
-0.5	-0.0011	-0.0025 (0.0013)	-46.53 (16.41)	413.72 (145.91)

Notes: Column 5 of this table reports estimates of the upper bound on quality willingness to pay (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 3 (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 3. Column 4 translates the forecast coefficient estimate to an estimate of the quality utility coefficient bound described in the text. Standard errors are clustered by county and reported in parentheses.

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)).⁴¹ Column 4 of Table 5 uses these estimates to compute our upper bound on τ , while column 5 reports our corresponding estimates of the WTP for a one percentage point increase in plan quality.

For the wide range of possible premium elasticities, we obtain estimated WTP bounds of around \$275–\$416, implying that consumers are willing to pay no more than this amount to offset a one percentage point increase in one-year mortality.⁴² These estimates are around half of the

⁴¹Appendix Table A7 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 Blue plans) appear to supply higher quality plans. These are interesting avenues for future research.

⁴²Naïve 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.

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 \$2 million by age 80; see Kniesner and Viscusi (2019) and Murphy and Topel (2006)). In Appendix Table A9, 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 one 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 at least an order of magnitude lower.

The finding that consumers are insensitive to plan mortality effects is broadly consistent with a literature demonstrating that consumers overweigh 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. While 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 4.2 suggest that MA plan mortality effects are enormously variable within 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 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: 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 pp per year. Given our VSL estimates, this implies that the potential benefits from reduced mortality risk in such scenarios could be quite substantial.⁴³

6 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 then show that this variation unbiasedly predicts true plan mortality effects with a novel quasi-experimental design. Publicly available quality measures are uncorrelated with true mortality effects. Perhaps as a result,

⁴³We note that the simulations are partial equilibrium and do not consider other important plan attributes, including financial characteristics. We discuss additional limitations in Appendix D.

consumer demand is under-responsive 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 financial consequences 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 is needed to fully investigate the role of plan networks.

Methodologically, this paper 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 which 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 willingness to pay 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 Koijen 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.⁴⁴ Furthermore, our model does not allow for capacity constraints or for premiums and quality to adjust with demand. Such effects could offset our implied gains, although the health effects are large enough that they are likely to be first-order.⁴⁵ 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.

⁴⁴Existing programs subsidize plans that score better on measures like star ratings, which we find to be uncorrelated with causal mortality effects. While such programs might be improved by targeting risk-adjusted mortality, though this could lead to insurer gaming. See Decarolis and Guglielmo (2017) for an analysis of the MA Quality Bonus Payment Demonstration program.

⁴⁵Nevertheless, 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, J. and A. Adams (2017). What do consumers consider before they choose? Identification from asymmetric demand responses. NBER Working Paper No. 23566.
- Abaluck, J. and J. Gruber (2011). Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program. *American Economic Review* 101(4), 1180–1210.
- Abaluck, J. and J. Gruber (2016a). Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review* 106(8), 2145–84.
- Abaluck, J. and J. Gruber (2016b). Improving the quality of choices in health insurance markets. NBER Working Paper No. 22917.
- Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High wage workers and high wage firms. *Econometrica* 67(2), 251–333.
- Allende, C. (2019). Competition under social interactions and the design of education policies. Working Paper.
- Altonji, J., T. Elder, and C. Taber (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy* 113(1), 151–184.
- Angrist, J., P. Hull, P. Pathak, and C. Walters (2016). Interpreting tests of school VAM validity. *American Economic Review: Papers & Proceedings* 106(5), 388–392.
- Angrist, J., P. Hull, P. Pathak, and C. Walters (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics* 132(2), 871–919.
- Atherly, A., R. D. Feldman, B. Dowd, and E. Van Den Broek-Altenburg (2020). Switching costs in Medicare Advantage. In *Forum for Health Economics & Policy*, Volume 1. De Gruyter.
- Berry, S., A. Gandhi, and P. Haile (2013). Connected substitutes and invertibility of demand. *Econometrica* 81(5), 2087–2111.
- Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63(4), 841–890.
- Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics* 25(2), 242–262.
- Brown, J., M. Duggan, I. Kuziemko, and W. Woolston (2014). How does risk-selection respond to risk-adjustment: Evidence from the Medicare Advantage program. *American Economic Review* 104(10), 3335–64.
- Cabral, M., M. Geruso, and N. Mahoney (2018). Do larger health insurance subsidies benefit patients or producers? Evidence from Medicare Advantage. *American Economic Review* 108(8), 2048–2087.

- Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics* 36(S1), S13–S70.
- Card, D., C. Dobkin, and N. Maestas (2008). The impact of nearly universal insurance coverage on health care utilization: Evidence from Medicare. *American Economic Review* 98(5), 2242–58.
- Carroll, C. (2019). Impeding access or promoting efficiency? Effects of rural hospital closure on the cost and quality of care. Working Paper.
- Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2015). Healthcare exceptionalism? Performance and allocation in the U.S. healthcare sector. *American Economic Review* 106(8), 2110–44.
- Chernew, M. E., W. P. Wodchis, D. P. Scanlon, and C. G. McLaughlin (2004). Overlap in HMO physician networks. *Health Affairs* 23(2), 91–101.
- Chetty, R., J. Friedman, and J. Rockoff (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review* 104(9), 2593–2632.
- Chetty, R. and N. Hendren (2018). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Quarterly Journal of Economics* 133(3), 1107–1162.
- Chyn, E. (2018). Moved to opportunity: The long-run effects of public housing demolition on children. *American Economic Review* 108(10), 3028–56.
- Commission, M. P. A. (2001). Report to the congress: Medicare in rural America. Technical report, Medicare Payment Advisory Commission.
- Commission, P. P. A. (1998). Medicare prospective payment and the American health care system report to the congress. Technical report, Prospective Payment Assessment Commission.
- Curto, V., L. Einav, J. Levin, and J. Bhattacharya (2015). Can health insurance competition work? Evidence from Medicare Advantage. NBER Working Paper No. 20818.
- Dafny, L. and D. Dranove (2008). Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. *RAND Journal of Economics* 39(3), 790–821.
- Darden, M. and I. McCarthy (2015). The star treatment estimating the impact of star ratings on Medicare Advantage enrollments. *Journal of Human Resources* 50(4), 980–1008.
- Decarolis, F. (2015). Medicare Part D: Are insurers gaming the low income subsidy design? *American Economic Review* 105(4), 1547–80.
- Decarolis, F. and A. Guglielmo (2017). Insurers' response to selection risk: Evidence from Medicare enrollment reforms. *Journal of Health Economics* 56, 383–396.
- Deryugina, T. and D. Molitor (2020). Does when you die depend on where you live? Evidence from Hurricane Katrina. *American Economic Review* 110(11), 3602–3633.

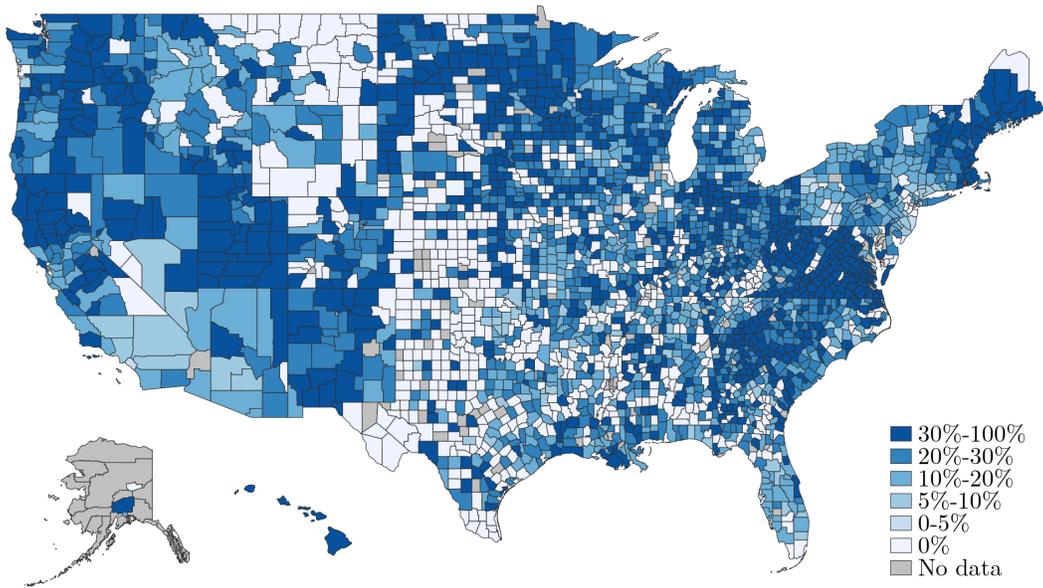
- Dillender, M. (2018). What happens when the insurer can say no? Assessing prior authorization as a tool to prevent high-risk prescriptions and to lower costs. *Journal of Public Economics* 165, 170–200.
- Doyle, J., J. Graves, and J. Gruber (2019). Evaluating measures of hospital quality: Evidence from ambulance referral patterns. *Review of Economics and Statistics* 101(5), 841–852.
- Doyle, J., J. Graves, J. Gruber, and S. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy* 123(1), 170–214.
- Drake, C., C. Ryan, and B. E. Dowd (2020). Sources of inertia in health plan choice in the individual health insurance market. Working Paper.
- Duggan, M., A. Starc, and B. Vabson (2016). Who benefits when government pays more? Pass-through in the Medicare Advantage program. *Journal of Public Economics* 141, 50–67.
- Ericson, K. and A. Starc (2016, December). How product standardization affects choice: Evidence from the Massachusetts health insurance exchange. *Journal of Health Economics* 50, 71–85.
- Finkelstein, A., M. Gentzkow, P. Hull, and H. Williams (2017). Adjusting risk adjustment – accounting for variation in diagnostic intensity. *New England Journal of Medicine* 376(7), 608–610.
- Finkelstein, A., M. Gentzkow, and H. L. Williams (2019). Place-based drivers of mortality: Evidence from migration. NBER Working Paper No. 25975.
- Fletcher, J., L. Horwitz, and E. Bradley (2014). Estimating the value added of attending physicians on patient outcomes. NBER Working Paper No. 20534.
- Gaynor, M., R. Moreno-Serra, and C. Propper (2013). Death by market power: Reform, competition, and patient outcomes in the National Health Service. *American Economic Journal: Economic Policy* 5(4), 134–66.
- Geruso, M., T. J. Layton, and J. Wallace (2020). Are all managed care plans created equal? evidence from random plan assignment in medicaid. NBER Working Paper No. 27762.
- Geweke, J., G. Gowrisankaran, and R. J. Town (2003). Bayesian inference for hospital quality in a selection model. *Econometrica* 71(4), 1215–1238.
- Gibbons, R. and L. Katz (1992). Does unmeasured ability explain inter-industry wage differentials? *The Review of Economic Studies* 59(3), 515–535.
- Goldin, J., I. Z. Lurie, and J. McCubbin (2021). Health insurance and mortality: Experimental evidence from taxpayer outreach. *Quarterly Journal of Economics* 136(1), 1–49.
- Gruber, J., B. R. Handel, S. H. Kina, and J. T. Kolstad (2020). Managing intelligence: Skilled experts and AI in markets for complex products. NBER Working Paper No. 27038.
- Handel, B. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review* (103), 2643–2682.

- Handel, B. R. and J. T. Kolstad (2015). Getting the most from marketplaces: Smart policies on health insurance choice. Brookings Hamilton Project Discussion Paper 2015-08.
- Heiss, F., D. McFadden, J. Winter, A. Wuppermann, and B. Zhou (2016). Inattention and switching costs as sources of inertia in Medicare Part D. NBER Working Paper No. 22765.
- Ho, K., J. Hogan, and F. S. Morton (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *RAND Journal of Economics* 48(4), 877–905.
- Hull, P. (2018). Estimating treatment effects in mover designs. Working Paper.
- Hull, P. (2020). Estimating hospital quality with quasi-experimental data. Working Paper.
- Imbens, G. and J. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.
- Jacob, B. A. (2004). Public housing, housing vouchers, and student achievement: Evidence from public housing demolitions in Chicago. *American Economic Review* 94(1), 233–258.
- Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. NBER Working Paper No. 14607.
- KFF (2019). A dozen facts about Medicare Advantage in 2019. <https://www.kff.org/medicare/issue-brief/a-dozen-facts-about-medicare-advantage-in-2019/>. Accessed: 2021-03-05.
- KFF (2021). Medicare Advantage 2021 spotlight: First look. <https://www.kff.org/medicare/issue-brief/medicare-advantage-2021-spotlight-first-look/>. Accessed: 2021-03-05.
- Kniesner, T. J. and W. K. Viscusi (2019). The value of a statistical life. *Oxford Research Encyclopedia of Economics and Finance*, 19–15.
- Koijen, R. S. and S. Van Nieuwerburgh (2020). Combining life and health insurance. *Quarterly Journal of Economics* 135(2), 913–958.
- Krueger, A. B. and L. H. Summers (1988). Efficiency wages and the inter-industry wage structure. *Econometrica* 56(2), 259–293.
- McGuire, T., J. Newhouse, and A. Sinaiko (2011). An economic history of Medicare Part C. *Milbank Quarterly* 89(2), 289–332.
- Miller, K. S., A. Petrin, R. Town, and M. Chernew (2019). Optimal managed competition subsidies. NBER Working Paper No. 25616.
- Miller, S., S. Altekruuse, N. Johnson, and L. R. Wherry (2019). Medicaid and mortality: New evidence from linked survey and administrative data. NBER Working Paper No. 26081.
- Morris, C. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association* 78(381), 47 – 55.

- Murphy, K. M. and R. H. Topel (1987). Unemployment, risk, and earnings: Testing for equalizing wage differences in the labor market. *Unemployment and the Structure of Labor Markets*, 103–140.
- Murphy, K. M. and R. H. Topel (2006). The value of health and longevity. *Journal of Political Economy* 114(5), 871–904.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business and Economic Statistics* 37(2), 187–204.
- Pakes, A., J. Porter, K. Ho, and J. Ishii (2015). Moment inequalities and their application. *Econometrica* 83(1), 315–334.
- Pelech, D. (2018). Paying more for less? Insurer competition and health plan generosity in the Medicare Advantage program. *Journal of Health Economics* 61, 77–92.
- Polyakova, M. (2016). Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D. *American Economic Journal: Applied Economics* 8(3), 165–195.
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy* 4(4), 537–571.
- Sabety, A. (2020). The value of relationship-specific capital in health care. Working Paper.
- Sommers, B. D., A. A. Gawande, and K. Baicker (2017). Health insurance coverage and health—what the recent evidence tells us. *New England Journal of Medicine* 377(6), 586–593.
- Staiger, B. (2020). Disruptions to the provider-patient relationship and patient utilization and outcomes: Evidence from Medicaid Managed Care. Working Paper.
- Starc, A. (2015). Insurer pricing and consumer welfare: Evidence from Medigap. *RAND Journal of Economics* 45(1), 198–220.
- Starc, A. and R. J. Town (2020). Externalities and benefit design in health insurance. *The Review of Economic Studies* 87(6), 2827–2858.
- Yakusheva, O., R. Lindrooth, and M. Weiss (2014). Nurse mortality-added and patient outcomes in acute care. *Health Services Research* 49(6), 1767–1786.
- Yang, Z., D. B. Gilleskie, and E. C. Norton (2009). Health insurance, medical care, and health outcomes a model of elderly health dynamics. *Journal of Human Resources* 44(1), 47–114.

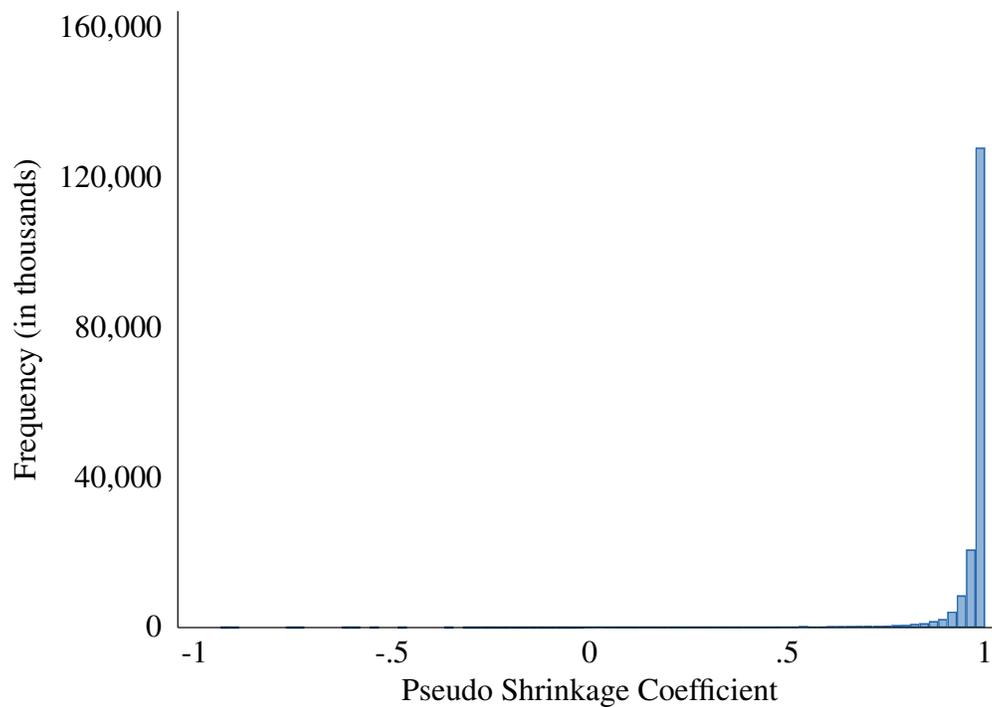
A Appendix Figures and Tables

Figure A1: Geographic Distribution of Plan Terminations



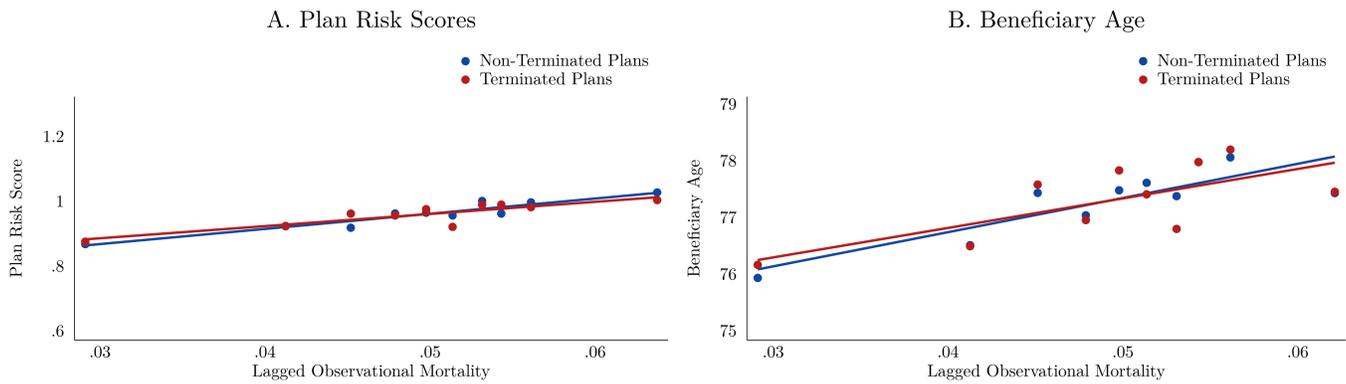
Notes: This map shows the fraction of plans in a county that were terminated over 2008-2011, with counties shaded according to the quantiles reported in the legend.

Figure A2: Distribution of Observational Mortality Shrinkage Coefficients



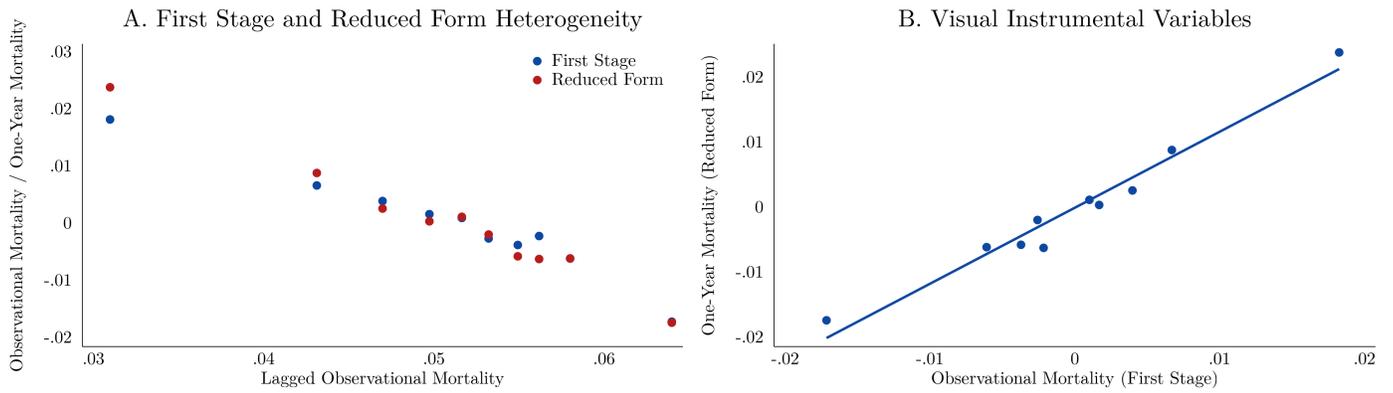
Notes: This figure shows the distribution of “pseudo shrinkage coefficients” for observational mortality, given by the ratio of each plan’s de-meaned posterior to the de-meaned OLS estimate, across beneficiary-years in our main sample. A coefficient close to one thus implies minimal shrinkage. This coefficient can be negative under the hierarchical shrinkage procedure described in Appendix C.1.

Figure A3: Balance on Plan Risk Scores and Beneficiary Age



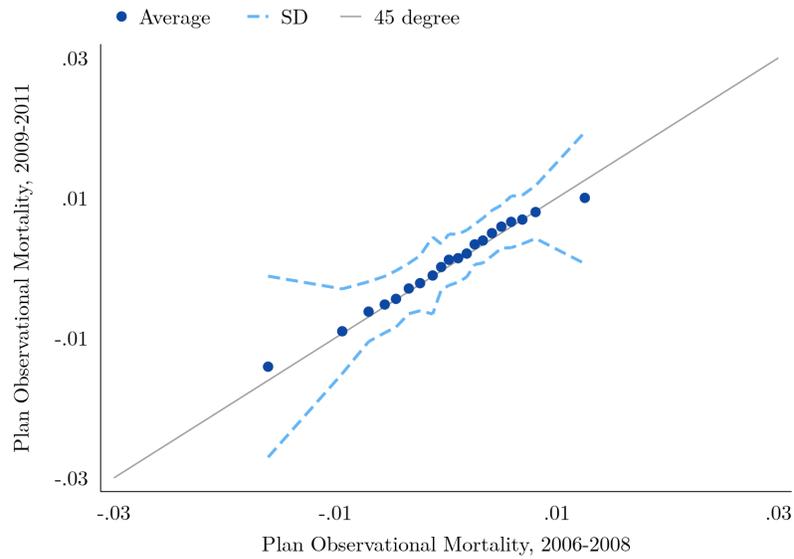
Notes: This figure shows average beneficiary risk scores and beneficiary age by deciles of lagged observational mortality among non-terminated and terminated plans (with non-missing risk score averages), controlling for county-by-year fixed effects and other observables in our baseline specification. The points in each panel are given by the average dependent variable, 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 as in Table 2, including decile main effects. Coefficients are normalized to remove termination main effects.

Figure A4: Forecast Regression First-Stage and Reduced-Form Estimates



Notes: Panel A shows the estimated effects of lagged plan termination on enrolled plan observational mortality (first stage) and average one-year mortality (reduced form) by deciles of lagged observational mortality. Panel B plots these estimates against each other. The slope of the line of best fit through the origin gives an IV estimate of the forecast coefficient. Estimates come from Panel A and D of Figure 4; see the notes from that figure for computation details.

Figure A5: Observational Mortality Persistence: 2006-08 vs. 2009-11



Notes: This figure shows the average observational mortality of plans measured in 2006-2008 by ventiles of observational mortality measured in 2009-2011. We restrict to plans found in both halves of the data. The dashed line indicates the standard deviation of 2009-2011 observational mortality by 2006-2008 ventiles.

Table A1: County Characteristics

	No Terminations	Terminations
Median Income	40,969	45,313
% of Pop >65	17.15	16.32
% Dual	10.13	10.02
% White	83.10	84.33
% Black	9.59	8.81
% Asian	0.68	1.42
Population	34,281	124,512
Population Density	82	326
Counties	736	2,290
Number of Beneficiaries	178,527	4,379,646

Notes: This table compares the demographics of counties with and without terminations, using data taken from the 2011 American Community Survey (ACS). Population density is calculated as population per square mile. All percentage variables are calculated using total population as the denominator.

Table A2: Switching Behavior Summary Statistics

	MA (1)	Terminated (2)
% Do Not Switch	85.9	0.1
% Switch Plans within Same Insurer	2.7	18.6
% Switch Insurer	11.4	81.3
% Switch Into TM Plan	2.4	20.6
% Switch Into PFFS Plan	1.6	17.4
N Beneficiary-Years	11,442,053	322,928

Notes: This table compares choice behavior of consumers in MA plans to those in a MA plan that terminates. Market shares sum to more than one due to rounding.

Table A3: Predicted Mortality Model

	(1)
Age 70-74	0.0050 (0.0002)
Age 75-79	0.0182 (0.0002)
Age 80-84	0.0408 (0.0002)
Age 85-90	0.0799 (0.0003)
Age 90-94	0.1421 (0.0004)
Age 95+	0.2381 (0.0007)
Female	-0.0188 (0.0001)
White	0.0079 (0.0023)
Black	0.0068 (0.0023)
Other	-0.0041 (0.0023)
Asian	-0.0092 (0.0023)
Hispanic	-0.0056 (0.0023)
Native American	0.0099 (0.0027)
Dual	0.0469 (0.0002)
R^2	0.040
N Beneficiary-Years	11,442,053

Notes: This table reports coefficients of our predicted mortality regression model. Standard errors are clustered by county and reported in parentheses. The sample includes 983 singleton observations which are dropped from the main IV sample for being perfectly collinear with the fixed effects.

Table A4: Additional Balance Tests

	(1)	(2)
Age	−3.779 (1.854)	−0.245 (0.100)
Male	−0.047 (0.088)	0.002 (0.005)
White	0.123 (0.134)	0.012 (0.007)
Dual-Eligible	0.071 (0.124)	−0.003 (0.006)
Plan Risk Score	−0.617 (0.121)	−0.047 (0.009)
Predicted Mortality	−0.020 (0.013)	−0.001 (0.001)
Predicted Mortality (with Risk Score)	−0.032 (0.013)	−0.002 (0.001)
Specification	Linear	Median
N Beneficiary-Years	11,441,205	

Notes: This table follows Panel B of Table 2 in presenting OLS coefficient in a balance regression of observable controls on the instrument and controls. 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.

Table A5: Forecast Coefficient Sensitivity to Omitted Variables Bias

Max. Reduced Form R^2 , Relative to R^2 with Controls	Selection Proportionality Constant	Forecast Coefficient Bounds	
		With Demographic Controls	Demographics and Plan Characteristics
(1)	(2)	(3)	(4)
1.3	1	[1.020,1.038]	[0.996,1.026]
1.3	3	[1.001,1.057]	[0.966,1.057]
5	1	[0.904,1.154]	[0.809,1.213]
5	3	[0.654,1.404]	[0.404,1.618]

Notes: Columns 3 and 4 of this table report estimated bounds on the observational mortality forecast coefficient under different assumptions on the relationship between observable and unobservable confounders, following the framework of Altonji et al. (2005) and Oster (2019). Column 1 lists the assumed maximum reduced form R^2 , as a multiple of the reduced form R^2 when all controls are included. Column 2 lists the assumed selection proportionality constant, parameterizing how “important” the unobservable confounders are relative to the controls. Column 3 reports bounds using the beneficiary age, sex, race, and dual-eligibility status as controls. Column 4 further controls for the star rating, premium, an indicator for donut hole coverage, and medical loss ratio of the beneficiary’s enrolled plan after residualizing each plan characteristic on observational mortality. See Appendix C.7 for details.

Table A6: Forecast Coefficient Robustness Checks

	(1)	(2)
Counties With Terminations (<i>N</i> =11,156,559)	1.003 (0.098)	1.051 (0.107)
No TM Enrollments (<i>N</i> =10,910,294)	1.106 (0.092)	1.055 (0.099)
PFFS Terminations (<i>N</i> =11,334,799)	1.076 (0.113)	1.141 (0.120)
Non-PFFS Terminations (<i>N</i> =11,224,729)	1.021 (0.170)	1.319 (0.237)
No Dual-Eligibles (<i>N</i> =10,067,443)	1.046 (0.091)	1.054 (0.101)
Age-Specific Effects (<i>N</i> =11,441,205)	1.029 (0.098)	1.073 (0.106)
Risk Score Control (<i>N</i> =11,441,205)	1.024 (0.098)	1.067 (0.106)
No Shrinkage (<i>N</i> =11,441,205)	0.969 (0.053)	0.814 (0.055)
No Small Plans (<i>N</i> =11,372,517)	0.842 (0.112)	0.954 (0.135)
Two-Year Mortality (<i>N</i> =11,441,205)	1.170 (0.117)	1.126 (0.141)
Specification	Linear	Median
Demographic Controls	Yes	Yes

Notes: This table reports second-stage coefficient estimates from Equation (10). The dependent variable is one-year mortality in the first ten rows and two-year mortality in the final row. 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. The first row drops counties with no terminations over 2008-2011. The second row drops beneficiaries who switch to a TM plan. The third row drops non-PFFS plans that terminate and the fourth row drops PFFS plans that terminate. The fifth row drops dual-eligible beneficiaries. The sixth row allows observational mortality to depend on beneficiary age, as described in the text. The seventh row adds lagged plan risk score as a control. The eighth row does not use the empirical Bayes procedure to construct observational mortality. The ninth row drops lagged plans with fewer than 50 beneficiaries. The tenth row uses two-year observational mortality. 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), and beneficiary demographics. Standard errors are clustered by county and reported in parentheses.

Table A7: Alternative Plan Characteristics Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: OLS						
Star Rating	0.0008 (0.0003)					
1(Star Rating \geq 4)		-0.0009 (0.0005)				
1(Star Rating \geq 3)			0.0016 (0.0005)			
1(Plan-Type = PPFS)				-0.0054 (0.0004)		
1(Plan-Type = HMO)					-0.0011 (0.0003)	
1(Plan-Type = PPO)						-0.0056 (0.0004)
Panel B: IV						
Star Rating	-0.0011 (0.0012)					
1(Star Rating \geq 4)		-0.0028 (0.0023)				
1(Star Rating \geq 3)			0.0004 (0.0022)			
1(Plan-Type = PPFS)				-0.0169 (0.0067)		
1(Plan-Type = HMO)					0.0044 (0.0041)	
1(Plan-Type = PPO)						0.0035 (0.0036)
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 3. The IV specifications instrument by the interaction of lagged plan characteristics and terminations, controlling for main effects. 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 3.

Table A8: Mortality and Plan Choice Simulations

	Change Among Reassigned (1)	% of Mean Mortality (2)	Unconditional Change (3)	% of Mean Mortality (4)
Assignment to Plans at Random	-0.001 [-0.002, 0.001]	-1.8 [-5.2, 1.7]	-0.001 [-0.002, 0.001]	-1.8 [-5.2, 1.7]
Assignment from Top- to Bottom-Quartile Plans	-0.006 [-0.007, -0.005]	-13.4 [-15.8, -10.9]	-0.0025 [-0.0030, -0.0021]	-5.4 [-6.4, -4.4]
Assignment from Top 5% of Plans	-0.013 [-0.015, -0.010]	-26.6 [-31.5, -21.6]	-0.0006 [-0.0007, -0.0005]	-1.3 [-1.6, -1.1]

Notes: Each row of the table summarizes the simulated change in observational mortality posteriors when MA beneficiaries are reassigned to plans, as described in the text. Given the observational mortality forecast estimate in Column (1) of Panel C of Table 3, these results imply commensurate changes in mortality. 95% confidence intervals reflecting the precision of the forecast coefficient are reported in brackets. All simulations are conducted on the MA sample, excluding plans with fewer than 12 beneficiaries in a given year.

Table A9: Discounted Value of Statistical Life

Annual VSLY (\$1,000s)	Discount Factor	$Pr(\text{death})$ Multiplier	Discounted VSL (\$1,000s)
369	0.95	×1	2,171
		×2	1,305
	0.98	×1	2,646
		×2	1,520
184.5	0.95	×1	1,085
		×2	652
	0.98	×1	1,323
		×2	760

Notes: The table estimates the expected VSL by computing the discounted expected stream of a VSLY. Specifically for individual i , we compute $\sum_{t=\text{age}_i}^{100} \text{VSLY} \cdot \delta^{t-\text{age}_i} \prod_{s=\text{age}_i}^t (1 - P(\text{death at age } s))$. The second panel considers cutting the VSLY in half, assuming all the elderly in our sample are in poor health. $P(\text{death})$ is taken from the SSA actuarial life tables; in the second row of each panel we consider doubling the mortality rate for each year. We consider discount rates of 2% and 5%. For comparison, the OMB uses a social discount rate of 3% for private consumption https://obamawhitehouse.archives.gov/omb/circulars_a004_a-4/.

B Data Appendix

We use the 100% CMS Master Beneficiary Summary Files for 2007–2011 as the basis of our analysis. We apply a series of sample selection criteria throughout. We restrict to Medicare beneficiaries who are 65 years or older and who reside in the 50 United States or the District of Columbia. We drop beneficiaries that are ever observed in a small minority of plans or contracts with more than 50% dual-eligible beneficiaries, which tend to be outliers with high mortality rates. We further drop beneficiaries with incomplete enrollment or location data, beneficiaries with gap years in their enrollment, beneficiaries with contract and plan data missing for every month of a given year, beneficiaries with enrollment data in years after the year of their death, and beneficiaries with multiple years of death records.

Our IV analysis is based on a subsample of beneficiaries enrolled in a MA plan from 2008–2011. We define MA plans as those with types of HMO, non-HMO/POS, Local PPO, Local HMO, PFFS, or Regional PPO. We exclude 800-series plans, special needs plans, and demonstration plans. We define terminations by the CMS Landscape file for Medicare Advantage and Cost Plans.

Star rating data become available from CMS in 2008. We take average star ratings in 2008–2011 and merge these characteristics by plan contracts.

We collect premium data also from the CMS Landscape files. The variable includes Medicare Part C and Part D. The average premium is taken at the state, county, and plan contract level. Premium data from the Landscape files are merged onto our observational mortality estimates first using state, county, plan, and contract. If an observation has a missing premium value after this first merge, then a second merge is performed to the Landscape files using state, county, and contract, where contracts with the lowest plan ID in the Landscape files are used.

We construct Medical Loss Ratios (MLRs) from data provided by CMS. These data are only publicly available online from 2011–2017, so the 2011 data are used, subset to the government market segment. MLR is calculated as $(\text{total claims with permitted adjustments} + \text{total expenses for activities to improve healthcare quality}) / (\text{total premium adjusted for payments to or from the federal and state high risk pools} - \text{total federal and state payments as adjustment to premium})$. MLRs that are negative or greater than 2 are excluded. We merge MLR values to the observational mortality dataset, first by state, county, plan, and contract. If an observation has a missing star rating after this first merge, then a second merge is performed to the Landscape files using state, county, and contract, where contracts with the lowest plan ID in the Landscape files are used. If there were multiple organization names associated with the same plan and contract within a given state and county, the longest organization name was used. Then a manual mapping between the company name in the MLR data and the organization name from the Landscape files was constructed.

C Econometric Appendix

C.1 Empirical Bayes Shrinkage

This appendix describes our empirical Bayes approach to account for noise in our estimates of observational mortality μ_j . We specify a hierarchical linear model in which μ_j is clustered across plans in the same contract, $c(j)$. We further allow the distribution of μ_j to vary across plan size bins.⁴⁶ Throughout we normalize the mean μ_j to be zero within each county. For notational simplicity we here abstract away from the latter two implementation details, imagining a set of mean-zero observational mortality levels $\underline{\mu} = (\mu_1, \dots, \mu_J)$ of a given size in a given county.

Our hierarchical linear model specifies the observational mortality effects as the sum of *iid* contract- and plan-level random effects

$$\mu_j = w_{c(j)} + u_j, \quad (\text{C1})$$

where $E[w_c] = E[u_j] = 0$, $\text{Var}(w_c) = \sigma_w^2$, $\text{Var}(u_j) = \sigma_u^2$. Medicare assigns both “contract IDs” and “plan IDs” within a contract. Throughout, we consider a product a contract-plan-county; observational mortality μ_j is time invariant. We estimate σ_w^2 and σ_u^2 from a vector of estimates $\hat{\underline{\mu}}$, where $\hat{\mu}_j = \mu_j + e_j$ with e_j denoting mean-zero and uncorrelated estimation error with a j -specific variance $\sigma_{e,j}^2$. These estimates are given by (recentered) OLS coefficients, and we estimate σ_w^2 , σ_u^2 , and σ_e^2 by a conventional random effects procedure (Morris, 1983). To minimize small-sample biases, we exclude from this procedure the small minority of contracts with fewer than 100 beneficiary-years. Our estimates of $\text{Var}(w_c) = \sigma_w^2$ and $\text{Var}(u_j) = \sigma_u^2$ yield our overall estimate of the standard deviation of observational mortality, according to Equation (C1).

“Shrunk” empirical Bayes posteriors of observational mortality are given by Equation (C1) and our estimates of $\text{Var}(w_c) = \sigma_w^2$, $\text{Var}(u_j) = \sigma_u^2$, and $\text{Var}(u_j) = \sigma_e^2$. Formulas for these posteriors are derived from the regression of $\underline{\mu}$ on $\hat{\underline{\mu}}$ and give the best linear unbiased prediction of $\underline{\mu}$ from $\hat{\underline{\mu}}$ by standard Gauss-Markov logic. To illustrate this procedure, suppose there are only three plans (A , B , and C) in two contracts, with $c(A) = c(B)$. Then the posterior vector is given by

$$\begin{aligned} \underline{\mu}^* &= \text{Cov}([\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C], [\mu_A, \mu_B, \mu_C]') \text{Var}([\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C]')^{-1} [\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C]' \\ &= V(V + \text{diag}([\sigma_{e,A}^2, \sigma_{e,B}^2, \sigma_{e,C}^2])^{-1} [\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C]'), \end{aligned} \quad (\text{C2})$$

⁴⁶The sample is split into bins with roughly equal-sized numbers of *terminated* enrollees: ≤ 50 , 50-150, 150-350, 350-1,000, 1,000-4,000, and $\geq 4,000$.

where

$$V = \text{Var}(\underline{\mu}) = \begin{bmatrix} \sigma_w^2 + \sigma_u^2 & \sigma_w^2 & 0 \\ \sigma_w^2 & \sigma_w^2 + \sigma_u^2 & 0 \\ 0 & 0 & \sigma_w^2 + \sigma_u^2 \end{bmatrix}. \quad (\text{C3})$$

This shows that the posterior for the third plan C is

$$\mu_C^* = \frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2 + \sigma_u^2 + \sigma_{e,C}^2} \hat{\mu}_C = \frac{\text{Var}(\mu_C)}{\text{Var}(\mu_C) + \text{Var}(e_C)} \hat{\mu}_C, \quad (\text{C4})$$

as in a standard empirical Bayes shrinkage procedure. The formula also shows that the posteriors of the two clustered plans A and B are determined by the relative variances at the contract and plan level. When σ_w is small, μ_A^* and μ_B^* will be similar to the conventional non-clustered shrinkage formula (C4). Otherwise, noisy observational mortality estimates of plans in the same contract are implicitly shrunk towards one another, as well as towards the grand mean of zero.

In practice, the typical estimate of a plan's observational mortality is very precise (i.e. the typically $\text{Var}(e_j)$ is very small), making $\underline{\mu}^*$ close to $\hat{\underline{\mu}}$. This fact is summarized in Appendix Figure A2, which shows the distribution of a “pseudo shrinkage coefficient,” $\hat{\mu}_j/\mu_j^*$ given our estimates of the variance parameters in Equation (C1). The median coefficient is one, with nearly all coefficients found to be larger than 0.75.

C.2 Plan-Level Balance Assumption

This appendix shows how our balance assumption (2) can be written in terms of an infeasible plan-level difference-in-differences regression. We first note that by the Frisch-Waugh-Lovell Theorem, $\text{Cov}(\tilde{Z}_{it}, \varepsilon_{it}) = 0$ if and only if $\phi_Z = 0$ in the beneficiary-level regression of

$$\varepsilon_{it} = \phi_Z Z_{it} + X_{it}' \phi_X + e_{it}. \quad (\text{C5})$$

We next note that since the regressors of this equation, $Z_{it} = \mu_{i,t-1} T_{i,t-1} = \sum_j \mu_j T_{j,t-1} D_{ij,t-1}$ and $X_{it} = \sum_j X_{j,t-1} D_{ij,t-1}$, only vary at the level of lagged enrollment group indicators $D_{ij,t-1}$, the coefficients of this equation are equivalently obtained by a $\text{Pr}(D_{ij,t-1} = 1)$ -weighted plan-level regression of

$$\bar{\varepsilon}_{jt} = \phi_Z \mu_j T_{j,t-1} + X_{j,t-1}' \phi_X + e_{jt}, \quad (\text{C6})$$

where $\bar{\varepsilon}_{jt} = E[\varepsilon_{it} \mid D_{ij,t-1} = 1]$. Thus $\text{Cov}(\tilde{Z}_i, \varepsilon_i) = 0$ if and only if $\phi_Z = 0$ in this regression, which coincides with Equation (7).

C.3 Alternative Monotonicity Condition for Interpreting Forecast IVs

This appendix discusses an alternative to the fallback condition, similar to the monotonicity condition of Imbens and Angrist (1994). The condition permits an alternative interpretation of the forecast IV regression coefficient as a proper weighted regression of causal effects β_j on observational mortality μ_j . Again letting \tilde{Z}_{it} be the residualized instrument, and treating the (β_j, μ_j) as fixed, note that under Assumptions 1 and 2 (only) the IV coefficient can be written

$$\frac{Cov(\tilde{Z}_{it}, Y_{it})}{Cov(\tilde{Z}_{it}, \mu_{it})} = \frac{\sum_j \beta_j Cov(\tilde{Z}_{it}, D_{ijt}) + Cov(\tilde{Z}_{it}, \varepsilon_{it})}{\sum_j \mu_j Cov(\tilde{Z}_{it}, D_{ijt})} = \frac{\sum_j \omega_j \beta_j \mu_j}{\sum_j \omega_j \mu_j^2}, \quad (C7)$$

where we write $\omega_j = Cov(\tilde{Z}_{it}, D_{ijt})/\mu_j$ for j such that $\mu_j \neq 0$ and $\omega_j = 0$ otherwise. Further noting that we can without-loss normalize $\sum_j \omega_j \mu_j = 0$, this shows that the IV coefficient can be written as a weighted regression of β_j on μ_j with weights ω_j . It is furthermore a proper (convex) weighting scheme under the following condition:

Assumption 4. (*Monotonicity*): $\omega_j \geq 0$ for all j .

Here monotonicity holds when the covariance between the residualized instrument \tilde{Z}_{it} and plan enrollment D_{ijt} is always of the same sign as the plan's observational mortality μ_j .

Consider our “median” specification where Z_{it} is the interaction of an indicator for lagged plan observational mortality being above-median in its market and a lagged termination indicator, where we control for above-median and termination main effects. Further suppose the ω_j -weighted distribution of μ_j is symmetric, so the mean and median coincide. Then Assumption 4 holds when above-median observational mortality plans (with $\mu_j > 0$) always attract a weakly larger share of beneficiaries from above- vs. below-median plans following a termination, relative to below-median observational mortality plans. This condition is strong; it rules out, for example, plans that are just above-median differentially attracting beneficiaries from below-median plans because they share some non- μ_j characteristic. We further note that Assumption 4 is testable, since each ω_j is a function of observable data.

In general, the monotonicity condition and our baseline fallback condition are non-nested assumptions: we may have $\omega_j \geq 0$ for all j while $Cov(\tilde{Z}_{it}, \eta_{it}) \neq 0$ and vice-versa. A virtue of the fallback condition is that it can be microfounded with classic discrete choice models, per Appendix C.4. The forecast IV furthermore identifies the coefficient from an unweighted regression of β_j on μ_j when Assumption 3 holds. In contrast, the forecast IV under Assumption 4 identifies a weighted regression with weights that are functions of the instrument and which may not be policy-relevant.

C.4 Discrete Choice Microfoundation of the Fallback Condition

This appendix develops a simple discrete choice model which satisfies Assumption 3. We then discuss extensions to this approach of microfounding the fallback condition. The simplest version of the model assumes that plan terminations are as good-as-randomly assigned in each period $t - 1$, that consumers in non-terminated plans are fully inertial in the next period t , and that consumers in terminated plans make an unrestricted choice in period t to maximize a latent utility of

$$U_{ijt} = \alpha'_{it} W_j + \xi_j + u_{ijt}. \quad (\text{C8})$$

Here α_{it} captures potentially heterogeneous preferences over observed plan characteristics W_j and ξ_j denotes a plan unobservable. We follow the standard convention in such models (e.g. Berry et al. (1995)) of treating residual utility u_{ijt} as an idiosyncratic *iid* shock, though we do not require any parametric assumptions on the distribution of α_{it} or u_{ijt} .⁴⁷ We complement this model for plan choice with our baseline outcome model $Y_{ijt} = \beta_j + \varepsilon_{it}$, allowing the unobserved ε_{it} to be arbitrarily correlated with both α_{it} and u_{ijt} . Such correlation with the choice process (C8) will tend to generate endogeneity in plan choice and bias in the observational mortality measure μ_j .

A sufficient condition for Assumption 3 is that the beneficiaries previously enrolled in terminated plans select new plans similarly to those previously enrolled in non-terminated plans, given the regression controls. When consumers in non-terminated plans are fully inertial, this condition means that the fallback choice probability of consumers in terminated plans does not systematically depend on the identity of their previous plan, making their choices representative of the initial choices of non-terminated consumers. Formally, we consider the sufficient condition of

$$\Pi'_{k \rightarrow j}(X_{k,t-1}) \equiv Pr(D_{ijt} = 1 \mid D_{ik,t-1} = 1, X_{k,t-1}) = Pr(D_{ijt} = 1 \mid X_{k,t-1}) \equiv \pi'_j(X_{k,t-1}), \quad (\text{C9})$$

where $X_{j,t-1}$ are lagged plan characteristics which the IV regression flexibly controls for (via the transformation of $X_{it} = \sum_j X_{j,t-1} D_{ij,t-1}$). Equation (C9) holds when fallback choice probabilities $\Pi'_{k \rightarrow j}(X_{k,t-1})$ do not depend on the identity of the lagged plan k , and are thus equal to the unconditional choice probabilities $\pi'_j(X_{k,t-1})$, given the observables in $X_{k,t-1}$.

To see that Equation (C9) is enough to satisfy the fallback condition, consider a version of the IV regression which conditions on the lagged plan characteristics in X_{it} . As in the full-sample case, we can without loss normalize the market share weighted average observational mortality to zero; here we use the weights in the $t - 1$ period, $\pi_j^{t-1}(x)$ for the conditioning value x . That is, $\sum_j: X_{j,t-1}=x \pi_j^{t-1}(x) \mu_j = 0$. The forecast residual η_j is furthermore defined to be mean-zero and uncorrelated with μ_j (when also weighted by $\pi_j^{t-1}(x)$) in this subsample. Then, when beneficiaries

⁴⁷In particular, and in contrast to the standard approach, we allow α_{it} and u_{ijt} to be correlated.

in non-terminated plans are fully inertial, the instrument and fallback residual are conditionally uncorrelated among beneficiaries in non-terminated plans:

$$\begin{aligned} \text{Cov}(Z_{it}, \eta_{it} \mid T_{i,t-1} = 0, X_{it} = x) &= \sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j \eta_j - \left(\sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j \right) \left(\sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \eta_j \right) \\ &= 0, \end{aligned} \quad (\text{C10})$$

where we use the fact that terminations are randomly assigned (at least conditional on X_{it}). Furthermore, when Equation (C9) holds we have the same relationship for terminated plans:

$$\begin{aligned} \text{Cov}(Z_{it}, \eta_{it} \mid T_{i,t-1} = 1, X_{it} = x) &= \sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j \left(\sum_k \eta_k \Pi_{j \rightarrow k}^t(x) \right) \\ &\quad - \left(\sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j \right) \left(\sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \left(\sum_k \eta_k \Pi_{j \rightarrow k}^t(x) \right) \right) \\ &= \left(\sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j \right) \left(\sum_k \eta_k \pi_k^t(x) \right) = 0, \end{aligned} \quad (\text{C11})$$

where we twice use the normalization of $\sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j = 0$. This extends to feasible IV regressions which control flexibly for the lagged plan characteristics that make Equation (C9) hold.⁴⁸

It remains to be shown that the discrete choice model (C8) admits a set of $X_{j,t-1}$ satisfying Equation (C9). We show this by building up to Equation (C8) in a series of special cases. First suppose $\alpha_{it} = 0$, such that beneficiaries in terminated plans resort to new plans in proportion to their market shares. The fallback condition is clearly satisfied in this case without any conditioning. Next, suppose α_{it} varies across beneficiaries but is *iid* over time like u_{ijt} . Then beneficiaries differ unobservably in their fallback choice probabilities, but this variation is still independent of lagged plan choice so the fallback condition again holds unconditionally. Finally, consider the case where α_{it} both varies across beneficiaries and is persistent across time. Then it is apparent that Equation (C9) holds provided the lagged plan characteristics that consumers exhibit heterogeneous and persistent preferences over, in W_k , are included in $X_{k,t-1}$.

We note two extensions of this simple microfoundation for the fallback condition. First, the basic logic of controlling for lagged plan characteristics governing persistent unobserved variation in fallback choice probabilities appears very general. Consider, for example, a version of Equation

⁴⁸It is worth emphasizing that we do not require the plan shares of non-terminated beneficiaries in time $t-1$, $\pi_j^{t-1}(\cdot)$, to equal the plan shares of terminated beneficiaries in time t , $\pi_j^t(\cdot)$. This allows for terminated plans to be unavailable in the following period. It also allows both the observable and unobservable plan characteristics, W_j and ξ_j , to be time-varying. We omit this extension to simplify notation and to ease exposition.

(C8) which allows for random coefficients on the unobserved plan characteristics:

$$U_{ijt} = \alpha_{it}' W_j + v_{it} \xi_j + u_{ijt}, \quad (\text{C12})$$

with v_{it} potentially correlated with ε_{it} , along with α_{it} and u_{ijt} . If ξ_j were observed, one could control for it in X_{it} to account for any persistent unobserved heterogeneity due to v_{it} that may cause the fallback condition to fail. With ξ_j unobserved, it may still be possible to implicitly condition on it by conditioning on the market share variation that is sufficient to identify these random coefficients. If v_{it} is almost-surely positive, for example, the market shares implied by equation (C12) are typically invertible in ξ_j , yielding such identification. Berry et al. (2013) provide weaker conditions for such invertibility in a general class of utility specifications nesting (C12).

Second, we note that the microfoundation can be extended to allow for partial inertia among non-terminated beneficiaries. Consider, for example, the model of consumer inattention in Ho et al. (2017) in which beneficiaries in non-terminated plans re-optimize their plan choice when an attention shock v_{it} exceeds some threshold \tilde{v}_{it} . In Ho et al. (2017) the threshold is a function of beneficiary observables and fixed effects which could be included in X_{it} . When the shocks v_{it} are idiosyncratic, in the sense of being uncorrelated with the residual determinants of plan choice, then attentive consumers will re-optimize in the same way as consumers in terminated plans while inattentive consumers will remain in the choices representative of consumers in non-terminated plans. In this scenario we will still have Z_{it} and η_{it} conditionally uncorrelated among terminated plans, i.e. $\text{Cov}(Z_{it}, \eta_{it} \mid T_{i,t-1} = 1, X_{it} = x) = 0$, since post-terminated behavior does not change. Furthermore, with $E[Z_{it} \mid T_{i,t-1} = 0, X_{it} = x] = \sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j = 0$, we still have

$$\begin{aligned} \text{Cov}(Z_{it}, \eta_{it} \mid T_{i,t-1} = 0, X_{it} = x) &= E[Z_{it} \eta_{it} \mid T_{i,t-1} = 0, X_{it} = x] \\ &= E[Z_{it} \eta_{it} \mid T_{i,t-1} = 0, v_{it} > \tilde{v}_{it}, X_{it} = x] \Pr(v_{it} > \tilde{v}_{it} \mid T_{i,t-1} = 0, X_{it} = x) \\ &\quad + E[Z_{it} \eta_{it} \mid T_{i,t-1} = 0, v_{it} \leq \tilde{v}_{it}, X_{it} = x] \Pr(v_{it} \leq \tilde{v}_{it} \mid T_{i,t-1} = 0, X_{it} = x) \\ &= E[Z_{it} \eta_{it} \mid T_{i,t-1} = 1, X_{it} = x] \Pr(v_{it} > \tilde{v}_{it} \mid T_{i,t-1} = 0, X_{it} = x) \\ &\quad + E[Z_{it} \eta_{it} \mid T_{i,t-1} = 0, v_{it} \leq \tilde{v}_{it}, X_{it} = x] \Pr(v_{it} \leq \tilde{v}_{it} \mid T_{i,t-1} = 0, X_{it} = x) \\ &= \left(\sum_{j: X_{j,t-1}=x} \pi_j^{t-1}(x) \mu_j \eta_j \right) \Pr(v_{it} \leq \tilde{v}_{it} \mid T_{i,t-1} = 0, X_{it} = x) = 0, \end{aligned} \quad (\text{C13})$$

using the fact that $E[Z_{it} \eta_{it} \mid T_{i,t-1} = 0, v_{it} > \tilde{v}_{it}, X_{it} = x] = E[Z_{it} \eta_{it} \mid T_{i,t-1} = 1, X_{it} = x] = 0$ and the normalization of $\sum_{j: X_{j,t-1}=x} \tilde{\pi}_j(x) \mu_j \eta_j = 0$ as before. Intuitively, the conditionally idiosyncratic attentiveness shocks randomly move some individuals from the $T_{i,t-1} = 0$ group to the $T_{i,t-1} = 1$ group, leaving the equality of covariances across these groups unchanged. We can further relax the assumption of idiosyncratic v_{it} by including plan characteristics affecting the distribution of

shocks (such as out-of-pocket costs, as in Ho et al. (2017)) in X_{it} . The threat to the fallback condition in the partial inertia case is some unobserved attentiveness shock which causes non-terminated individuals to re-optimize in a systematically different way than the set of individuals who are forced to re-optimize by a plan termination.

C.5 Forecast IVs with Estimated Observational Mortality

This appendix discusses feasible forecast IV regressions when observational mortality μ_j is not known and must be estimated. To understand the problem with a naïve regression of mortality Y_{it} on unadjusted regression estimates $\hat{\mu}_{it} = \sum_j \hat{\mu}_j D_{ijt}$, instrumented by some \tilde{Z}_{it} satisfying our first-stage, balance, and fallback conditions, write $\hat{\mu}_j = \mu_j + e_j$ for idiosyncratic estimation error satisfying $E[e_j] = 0$, $\text{Var}(e_j) = \sigma_{e,j}^2$, and $\text{Cov}(e_j, \mu_j) = \text{Cov}(e_j, \eta_j) = 0$. Suppose for simplicity that $E[\beta_j] = E[\mu_j] = 0$ and $\text{Var}(\mu_j) = \sigma_\mu^2$. Then, under Assumptions 1–3 this IV regression identifies an attenuated forecast coefficient, given by

$$\tilde{\lambda} = \frac{\frac{1}{J} \sum_j \text{Cov}(\beta_j, \hat{\mu}_j)}{\frac{1}{J} \sum_j \text{Var}(\hat{\mu}_j)} = \frac{\frac{1}{J} \sum_j \text{Cov}(\lambda \mu_j + \eta_j, \mu_j + e_j)}{\frac{1}{J} \sum_j \text{Var}(\mu_j + e_j)} = \lambda \frac{\sigma_\mu^2}{\sigma_\mu^2 + \frac{1}{J} \sum_j \sigma_{e_j}^2}. \quad (\text{C14})$$

Intuitively, under Assumptions 1–3 the IV procedure recovers the regression of β_j on $\hat{\mu}_j$, which suffers from classic attenuation bias due to the measurement error in $\hat{\mu}_j$.

As with classic attenuation bias, this attenuation bias can be addressed by replacing $\hat{\mu}_j$ with a posterior mean μ_j^* like those considered in Appendix C.1. To see this simply suppose $\mu_j^* = \frac{\sigma_\mu^2}{\sigma_{e_j}^2 + \sigma_\mu^2} \hat{\mu}_j$ as in a conventional shrinkage procedure. Then, again under Assumptions 1–3, and IV regression of Y_{it} on the corresponding μ_{it}^* instrumented by a valid Z_{it} identifies

$$\begin{aligned} \frac{\frac{1}{J} \sum_j \text{Cov}(\beta_j, \mu_j^*)}{\frac{1}{J} \sum_j \text{Var}(\mu_j^*)} &= \frac{\frac{1}{J} \sum_j \text{Cov}\left(\lambda \mu_j + \eta_j, \frac{\sigma_\mu^2}{\sigma_{e_j}^2 + \sigma_\mu^2} (\mu_j + e_j)\right)}{\frac{1}{J} \sum_j \text{Var}\left(\frac{\sigma_\mu^2}{\sigma_{e_j}^2 + \sigma_\mu^2} (\mu_j + e_j)\right)} \\ &= \lambda \frac{\frac{1}{J} \sum_j \frac{\sigma_\mu^4}{\sigma_{e_j}^2 + \sigma_\mu^2}}{\frac{1}{J} \sum_j \frac{\sigma_\mu^4}{(\sigma_{e_j}^2 + \sigma_\mu^2)^2} \text{Var}(\mu_j + e_j)} = \lambda. \end{aligned} \quad (\text{C15})$$

Intuitively, the shrinkage adjustment in μ_j^* undoes the attenuation bias due to e_j , as it would if we were to estimate directly the regression of β_j on μ_j^* .

C.6 Treatment Effect Heterogeneity

This appendix shows how our IV framework accommodates unobserved treatment effect heterogeneity. The general model allows for heterogeneous treatment effects by writing

$$Y_{ijt} = \beta_j + X'_{it}\gamma + \varepsilon_{it} + \zeta_{ijt}, \quad (\text{C16})$$

where β_j is normalized such that $\frac{1}{j}\sum_j \beta_j = 0$, ε_{it} is normalized such that $E[X_{it}(Y_{ijt} - \beta_j - \varepsilon_{it})] = 0$ for each j , and ζ_{ijt} is a residual from this projection. In our baseline model $\zeta_{ijt} = 0$ and ε_{it} captures the relevant unobserved health of beneficiary i in year t . Otherwise, ζ_{ijt} captures the relative unobserved appropriateness of beneficiary i for plan j in year t : when $\zeta_{ijt} < 0$ then (i, t) derives a better-than-average reduction in mortality from selecting plan j relative to the typical beneficiary-year with similar observables X_{it} .

We can continue to project β_j on μ_j in this more general model to define a forecast coefficient λ and forecast residual η_j . This projection yields a second-stage equation of

$$\begin{aligned} Y_{it} &= \sum_j Y_{ijt} D_{ijt} = \sum_j \beta_j D_{ijt} + X'_{it}\gamma + \varepsilon_{it} + \zeta_{it} \\ &= \lambda \mu_{it} + X'_{it}\gamma + \varepsilon_{it} + \eta_{it} + \zeta_{it}, \end{aligned} \quad (\text{C17})$$

where $\mu_{it} = \sum_j \mu_j D_{ijt}$ and $\eta_{it} = \sum_j \eta_j D_{ijt}$ as before, and now $\zeta_{it} = \sum_j \zeta_{ijt} D_{ijt}$. This latter term captures the selected-on-gains of beneficiary i in year t : here $\zeta_{it} < 0$ implies that (i, t) has selected a plan which is relatively more appropriate for her than the typical beneficiary-year.

The first-stage, balance, and fallback conditions continue to be necessary for estimation of Equation (C17) with an instrument Z_{it} to identify λ . With $\zeta_{it} \neq 0$ we also require a fourth condition, that $Cov(\tilde{Z}_{it}, \zeta_{it}) = 0$. This condition says that the conditional variation in the instrument does not predict variation in the relative extent of selection-on-gains captured by ζ_{it} . As with Assumptions 2 and 3, it can be interpreted via an infeasible plan-level difference-in-differences regression, of

$$\bar{\zeta}_{jt} = \phi_Z \mu_j T_{j,t-1} + X'_{j,t-1} \phi_X + e_{jt}, \quad (\text{C18})$$

where $\bar{\zeta}_{jt} = E[\zeta_{it} | D_{ij,t-1} = 1]$ captures the average selection-on-gains among beneficiaries previously enrolled in plan j at time $t - 1$. For $\phi_Z = 0$ in this expression, satisfying $Cov(\tilde{Z}_{it}, \zeta_{it}) = 0$, the conditional relationship between observational mortality and average selection-on-gains in terminated and non-terminated plans should be similar.

This new condition mirrors the logic of the fallback condition, as it is satisfied when beneficiaries choose similarly following a plan termination to new consumers in a given market. Formally, note that the microfoundation in Appendix C.4 easily generalizes to allow for treatment effect het-

erogeneity by replacing ε_{it} with $\varepsilon_{it} + \zeta_{ijt}$. That is, the fallback condition holds given no persistent unobserved heterogeneity in plan choice that is correlated with either beneficiary health or the beneficiary’s appropriateness for certain plans. As with the fallback condition, this new restriction can also be empirically investigated by using beneficiary observables to proxy for unobservables that might drive treatment effect heterogeneity.

C.7 Forecast Coefficient Sensitivity to Omitted Variables Bias

This appendix discusses how we gauge sensitivity of our baseline forecast coefficient estimates to omitted variables bias, using the frameworks of Altonji et al. (2005) and Oster (2019). We consider bias in the reduced-form regression due to the failure of both Assumptions 2 and 3, and compute bounds on the IV coefficient under moderate deviations from these identifying restrictions.

We follow Oster (2019) in considering “short” and “long” regressions, of the form

$$Y_{it} = \varphi_Z Z_{it} + X'_{1it} \varphi_{X1} + v_{it} \quad (\text{C19})$$

$$Y_{it} = \tilde{\varphi}_Z Z_{it} + X'_{1it} \tilde{\varphi}_{X1} + X'_{1it} \tilde{\varphi}_{X2} + \tilde{v}_{it}, \quad (\text{C20})$$

where Y_{it} is beneficiary mortality, Z_{it} is our baseline (linear) instrument, X_{1it} is a vector of maintained controls (e.g. year-by-county fixed effects), and X_{2it} is a vector of auxiliary controls (e.g. beneficiary demographics) thought to be uncorrelated with Z_{it} given X_{1it} under Assumptions 2 and 3. Thus, under these conditions we expect $\varphi_Z = \tilde{\varphi}_Z$.

We use the difference in estimates of φ_Z and $\tilde{\varphi}_Z$ and the sample R^2 from each regression to bound the true reduced-form coefficient. Formally, we use the approximation in Oster (2019):

$$\varphi_Z^* \approx \tilde{\varphi}_Z + \delta(\tilde{\varphi}_Z - \varphi_Z) \frac{R_{max} - \tilde{R}}{\tilde{R} - R}, \quad (\text{C21})$$

where φ_Z^* is the true instrument coefficient once all confounding variables are controlled for, R and \tilde{R} are the R^2 from Equations (C19) and (C20) respectively, and R_{max} is the R^2 from the regression that controls for all confounding variables. The constant δ is the ratio of coefficients from regressing the instrument on the included control component $X'_{1it} \tilde{\varphi}_{X1} + X'_{2it} \tilde{\varphi}_{X2}$, and from regressing the instrument on the full component of included and omitted variables.⁴⁹ Oster (2019) interprets this as how “important” the observables are as a source of omitted variables bias relative to unobservables.

We use Equation (C21) to bound the true forecast coefficient λ in the following way. First, we estimate the “short” reduced-form and first-stage coefficients φ_Z by including in X_{1it} all of

⁴⁹We note that δ here is different than mean utility in the discrete choice microfoundation. We are adopting the notation common to each literature.

our baseline controls (i.e. those in Columns 1 and 3 of Table 3). We further obtain the R^2 from this regression, denoted R . Next, we add to these regressions a X_{2it} which captures all observable proxies of beneficiary-level and plan-level confounders which may violate Assumptions 2 and 3. Specifically, we include all of the demographic controls (i.e. those in Columns 2 and 4 of Table 3) as well as all of the enrolled plan characteristics in Table 4. We residualize plan characteristics on enrolled plan observational mortality to ensure they proxy for the residual η_j . We obtain from this specification the "long" reduced-form and first-stage coefficients $\tilde{\varphi}_Z$ and corresponding R^2 , denoted \tilde{R} . We then compute the reduced-form coefficients from Equation C21 with different values of $|\delta|$ and R_{max} . Finally, we scale the reduced-form estimates by our baseline first-stage estimate to compute bounds on the forecast coefficient.⁵⁰

We first follow Altonji et al. (2005) and Oster (2019) in using $|\delta| = 1$. This corresponds to an assumption that the unobserved confounders are equally "important" as the observables and bias the reduced-form and first-stage coefficients in either the same (for $\delta = 1$) or opposite (for $\delta = -1$) directions. We then consider a scenario in which the unobservables are three times more important, i.e. $|\delta| = 3$. In each scenario, we vary R_{max} from $1.3\tilde{R}$ and $5\tilde{R}$. The former is what Oster (2019) recommends for computing bounds, based on a calibration to experimental data. The latter value of R_{max} corresponds to a much more aggressive robustness check, in which we allow the relevant unobservables to explain up to five times the variation in the outcome than the instrument and observable controls, or all remaining variation in the "long" regressions.

The results of this exercise are given in Table A5. We obtain relatively tight bounds on the forecast coefficient when assuming the observable and unobservable confounders are equally "important" ($\delta = 1$), even for relatively large values of R_{max} . When R_{max} equals the preferred value of Oster (2019) and all observables are included the bounds are $[1.00, 1.03]$, while for the more aggressive $R_{max} = 5\tilde{R}$ the bounds are $[0.81, 1.21]$. As expected, increasing $|\delta|$ increases the bounds. Yet even under the most conservative specification in the table we obtain a forecast coefficient bound that easily excludes zero, of $[0.40, 1.62]$. This implies that unobservable confounders would have to be massively more important, in both the sense of R_{max} and $|\delta|$, to explain our finding of a high forecast coefficient.

⁵⁰The first stage coefficient is unchanged with additional controls, as shown in Table 3.

D Policy Simulations

This appendix quantifies the potential gains from aligning consumer choice with plan mortality effects in a series of partial-equilibrium simulations. We first compare average one-year mortality among MA beneficiaries under their status quo choices to a benchmark of random assignment to plans within a market. Random assignment is used for low-income subsidy enrollees in Medicare Part D (Decarolis, 2015) and in some state Medicaid programs (including California, New York, and South Carolina). If MA consumers are more likely to choose plans with better mortality effects, as we found in Section 5.3, then random assignment could increase mortality relative to the status quo. This first simulation quantifies the change in β_j , and thus the change in average mortality under active choice.⁵¹ In practice we compute this value by first obtaining a forecast coefficient that implicitly regresses β_j on log plan market shares by our baseline IV approach. We then multiply this coefficient by the change in market shares obtained under random assignment.

The results are reported in the first row of Appendix Table A8. While we find in Section 5.3 that consumers are only modestly attentive to plan mortality effects, plans are sufficiently differentiated that this modest attentiveness produces large benefits in our simulation. We find that redirecting consumers to plans at random decreases average mortality by an average of -0.1 percentage points, or 1.8% of the average one-year mortality rate in the sample. In other words, existing choices are no better than random with respect to plan mortality effects.

We next consider the scope for improving on existing choices by leveraging observational predictions of plan quality. These simulations proceed in three steps. First, we simulate draws of observational mortality μ_j and posteriors μ_j^* given the variance parameters and distribution of estimation error that we estimate by the empirical Bayes procedure in Section 2, and a normality assumption on the underlying distribution of μ_j . Next, within each county, we simulate a policy of reassigning beneficiaries in plans with an observational mortality posterior $\tilde{\mu}_j^*$ (e.g. the maximum) to plans with a better prediction of $\bar{\mu}_j^*$ (e.g. the average observational mortality). Our estimate of the average mortality improvement from such reassignment is then the average $\hat{\lambda}(\bar{\mu}_j - \tilde{\mu}_j)$, where $\hat{\lambda}$ is an estimate of the forecast coefficient and $(\bar{\mu}_j, \tilde{\mu}_j)$ denotes the actual observational mortality of plans with posteriors of $(\bar{\mu}_j^*, \tilde{\mu}_j^*)$. With $\hat{\lambda} = 1$, this simulation effectively predicts the potential mortality effect by the average change in observational mortality.⁵²

In the second row of Table A8 we instead reassign beneficiaries from the 75th percentile plan of observational mortality posteriors to those in the 25th percentile plan. Given a forecast coef-

⁵¹All of the exercises in this section are partial equilibrium in that we assume plans do not have capacity constraints, do not strategically enter or exit, and do not change plan characteristics affecting mortality.

⁵²This exercise may understate the gains from active choice to the extent status quo choices reflect positive selection on unobserved treatment effect heterogeneity (Hull, 2020). As noted in Section 4.3 the role of such Roy selection appears small in this setting.

ficient near one, the reassignment would reduce mortality among the affected consumers by 0.6 percentage points, or 13.4% of the average mortality rate.

Finally, the third row of Table A8 presents a simulation in the spirit of Chetty et al. (2014) (who simulate the effect of removing the observably lowest quality teachers on student test scores) and Abaluck and Gruber (2016b) (who simulate the effect of removing the financially worst health insurance plans on beneficiary costs). We consider the impact of removing plans with the worst observational mortality by randomly reassigning beneficiaries in the observably worst plans to other plans at random (top 5% of beneficiaries by observational mortality). This reassignment rule reduces observational mortality posteriors by 0.06 percentage points (1.3 percentage points for affected consumers), or 1.3% of the sample mean. With more than 20 million MA enrollees each year, even this small change in mortality would have a large impact, averting around 12,000 elderly deaths each year given a forecast coefficient of one.⁵³

The reduction in mortality is quite valuable. Table A9 computes the value of a statistical life for marginal enrollees in our sample under a variety of different assumptions. If we assume a plan extends life by exactly one year, we can apply a VSLY of \$369,000, which would imply that a 1.3% reduction in mortality is worth at least \$4,800 for impacted beneficiaries. This is likely too conservative. An alternative is to compute the number of life-years remaining for beneficiaries given assumptions about their age and sickness. To do so, we compute: $VSL = \sum_{t=age_i}^{100} VSLY \cdot \delta^{t-age_i} \prod_{s=age_i}^t (1 - P(\text{death at age } s))$. In our baseline assessments, mortality rates are taken from Social Security Administration actuarial life tables. The second panel of Table A9 considers cutting the VSLY in half, assuming all the elderly in our sample are in poor health and remaining years of life are worth less. Within each panel, we also consider doubling mortality rates, assuming marginal consumers may be sicker than average. Finally, we consider discount rates of 2% and 5%. The Office of Management and Budget (OMB) uses a social discount rate of 3% for private consumption.⁵⁴ The analysis implies VSL ranges from \$0.76-\$2.17 million, which would imply benefits of \$10,000 or more per beneficiary reassigned from the worst 5% of plans.

While suggestive of potentially large mortality reductions, these partial-equilibrium simulations should be interpreted with care. Any policy that reassigns beneficiaries to plans is likely to impact consumer well-being through many channels other than the mortality effects we consider. On one hand, plans that reduce mortality likely also produce better morbidity outcomes such that our estimates understate the health benefits. On the other hand, consumers may be made worse off by having to switch providers, though we find no evidence that terminations directly raise mortal-

⁵³This number is not additive across years, due to competing risks. In other words, some of the 12,000 deaths that are averted in any given year will still die the following year.

⁵⁴See OMB Circular A-4 for guidance on discount rates: https://obamawhitehouse.archives.gov/omb/circulars_a004_a-4/.

ity (through a channel other than plan choice) in any specification.⁵⁵ Our simulations also do not account for the possibility that plans with lower mortality may differ systematically on financial dimensions that consumers value. Our findings in Section 5.1 suggest that plans with lower observational mortality also have lower premiums, although the general equilibrium consequences we have not modelled are likely especially important for premiums. In a sample of MA plans from 2015 to 2017, the average standard deviation of total costs (premiums + out of pocket costs) in MA is around \$1,000 (Gruber et al., 2020).⁵⁶ With a \$1 million VSL, the health benefits of the reassignments in row 3 of Table A8 would be \$13,000 per reassigned beneficiary-year, likely dwarfing any effects via switching costs or financial plan characteristics. However, the calculation is not meant to be a complete welfare accounting of the impact of the policy change. Such a calculation would consider all important attributes, including financial characteristics. Because we do not estimate primitive preferences over such characteristics, this is outside the scope of the paper.⁵⁷

⁵⁵Sabety (2020) and Staiger (2020) argue that switching providers could be harmful, as relationships with primary care physicians improve patient health. When we estimate Equation (10) with a single direct termination effect, we nevertheless find precise zeros estimate of the direct impact of terminations on mortality. For example, when we estimate the specification in column 1 of Table 3 with only a direct effect of terminations (rather than interactions by year and county), our estimate is a 0.2 percentage point reduction in mortality with a standard error of 0.1 percentage points. This rules out even a 0.1 percentage point increase in mortality, compared to the 1.3 percentage point reduction we simulate for eliminating the 5% of worst plans. Atherly et al. (2020) estimate switching costs in Medicare Advantage of \$2,800, which still would not fully offset these benefits, although other analyses suggest that most inertia in health plan choice results from inattention, so utility-relevant switching costs might be as much as 85% lower (Heiss et al., 2016; Abaluck and Adams, 2017; Drake et al., 2020).

⁵⁶We thank Ben Handel and Sam Kina for providing this information.

⁵⁷We also do not consider insurer responses. For example, gaming of newly incentivized metrics should be expected.