Mortality Effects and Choice Across Private Health Insurance Plans

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

July 2020

Abstract

Competition in health insurance markets may fail to improve health outcomes if consumers are not willing to pay for 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. Moving beneficiaries out of the bottom 5% of plans could save tens of thousands of elderly lives each year.

*Abaluck: jason.abaluck@yale.edu, Yale University School of Management and NBER; Caceres: mauricio.caceres.bravo@gmail.com, Brown University; Hull: hull@uchicago.edu, 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, Amit Gandhi, Craig Garthwaite, Marty Gaynor, Jon Gruber, Kate Ho, Tim Layton, Victoria Marone, Tom McGuire, Chris Ody, Molly Schnell, Fiona Scott Morton, Mark Shepard, Doug Staiger, Jeroen Swinkels, Bob Town, and numerous seminar participants for helpful comments. We thank Emily Crawford, Eilidh Geddes, Elise Parrish, Aaron Pollack, and Amy Tingle for excellent research assistance. 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, policymakers 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.

We estimate 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 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." If causal, the estimated variation in observational mortality would suggest that a one standard deviation higher quality plan decreases beneficiary mortality by 0.9 percentage points—a 19% reduction in mortality from a baseline rate of 4.7%, comparable to the sizable variation in mortality effects across hospitals (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 observational mortality measure may reflect unobserved sorting as well as causal plan health effects. To validate the measure, we 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 thus expect cohort mortality to decline (rise) when high- (low-) mortality plans exogenously exit the market, relative to mortality in similar plans that do not terminate. The magnitude of the relationship should furthermore reveal the relationship between observational estimates and causal plan effects. All else equal, cohort 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. 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 exclusion restriction 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, IV estimation of the forecast coefficient requires 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 highly reliable 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. However it shows given our assumptions that observational mortality accurately predicts causal mortality effects on average.

We then extend our approach to answer a series of policy-relevant questions. We first use 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 further show how our IV framework can be used to estimate the implicit willingness to pay for plan quality (WTP). 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.

Finally, we simulate how consumer health might change if MA enrollment decisions were more aligned with plan quality. While consumers are only weakly sensitive to plan quality, status quo choices are better than random. Redirecting consumers from the observably worst plans in a market may also dramatically improve their health. We find that randomly reassigning those in plans with the worst 5% observational mortality rates could avert around 10,000 elderly deaths each year. At conventional VSL estimates, such an effect would have a dollar-equivalent mortality benefit of $10,000 per reassigned enrollee.

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. (2019), 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, 2016a, 2011; 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, 2019; 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., 2017; 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).

We organize the remainder of the paper as follows. In Section 2, we describe the institutional setting and data, document large variation in observational mortality across MA plans, and motivate our quasi-experimental validation approach. In Section 3, we develop our econometric 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. In Section 6, we discuss the scope for quality-based enrollment policy. We conclude in Section 7.

[1] McGuire 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. (2015) 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. (2019), for example, study random assignment of low-income beneficiaries to alternative Medicaid Managed Care plans, finding large spending effects but lacking sufficient power to detect mortality differences.

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

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 opted to receive coverage through a set of diverse private managed care plans (34% in 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 2010, for example, 33 MA plans operated in the average county (KFF, 2009).

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 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

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

[4] The 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 (see Cabral et al. (2014); Duggan et al. (2015)).

revenue from different groups of beneficiaries, and policies affecting federal reimbursement when deciding what plans to offer. Duggan et al. (2015) 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 an 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 15,012,189 enrollees in 75,417 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 and outcomes for the universe of Medicare beneficiaries in 2008-2011. The average Medicare beneficiary is 78 years old; 86% are white, 42% are male, and 16% are low-income and eligible for Medicaid in addition to Medicare ("dual-eligibles"). In any given year of our sample, 8.1% of Medicare beneficiaries change plans and 5.6% die. Within a county-year, we find about 25 plans in the median beneficiary choice set (including both traditional Medicare and MA plans).

Columns 2-4 of Table 1 summarize the subpopulation of beneficiary-years who ended the previous year in an 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 (22.6%) and their annual mortality rate is somewhat lower than in the full sample (4.7%).

Columns 3 and 4 of Table 1 summarize the subpopulations of enrollees of MA plans that

Table 1: Summary Statistics

	All Medicare Plans (1)	IV Sample		
		All MA Plans (2)	Non-Terminated (3)	Terminated (4)
Beneficiary Age	77.5	77.3	77.3	77.0
% White	85.5	87.1	87.0	90.4
% Male	41.9	41.3	41.3	43.4
% Dual-Eligible	15.9	8.6	8.7	6.2
% Switched Plans	8.1	22.6	20.9	100.0
% Died	5.6	4.7	4.7	4.5
Median N Plans in Choice Set	25	33	33	25
Total Plans	226,460	75,417	65,768	9,649
N Beneficiary-Years	118,184,127	15,012,189	14,682,291	329,898

Notes: This table summarizes the analysis samples in 2008-2011. Column 1 reports average enrollee demographics, annual plan switching rates, and annual mortality of the full Medicare population. Column 2 restricts the sample to beneficiary-years who ended the previous year in an 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.

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.[5] The largest difference in these samples is the annual plan-switching rate: while all beneficiaries previously enrolled in a terminated plan are forced to change to a new MA plan, only 20.9% of beneficiaries in non-terminated plans switch.[6]

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

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

[6] Appendix Table A2 describes switching behavior in more detail. In the full sample, 77.4% of enrollees do not switch plans in any given year. Among those consumers, 7.8% enroll in a different plan offered by the same insurer and 14.8% 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 81.2% enroll in a plan offered by a different insurer. Thus the vast majority of termination-induced switches are to new insurers within a market.

least squares (OLS) regressions, of the form

$$Y_{it} = \sum_j \mu_j D_{ijt} + X_{it}'\omega + \varepsilon_{it},, \tag{1}$$

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-eligibile 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. Given the fixed effects and controls, variation in the observational mortality coefficients μ_j thus reflects within-county differences in one-year MA plan mortality rates among observably similar enrollees.

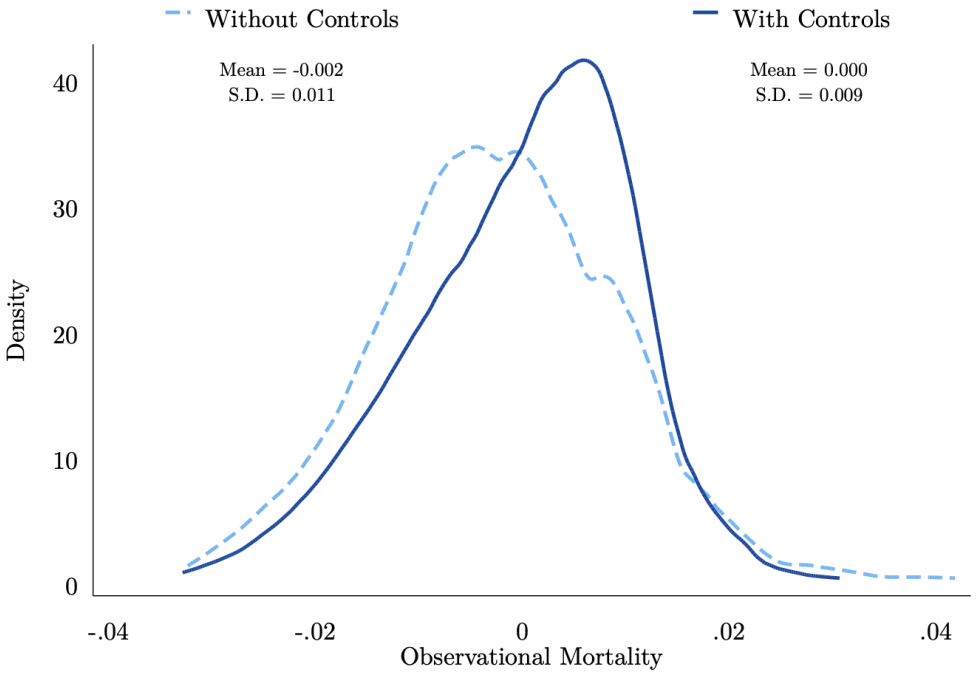
We account for statistical noise in the observational mortality estimates by applying a conventional empirical Bayes correction (Morris, 1983). This correction, detailed in Appendix C.1, "shrinks" the estimated μ_j towards their county- and plan size-level mean, in proportion to their expected degree 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 over 9,000 beneficiary-years. The average effective shrinkage coefficient is very close to one.[7]

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, after correcting for estimation error, is 0.9 percentage points or 19% 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. The model without controls has a slightly lower mean (implying that MA plans have observably healthier beneficiaries, on average) and a 20% larger standard deviation of 1.1 percentage points.

The changing mean and standard deviation of observational mortality when beneficiary demographic controls are included suggests some degree of selection-on-observables. In other words, the variation in observational mortality from the simpler specification appears to be in part driven by observable differences in 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 on the estimated distribution of observational mortality (e.g. the noise-adjusted stan-

[7] Appendix 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 (traditional Medicare 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 computed as described in Appendix C.1 and shown for each estimation procedure.

dard deviation of μ_j remains at 0.9 percentage points). Absent further observables, we are unable to directly test for remaining selection bias in our benchmark specification. 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 an MA plan with a high observational mortality rate μ_j. Since Medicare plan choice is highly inertial (only 23% 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 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. 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 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 reflect causal effects, then cohort 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.

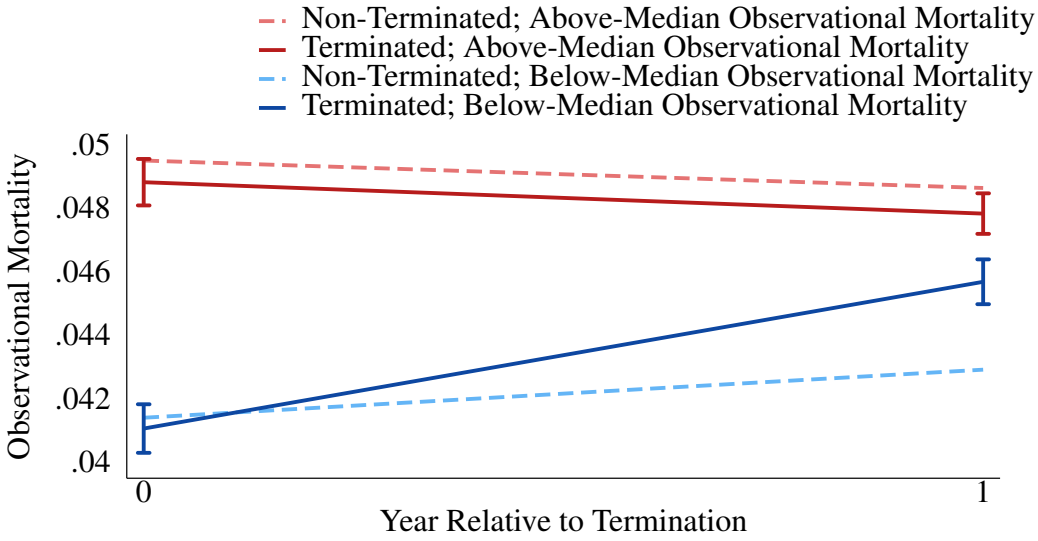
Figure 2 illustrates the relationship between plan mortality rates and termination status for high- and low-mortality plans in our IV sample. The solid lines indicate regression-adjusted trends in observational mortality for beneficiaries before and after a plan termination, separately for beneficiaries previously enrolled in plans with above-median (blue) and below-median (red) mortality.[8] 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 similar 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 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 rate 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, while average mortality prior to termination is more similar.

Together, the differential trends in Figures 2 and 3 suggest that a termination-induced move

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

Figure 2: Plan Terminations and Observational Mortality



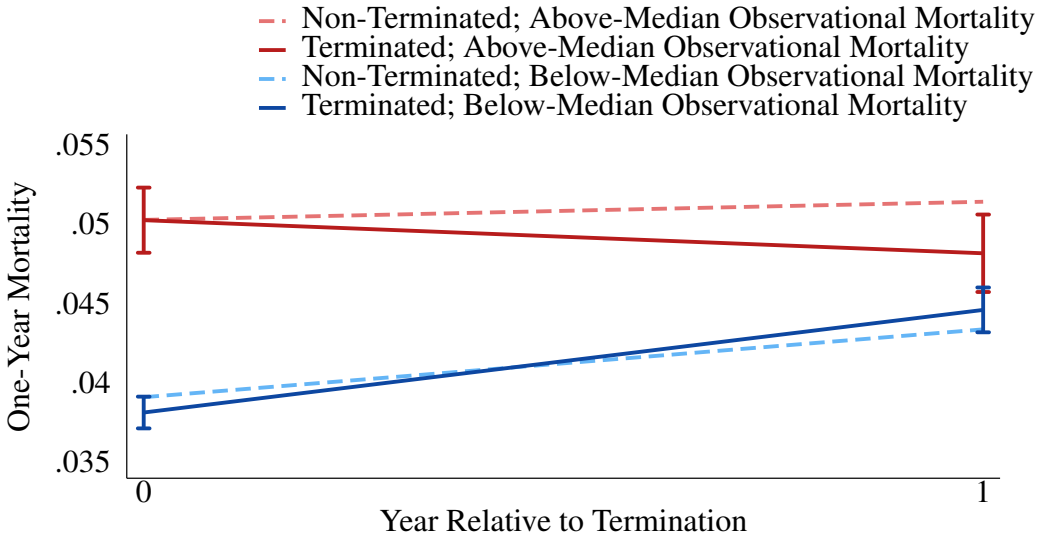
Notes: This figure shows regression-adjusted trends in the observational mortality of enrollees of non-terminated and terminated MA plans, separately for plans with above- and below-median observational mortality. The median is defined over the entire IV sample. Year 0 is defined as the last year prior to termination for terminated plans and year 1 is 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, lagged observational mortality, 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.

to MA plans with different observational mortality μ_j has a differential causal effect on actual mortality Y_{it}. This finding 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 reliability of observational mortality for such causal effects.

3 Econometric Framework

We develop an instrumental variables (IV) framework for using plan terminations to measure the reliability 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, the "fallback condition," which is novel to our paper.

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. Year 0 is defined as the last year prior to termination for terminated plans and year 1 is 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, lagged observational mortality, 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.

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 $Y_{ijt} = \beta_j + \upsilon_{it}$; 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 υ_{it} capturing beneficiary health. Projecting υ_{it} on a vector of observable characteristics X_{it} (which includes a constant) yields

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

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. Consumer mortality is then given by $Y_{it} = \sum_j Y_{ijt} D_{ijt}$.

Substituting in the previous expression for Y_{ijt} yields:

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

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.[9]

3.2 The Forecast Coefficient

Our first goal is to measure the reliability of observational mortality μ_j in predicting the variation in true MA mortality effects β_j. Formally, we seek to estimate the MA forecast coefficient λ, defined by the the projection of causal mortality effects β_j on observational mortality μ_j (here, without loss, also normalized to zero):

$$\beta_j = \lambda \mu_j + \eta_j, \tag{4}$$

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 reliability 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

[9] Estimating 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.

causal effects when λ is small.[10] 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. $\mu_j = \beta_j$).[11] 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, $Var(\beta_j) \geq \lambda^2 Var(\mu_j)$.

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

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

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.[12] To estimate the forecast coefficient, we will 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 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 μ_j of 0.05 and the third plan C has an observational mortality of 0.03. Suppose plan C exogenously terminates, and that

[10] This definition of the forecast coefficient makes $1 - \lambda$ align with the notion of "forecast bias" in the education value-added literature (Kane and Staiger, 2008; Chetty et al., 2014; Angrist et al., 2017).

[11] Chetty 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 $\lambda = 1$.

[12] In fact, when the control vector X_{it} is the same in these two models, OLS estimation of Equation (5) which uses the first-step estimates of μ_j from Equation (3) will mechanically give a λ estimate of 1, even when observational mortality is a badly biased predictor of true mortality effects. This follows by standard projection algebra.

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., 2017).

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

To see how the forecast residual can yield misleading quasi-experimental estimates of the forecast 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 observational reliability is quite general. It arises whenever an instrument is used to estimate the relationship between observational value-added estimates or other attributes of an entity (e.g. mortality for hospitals or test-scores for teachers) and (unknown) causal effects. Doyle et al. (2017), for example, leverage the quasi-experimental assignment of emergency Medicare patients to different ambulance companies in IV regressions of short-term mortality on the average spending of the hospital a patient is admitted to. In this case, interpreting the resulting IV coefficient as a measure of the correlation between hospital mortality effects and hospital spending requires more than random assignment of patients to ambulances. It also requires knowing that certain ambulance companies do not systematically

refer patients to hospitals which are higher quality than would be predicted by the quality-spending relationship. Likewise, even if teachers quasi-randomly move across schools (as in Chetty et al. (2014)), the within-school assignment of teachers to classrooms may matter for estimation of the predictive reliability of observational value-added estimates.[13]

While the three-plan example may make this challenge seem intractable, such pessimism is unwarranted in our setting. When pooling termination-induced choice set variation across many markets, the solution becomes weaker and more natural. We show below that it 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 exclusion restrictions required by our IV approach.

The First-Stage and Exclusion Restriction 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

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

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

[13] The 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)). Our approach shows that a previously overlooked condition is needed to estimate the forecast coefficient when such bias might be present.

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 exclusion restriction: that Z_{it} is conditionally uncorrelated with unobserved beneficiary health ε_{it}.

Assumption 2. *(Exclusion):* $Cov(\tilde{Z}_{it}, \varepsilon_{it}) = 0$.

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,

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

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.[14] 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 exclusion restriction in Section 4.1 below.

The Fallback Condition The third identification condition we consider is novel, and follows the previous 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:

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

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

Recall that $\eta_{it} = \sum_j D_{ijt}\eta_j$ is the unobserved 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}. Consumers in terminated plans may subsequently make systematically different choices than those in non-terminated plans, in a manner that induces correlation between $\tilde{Z}_{it}$ and 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 fallback condition by asking what restrictions on consumer choices of plans would lead it to hold.

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 $\bar{\varepsilon}_{jt}$ 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 should hold when 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.3 formalizes this intuition with a discrete choice model that satisfies the fallback condition. The simplest version of the model assumes that consumers in non-terminated plans are fully inertial, while consumers in terminated plans make an unrestricted choice that maximizes their latent utility U_{ijt}. We show that the fallback condition holds provided the IV control vector X_{it} includes any lagged chararacteristics of plans that lead to persistent unobserved heterogeneity

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

$$U_{ijt} = \alpha_{it}' W_{jt} + \xi_j + u_{ijt}, \tag{8}$$

where α_{it} captures potentially heterogeneous preferences over observed plan characteristics W_{jt}, ξ_j denotes a fixed plan unobservable, and u_{ijt} captures unobserved idiosyncratic time-varying plan-specific preferences. We show in Appendix C.3 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 lagged plan characteristics are controlled for: i.e. 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.3 we discuss how any controls sufficient to capture persistent heterogeneity in plan choice probabilities can be included to satisfy Assumption 3 more generally.

The microfoundation suggests that 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 a school value-added model. 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.[15]

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 exclusion restriction, the fallback condition may be investigated empirically. Assumption 3 asserts that the forecast error of a beneficiary's plan, η_{it}, is conditionally uncorrelated with the instrument $\tilde{Z}_{it}$. We do not observe this residual directly, just as we do not observe the beneficiary residual ε_{it} which enters Assumption 2. However, just as standard IV falsification tests can investigate whether the instrument is correlated with observable proxies for 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

[15] Suppose, for example, that consumers in terminated high observational mortality plans learn to better identify plans with low β_j when forced to make an active choice. These consumers might choose plans with systematically smaller η_j following terminations. This may cause us to 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.

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, exclusion, 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.4 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 exclusion restriction 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 of course 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.5 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, exclusion, 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 exclusion restriction is analogous to before, 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:

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

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 lagged plan observational mortality 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).[16] 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.

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 documented 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 -0.35, implying that a consumer in a one percentage point higher observational mortality plan in the previous period switches to a plan with 0.35 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

[16] Plan 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.

Table 2: Tests of Assumptions

	(1)	(2)
Dep. Var.: Observational Mortality	A. First Stage	
Instrument	−0.349	−0.0055
	(0.037)	(0.0011)
F Statistic	89.6	24.3
Dep. Var.: Predicted Mortality	B. Exclusion	
Instrument	0.042	−0.0003
	(0.029)	(0.0007)
Dep. Var.: Predicted Forecast Residual	C. Fallback	
Instrument	0.002	−0.0001
	(0.002)	(0.0001)
Specification	Linear	Median
Demographic Controls	No	No
N Beneficiary-Years	15,012,189	

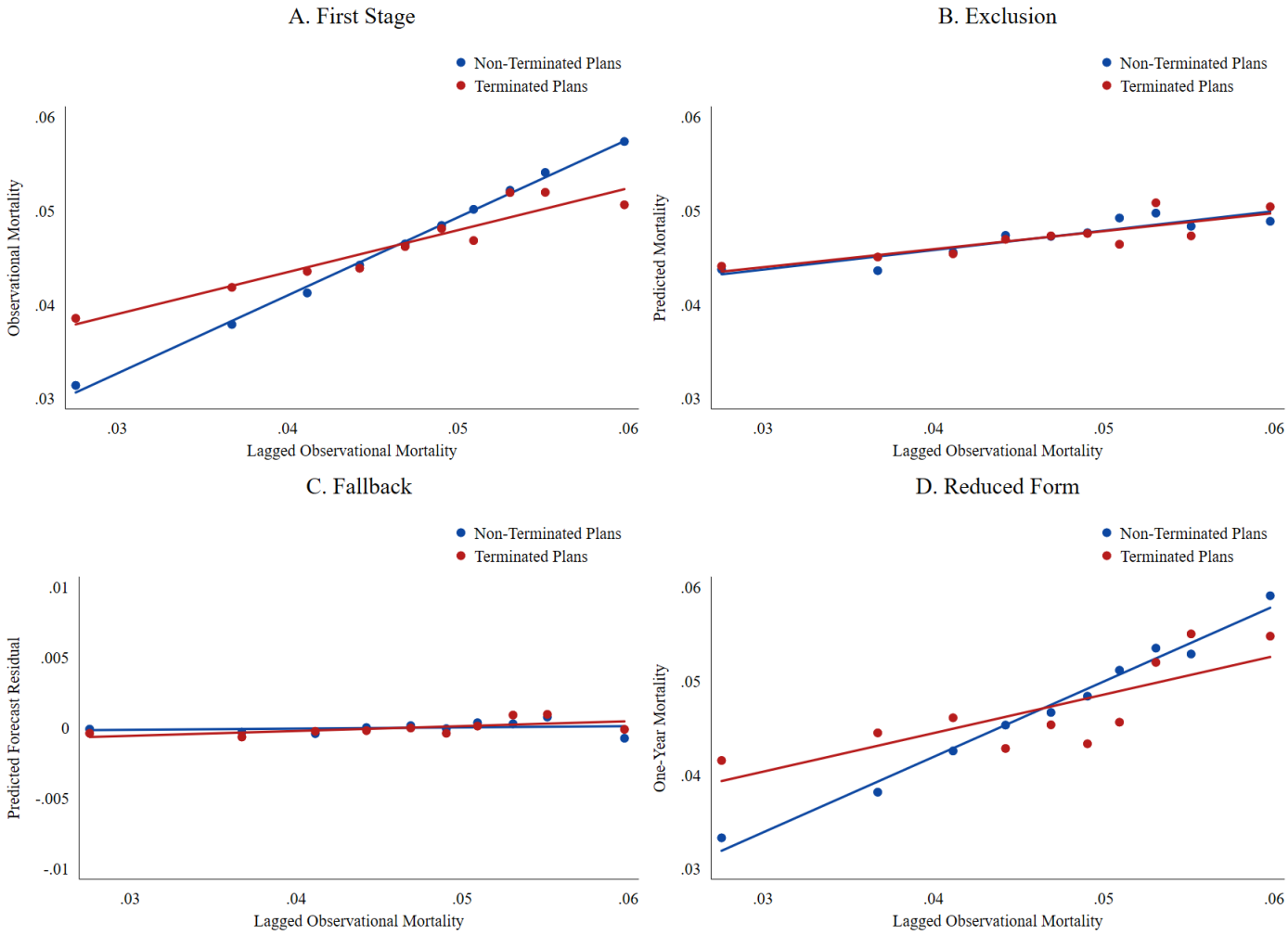
Notes: Panel A of this table is based on estimation of Equation (9) and presents the OLS coefficient in a first-stage regression of observational mortality on the instrument. Panel B replaces observational mortality as the dependent variable with a prediction of one-year mortality based on beneficiary 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 lagged plan observational mortality and a lagged plan termination indicator. In column 2 the instrument is the interaction of an indicator for above-median lagged plan observational mortality and a lagged plan termination indicator. In all specifications, we control for the lagged plan 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). Standard errors are clustered by county and reported in parentheses.

the observational mortality of year t plans of 0.55 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 Equation (9) with one based on deciles of lagged observational mortality (and including decile main effects). We then use this specification to plot the contemporaneous plan observational mortality of enrollees who, in the previous year, were enrolled in plans of different deciles of observational mortality that did and not terminate. The figure shows that while lagged plan observational mortality predicts current plan observational mortality among both groups, the relationship is much flatter 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.

We next build support for the IV exclusion restriction (Assumption 2) by testing whether the

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 combined with the decile-specific termination effects estimated from specifications of the form of Equation (9), with controls as in Table 2, including decile main effects. Coefficients are normalized to remove termination main effects.

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 0.042, while in the median specification we obtain an insignificant coefficent of -0.0003. Both of these estimates are more than an order of magnitude smaller than the corresponding first-stage estimates, with the linear specification coefficient of opposite sign. Finding balance for our instrument on predicted mortality is not surprising in light of the motivating Figure 3, which showed a lack of imbalance of terminations on lagged mortality. Appendix Figure A3 further shows that our instrument is balanced on average CMS risk scores, which attempt to predict enrollee costs based on demographics and diagnoses and are available for a subset of plan-years in our data.[17] All three findings are consistent with Assumption 2.

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.

Finally, we build support for the novel fallback condition (Assumption 3) by testing whether our instrument predicts an observable proxy for the forecast residual η_i. We construct the proxy by first regressing observational mortality on a set of observable plan characteristics (plan star ratings, premiums, medical loss ratios, and an indicator for donut hole coverage). We then take the residual from projecting the fitted values from this regression (as an observable proxy of β_j) on μ_j. This residual yields an observable proxy for η_j, and thus of $\eta_i = \sum_j \eta_j D_{ij}$ given a beneficiary's plan. Panel C of Table 2 reports the resulting instrument coefficients from replacing the outcome in Equation (9) with this proxy. As in Panel B, where we effectively constructed a proxy for the relevant beneficiary unobservable ε_i, we cannot reject the null of instrument balance on our proxy of the relevant plan unobservable. The linear specification yields an insignificant coefficient of 0.002, while with the median specification yields an insignificant coefficient of -0.0001.

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.

[17] The corresponding median specification from Table 2 for this observable proxy for beneficiary health gives an insignificant coefficient of 0.022 with a county-clustered standard error of 0.014.

This 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. These estimates are obtained from a second stage regression of

$$Y_{it} = \lambda \mu_{it} + X_{it}' \gamma_X + \varepsilon_{it} + \eta_{it}, \tag{10}$$

and a first stage given by Equation (9). The reduced form replaces the observational mortality outcome in Equation (9) with the actual mortality outcome in Equation (10). The second-stage coefficient λ estimates the observational mortality forecast coefficient under Assumptions 1-3. 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).

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 -0.34 and -0.39 for the linear specification (without and with demographic controls) and of -0.0068 and -0.0065 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 0.99-1.23, with standard errors in the range of 0.19-0.35. The point estimates are again similar with and without baseline demographic controls, which indeed tend to reduce the standard errors. The median specification yields a slightly higher forecast coefficient, although the estimates are not statistically distinguishable. Together, these IV estimates suggest observational mortality is a highly reliable predictor of true mortality effects.

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

Table 3: Forecast Coefficient Estimates

	(1)	(2)	(3)	(4)
Dep. Var.: Observational Mortality		A. First Stage		
Instrument	−0.349	−0.0055	−0.349	−0.0055
	(0.037)	(0.0011)	(0.037)	(0.0011)
F Statistic	89.6	24.3	89.4	24.3
Dep. Var.: One-Year Mortality		B. Reduced Form		
Instrument	−0.344	−0.0068	−0.386	−0.0065
	(0.099)	(0.0024)	(0.088)	(0.0020)
Dep. Var.: One-Year Mortality		C. Second Stage (Forecast Coefficient)		
Observational Mortality	0.986	1.230	1.107	1.183
	(0.230)	(0.353)	(0.187)	(0.310)
Specification	Linear	Median	Linear	Median
Demographic Controls	No	No	Yes	Yes
N Beneficiary-Years		15,012,189		

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 lagged plan observational mortality and a lagged plan termination indicator. In columns 2 and 4 the instrument is the interaction of an indicator for above-median lagged plan observational mortality 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.

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, while the relationship attenuates 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.

4.3 Robustness Checks

We verify the robustness of our forecast coefficient estimates in several exercises summarized in Appendix Table A4. 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. Panel A of Appendix Table A4 shows we obtain similar forecast coefficient estimates of around 1.09-1.15 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 Panel B of Appendix Table A4 with confidence intervals containing our baseline estimates.

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 increased PFFS operating costs. Panel C of Appendix Table A4 shows that these perhaps more plausibly exogenous plan terminations yield a similar forecast coefficient estimate of 1.15, though with larger standard error of 0.37. The corresponding median specification gives a larger but even more imprecise estimate, with a standard error of 0.78 (and a first-stage F statistic below 10).

Finally, we investigate the role of treatment effect heterogeneity. Panel D of Appendix Table A4 shows that we obtain similar estimates, of around 1.13-1.17, when we exclude dual-eligible beneficiaries from both the IV sample and the sample used to construct the observational mortality measure. Panel E further shows that our results are similar when we allow observational mortality to vary by beneficiary age, estimating Equation (1) separatelyby five-year age bins. This specification yields forecast coefficients of around 1.14-1.15, with 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.5 (that accommodates such heterogeneity) is likely to hold.

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 μ_j is a highly reliable predictor of β_j despite any such selection bias, in that $\lambda \approx 1$. 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.3),

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).[18] 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 0.9 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.[19] 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.[20] 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 for plan quality, capture financial generosity and potential mechanisms, or measure consumer willingness to pay for plan health effects.

[18] Formally, note that the forecast coefficient can be written as $\lambda = \frac{Cov(\beta_j,\mu_j)}{Var(\mu_j)} = \frac{Var(\beta_j)+Cov(\beta_j,b_j)}{Var(\beta_j)+Var(b_j)+2Cov(\beta_j,b_j)}$, 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.

[19] A growing literature also shows that insurance lowers mortality in the Medicaid program (Miller et al., 2019; Goldin et al., 2019). 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).

[20] Finkelstein 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 (2018) also find evidence of large place effects.

5.1 Proxies for 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.[21]

Suprisingly, we find that CMS star ratings are positively correlated with our observational mortality measure, suggesting higher-ranked plans have higher mortality rates.[22] The first column of Table 4, Panel A, shows that a one-star increase in a plan's ratings is associated with a 0.42 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),

$$Y_{it} = \theta W_{it} + X_{it}'\rho + \varepsilon_{it} + \eta_{W,it}, \tag{11}$$

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, exclusion 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 0.06 percentage point increase in plan effects, with a standard error of 0.14 percentage points. This suggests that the most commonly used measure of plan quality does not reliably predict which plans systematically reduce beneficiary mortality on average.

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

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

Table 4: Plan Characteristics Regressions

	(1)	(2)	(3)	(4)	(5)
	Panel A: OLS (Observational Mortality)				
Star Rating	0.0042				
	(0.0003)				
Premium		0.0048			0.0051
		(0.0005)			(0.0005)
Has Donut Hole Coverage			−0.0021		−0.0024
			(0.0003)		(0.0003)
Medical Loss Ratio				0.0142	0.0087
				(0.0035)	(0.0033)
	Panel B: IV (Plan Mortality Effect)				
Star Rating	0.0006				
	(0.0014)				
Premium		−0.0051			−0.0044
		(0.0020)			(0.0026)
Has Donut Hole Coverage			−0.0041		−0.0001
			(0.0016)		(0.0023)
Medical Loss Ratio				−0.0214	−0.0223
				(0.0044)	(0.0044)
First-Stage F Statistic	2,860.3	2,085.8	1,437.8	1,644.6	372.9
Maximum Forecast R^2	0.0005	0.0218	0.0214	0.0095	0.0298
N Beneficiary-Years	15,012,189				

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 0.5 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 0.5 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.[23]

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.[24] We find a maximum R^2 of 0.05% for star ratings and 2.18% 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.

[23] At a conservative $1 million VSL, a 0.5 percentage point reduction in mortality is worth $5,000.

[24] Formally, $Var(W_j'\theta)/Var(\beta_j) \leq Var(W_j'\theta)/Var(\lambda\mu_j)$ 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 (0.4 percentage points), on average. This contrast is consistent with earlier findings that sicker beneficiaries tend to to select into plans with donut hole coverage (e.g. Polyakova (2016)). The finding of large plan effect differences among plans which offer donut hole coverage suggests that lower cost-sharing may be more broadly beneficial.[25]

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.[26] 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 one standard deviation higher MLR (1.4 percentage points) is associated with a 2.1 percentage point reduction in the plan mortality effect. 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. (2014)), 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., 2014; 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.[27]

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

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

[27] Hospital network data is available from State Inpatient Databases, but consistent information on Medicare Advantage 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

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 2.98% in column 5 of Table 4, which includes all financial measures.[28] 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:

$$U_{ij} = \alpha p_j + \xi_j + u_{ij}, \tag{12}$$

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 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 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

fact that these data cover a relatively small number of markets makes it challenging to draw inferences.

[28] We 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.

(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.[29]

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:

$$ln\left(s_j\right) - ln\left(s_0\right) = \delta_j \equiv \alpha p_j + \tau\beta_j + \psi_j, \tag{13}$$

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

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

using the fact that $Cov(\beta_j, \psi_j) = 0$ by construction. Given Equation (13), $Var(\xi_j) = Var(\delta_j - \alpha p_j)$ is identified by market shares and the premium coefficient α. Our observational mortality forecast coefficient further identifies a lower bound on $Var(\beta_j) \geq \lambda^2 Var(\mu_j)$. The forecast coefficient κ then identifies a lower bound on $\tau = \kappa \frac{Var(\xi_j)}{Var(\beta_j)} \geq \kappa \frac{Var(\xi_j)}{\lambda^2 Var(\mu_j)}$ (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.[30] 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 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 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

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

[30] Curto 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	−0.0229	−0.0003	−403.95	176.38
		(0.0001)	(129.74)	(56.65)
-7	−0.0160	−0.0004	−284.07	177.19
		(0.0001)	(91.30)	(56.95)
-3.5	−0.0080	−0.0007	−144.81	180.66
		(0.0002)	(47.45)	(59.20)
-1	−0.0023	−0.0017	−46.43	202.75
		(0.0008)	(22.23)	(97.05)
-0.5	−0.0011	−0.0015	−25.23	220.30
		(0.0013)	(21.68)	(189.30)

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.

finding for hospitals in Chandra et al. (2015)).[31] 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 \$180-\$220, implying that consumers are willing to pay no more than this amount to offset a one percentage point increase in one-year mortality.[32] These estimates are around half of the average yearly premium in the sample (roughly \$500) 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, e.g., Kniesner and Viscusi (2019) and Murphy and Topel (2006)).

[31] In unreported regressions, we find that the largest insurers (Humana, United, and Blue plans) appear to supply higher quality plans. This is an interesting avenue for future research.

[32] Naïve 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 -167.89 with a standard error of 26.22, a nonsensical 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.

Conventional estimates imply a value of a one percentage point reduction in mortality of between $20,000 and $100,000. Although our WTP bounds increase and become more imprecise as we use a lower premium elasticity, our most conservative estimate is several orders of magnitude lower.[33]

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).

6 Plan Choice and Mortality

Our forecast coefficient estimates in Section 4.2 suggest that MA plan mortality effects are enormously variable within a market and can be reliably predicted by observational mortality differences. At the same time, our WTP estimates in Section 5.3 suggest that consumers place little weight on this dimension of plan quality when making enrollment decisions. Together, these findings imply that redirecting consumers from observably low-quality plans to plans with better observational mortality could substantially improve beneficiary health.

We quantify 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.[34] 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.

[33] Even if we assume that the deaths prevented by enrolling in lower mortality plans are delayed only a few years, with a $250,000 value of a statistical life-year (Cutler, 2005), the value of a one percentage point reduction in mortality should be at least an order of magnitude higher than we measure.

[34] All 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.

Table 6: Mortality and Plan Choice Simulations

	Change from Status Quo (1)	$\frac{\text{Change}}{\text{Mean Sample Mortality}}$ (2)
Random Assignment to Plans	0.0027	0.057
Assignment to Minimum-Mortality Plans	−0.0192	−0.408
Assignment from Top to Bottom Quartile	−0.0077	−0.163
Random Assignment from Top 5%	−0.0005	−0.011

Notes: The first row of this table uses a forecast coefficient estimate, obtained as in Table 4 by an IV regression of beneficiary mortality on the log market share of her plan, to predict how mortality would change if consumers were randomly assigned MA to plans within counties to the point of equalizing market shares. The remaining rows summarize the simulated change in observational mortality posteriors when MA beneficiaries are reassigned to plans, as described in the text. Given an observational mortality forecast coefficient near one, these results imply commensurate changes in mortality. All simulations are conducted on the MA sample, excluding plans with fewer than 12 beneficiaries in a given year.

The results are reported in the first row of Table 6. 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 increases average mortality by an average of 0.3 percentage points relative to the average one-year mortality rate of 4.7%. Scaling by a conservative VSL of one million dollars, this would imply that the partial equilibrium health benefits of active choice relative to (unconditional) random assignment are at least $2,700 per person.[35]

We next consider the scope for improving on random assignment 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 $\overline{\mu}_j^*$ (e.g. the average observational mortality). Our estimate of the average mortality improvement from such reassignment is then the average $\hat{\lambda}(\overline{\mu}_j - \tilde{\mu}_j)$, where $\hat{\lambda}$ is an estimate of the forecast coefficient and $(\overline{\mu}_j, \tilde{\mu}_j)$ denotes the actual observational mortality of plans with posteriors of $(\overline{\mu}_j^*, \tilde{\mu}_j^*)$. With $\hat{\lambda} = 1$, this simulation effectively predicts the potential mortality effect by the average change in observational mortality.[36]

[35] Of course, these benefits would be substantially lower if evaluated given the WTP estimates in the previous section rather than the $1 million VSL, but we do not think that doing so is normatively appropriate given that consumers are likely not fully informed about the significant differences in MA plan mortality effects.

[36] This exercise may understate the gains from active choice to the extent status quo choices reflect positive selection

The results are reported in the remaining rows of Table 6. In the second row we find that assigning beneficiaries to the observably best plan has a large impact. Given a forecast coefficient near one, the reassignment would reduce mortality by -1.9 percentage points, or -40.8%, relative to average mortality in the sample. This simulation helps illustrate the overall magnitude of variation in plan effects but is not especially realistic and may be driven by the assumed (normal) tails of the distribution of μ_j. In the third row of Table 6 we instead reassign beneficiaries from plans in the top quartile of observational mortality posteriors to those in the bottom quartile. Given a forecast coefficient near one, the reassignment would reduce mortality by a smaller (but still sizeable) 0.8 percentage points, or 16.3% of the average mortality rate.

Finally, row 4 of Table 6 presents a policy-relevant 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 (top 5% observational mortality) plans to other plans at random. This reassignment rule reduces reduces observational mortality posteriors by 0.05 percentage points, or 1.1%. With more than 20 million MA enrollees each year, even this small change in mortality would have a large impact, averting around 10,000 elderly deaths each year given a forecast coefficient of one.[37]

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 mortality (through a channel other than plan choice) in any specification.[38] Our simulations also do not

on unobserved treatment effect heterogeneity (Hull, 2020). As noted in Section 4.3 the role of such Roy selection appears small in this setting.

[37] This number is not additive across years, due to competing risks. In other words, some of of the 10,000 deaths that are averted in any given year will still die the following year. The 10,000 deaths averted per year thus corresponds to a minimum of 10,000 life-years saved per year.

[38] Sabety (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 0.06 percentage point reduction in mortality with a standard error of 0.08 percentage points. This rules out even a 0.1 percentage point increase in mortality, compared to the 1.5 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, 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).

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).[39] With a $1 million VSL, the health benefits of the reassignments in row 4 of Table 6 would be $10,000 per reassigned beneficiary-year, likely dwarfing any effects via switching costs or financial plan characteristics.

7 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 is a highly reliable predictor of true plan mortality effects with a novel quasi-experimental design. Publicly available quality measures are uncorrelated with true mortality effects. Perhaps as a result, consumer demand is under-responsive to this dimension of plan quality. In partial-equilibrium simulations we show that one-year mortality would fall significantly if beneficiaries were reassigned to lower observational mortality plans, suggesting 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 potentially large benefits from directing consumers to lower observational mortality plans. While the government does not currently release

[39] We thank Ben Handel and Sam Kina for providing this information.

risk-adjusted mortality information, we find that such information might be incredibly important. Existing programs subsidize plans that score better on measures like star ratings, which we find to be uncorrelated with causal mortality effects.[40] Such programs may therefore do a poor job of rewarding plans that improve beneficiary health and might do better if they targeted risk-adjusted mortality. 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. Our policy simulations are partial equilibrium; in general equilibrium, publishing observational mortality rates might induce plans to invest in selecting healthier beneficiaries rather than improving health. 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.[41] 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.

[40] See Decarolis and Guglielmo (2017) for an analysis of the MA Quality Bonus Payment Demonstration program.

[41] Nevertheless, the methods we develop here could help in adding these features: 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. Technical report, National Bureau of Economic Research.

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*.

Abaluck, J. and J. Gruber (2016b). Improving the quality of choices in health insurance markets. Technical report, National Bureau of Economic Research.

Allende, C. (2019). Competition under social interactions and the design of education policies. *Job Market Paper*.

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*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*(2).

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 (2014). Does privatized health insurance benefit patients or producers? Evidence from Medicare Advantage. NBER Working Paper No. 20470.

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.

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.

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.

Cutler, D. M. (2005). *Your money or your life: Strong medicine for America's health care system*. Oxford University Press.

Dafny, L. and D. Dranove (2008). Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. *The 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 (2018). Does when you die depend on where you live? Evidence from Hurricane Katrina. Technical report, National Bureau of Economic Research.

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 (2014). Paying for quality in healthcare. Working Paper.

Doyle, J., J. Graves, and J. Gruber (2017). Evaluating measures of hospital quality. NBER Working Paper No. 14607.

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. *Available at SSRN 3556857*.

Duggan, M., A. Starc, and B. Vabson (2015). Who benefits when government pays more? Pass-through in the Medicare Advantage program. NBER Working Paper No. 19989.

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. Technical report, National Bureau of Economic Research.

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 D. Prinz (2019). Screening in contract design: Evidence from the ACA health insurance exchanges. *American Economic Journal: Economic Policy*.

Geweke, J., G. Gowrisankaran, and R. J. Town (2003). Bayesian inference for hospital quality in a selection model. *Econometrica 71*(4), 1215–1238.

Goldin, J., I. Z. Lurie, and J. McCubbin (2019). Health insurance and mortality: Experimental evidence from taxpayer outreach. Technical report, National Bureau of Economic Research.

Gruber, J., B. R. Handel, S. H. Kina, and J. T. Kolstad (2020). Managing intelligence: Skilled experts and AI in markets for complex products. Technical report, National Bureau of Economic Research.

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. Technical report, National Bureau of Economic Research.

Hull, P. (2020). Estimating hospital quality with quasi-experimental data. Working Paper.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Technical report, National Bureau of Economic Research.

KFF (2009). Medicare advantage 2010 data spotlight: Plan availability and premiums.

KFF (2019). A dozen facts about medicare advantage in 2019.

Kniesner, T. J. and W. K. Viscusi (2019). The value of a statistical life. *Forthcoming, Oxford Research Encyclopedia of Economics and Finance*, 19–15.

Koijen, R. S. and S. Van Nieuwerburgh (2020). Combining life and health insurance. *The Quarterly Journal of Economics 135*(2), 913–958.

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. Technical report, National Bureau of Economic Research.

Miller, S., S. Altekruse, N. Johnson, and L. R. Wherry (2019). Medicaid and mortality: New evidence from linked survey and administrative data. Technical report, National Bureau of Economic Research.

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 (2006). The value of health and longevity. *Journal of Political Economy 114*(5), 871–904.

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. Technical report, Harvard University.

Sommers, B. D., A. A. Gawande, K. Baicker, et al. (2017). Health insurance coverage and health—what the recent evidence tells us. *N Engl J Med 377*(6), 586–593.

Staiger, B. (2020). Disruptions to the provider-patient relationship and patient utilization and outcomes: Evidence from Medicaid Managed Care.

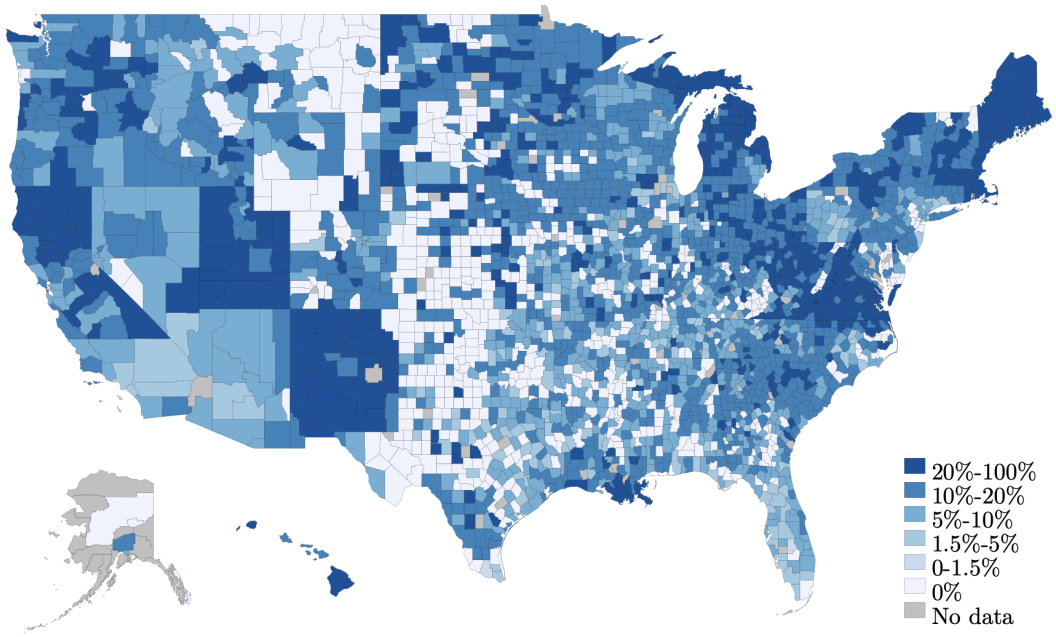
Starc, A. (2015, Spring). Insurer pricing and consumer welfare: Evidence from Medigap. *RAND Journal of Economics 45*(1), 198–220.

Starc, A. and R. Town (2019). Benefit integration in health insurance. *Forthcoming, Review of Economic Studies*.

Yakusheva, O., R. Lindrooth, and M. Weiss (2014). Nurse mortality-added and patient outcomes in acute care. *Health Services Research 49*(6), 1767–1786.

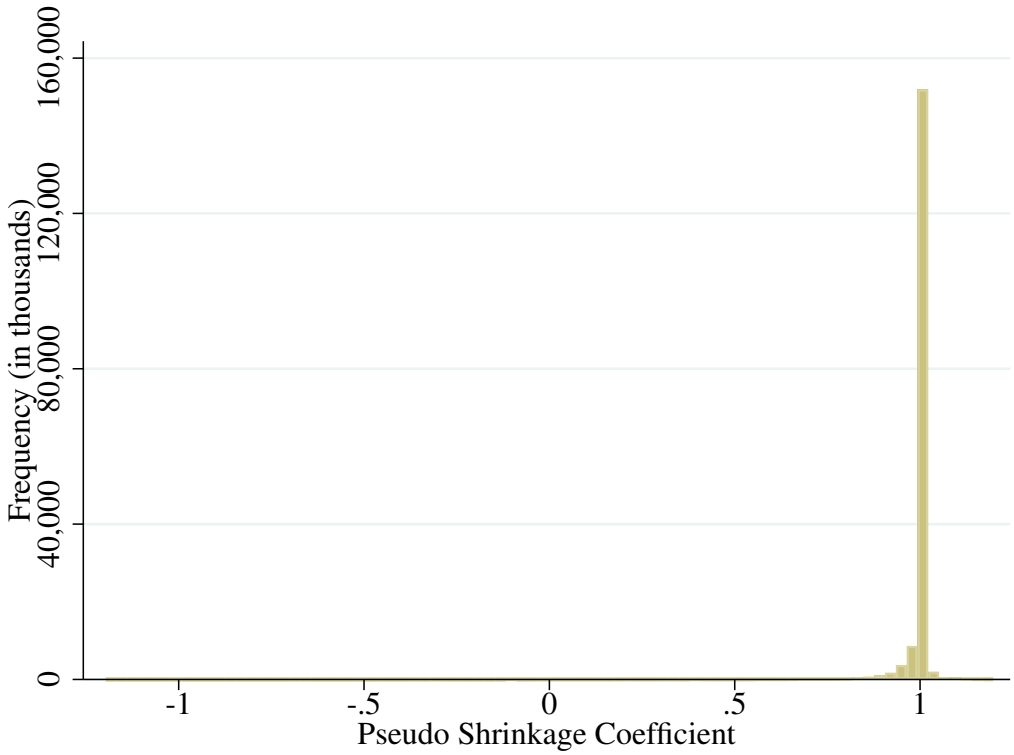
A Appendix Figures and Tables

Figure A1: Geographic Distribution of Plan Terminations



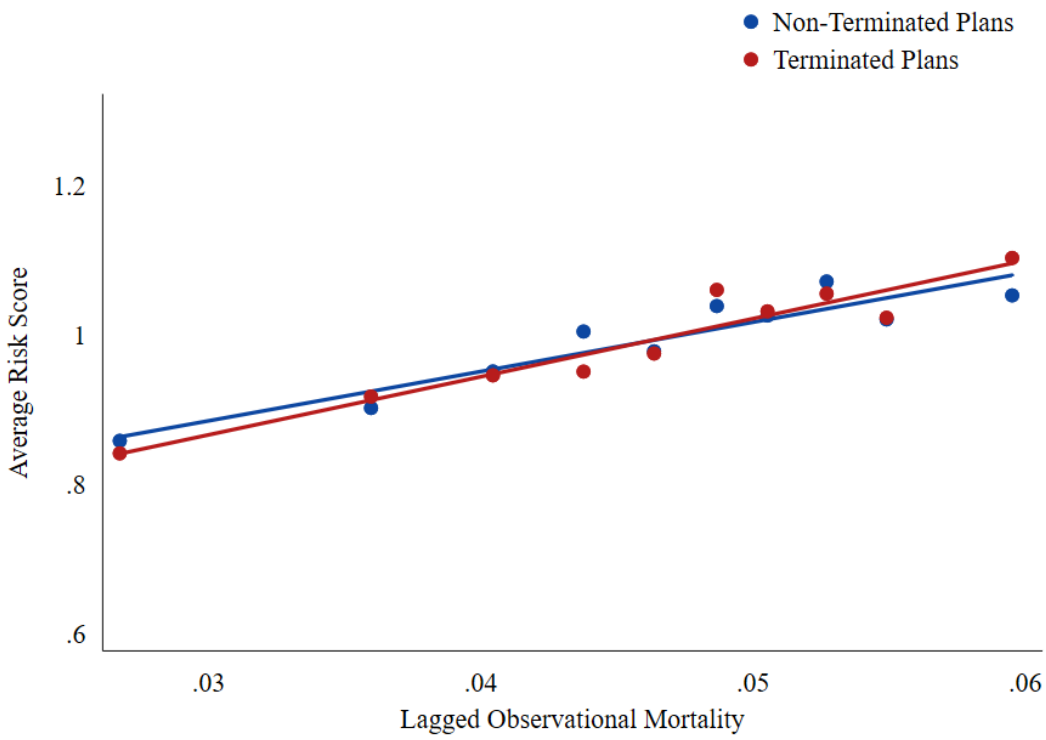
Notes: This map shows the fraction of plans in a county that were terminated over 2008-2011, with counties shaded according to the quantiles reported in the legend.

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: Plan Terminations and Beneficiary Risk Scores



Notes: This figure shows average beneficiary risk scores 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. Points are given by average risk scores combined with the decile-specific termination effects estimated from specifications of the form of Equation (9), with controls as in Table 2, including decile main effects. Coefficients are normalized to remove termination main effects.

Table A1: County Characteristics

	No Terminations	Terminations
Median Income	41,182	44,922
% of Pop >65	17.35	16.32
% Dual	13.80	14.06
% White	83.46	84.00
% Black	8.38	9.19
% Asian	0.61	1.38
Population	26,190	119,638
Population Density	52	315
Counties	595	2,466
Number of Beneficiaries	198,448	5,322,676

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

Sample:	MA	Terminated
% Do Not Switch	77.4	0.1
% Switch Plans within Same Insurer	7.8	18.6
% Switch Insurer	14.8	81.2
% Switch Into PFFS Plan	4.6	17.6
N Beneficiary-Years	15,012,189	329,898

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	0.0051
	(0.0002)
Age 75-79	0.0183
	(0.0002)
Age 80-84	0.0410
	(0.0002)
Age 85-90	0.0800
	(0.0002)
Age 90-94	0.1425
	(0.0003)
Age 95+	0.2385
	(0.0006)
Female	−0.0188
	(0.0001)
White	0.0068
	(0.0020)
Black	0.0057
	(0.0020)
Other	−0.0054
	(0.0021)
Asian	−0.0099
	(0.0021)
Hispanic	−0.0068
	(0.0020)
Native American	0.0090
	(0.0024)
Dual	0.0453
	(0.0002)
R^2	0.039
N Beneficiary-Years	15,013,172

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: Forecast Coefficient Robustness Checks

	(1)	(2)
	A. Counties With Terminations	
Observational Mortality	1.085	1.150
	(0.189)	(0.309)
First-Stage F Statistic	89.6	24.4
N Beneficiary-Years	14,644,200	
	B. No TM Enrollments	
Observational Mortality	1.380	1.325
	(0.219)	(0.289)
First-Stage F Statistic	122.0	32.9
N Beneficiary-Years	14,166,119	
	C. PFFS Terminations	
Observational Mortality	1.154	1.987
	(0.369)	(0.778)
First-Stage F Statistic	54.1	7.2
N Beneficiary-Years	14,904,951	
	D. No Dual-Eligibles	
Observational Mortality	1.132	1.169
	(0.207)	(0.313)
First-Stage F Statistic	107.1	30.5
N Beneficiary-Years	13,151,504	
	E. Age-Specific Effects	
Observational Mortality	1.146	1.135
	(0.088)	(0.140)
First-Stage F Statistic	829.2	231.1
N Beneficiary-Years	15,012,189	
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 column 1 the instrument is the interaction of lagged plan observational mortality and a lagged plan termination indicator. In column 2 the instrument is the interaction of an indicator for above-median lagged plan observational mortality and a lagged plan termination indicator. Panel A drops counties with no terminations over 2008-2011. Panel B drops beneficiaries who switch to a TM plan. Panel C drops non-PFFS plans that terminate. Panel D drops dual-eligible beneficiaries. Panel E allows observational mortality to depend on beneficiary age, as described in the text. In all specifications, we control for the lagged plan 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), and beneficiary demographics. Standard errors are clustered by county and reported in parentheses.

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, plnd 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 (total claims with permitted adjustments + total expenses for activities to improve healthcare quality) / (total premium adjusted for payments to or from the federal and state high risk pools - 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.[42] 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 $\underline{\mu} = (\mu_1, \ldots, \mu_J)$ 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

$$\mu_j = w_{c(j)} + u_j, \tag{15}$$

where $E[w_c] = E[u_j] = 0$, $Var(w_c) = \sigma_w^2$, $Var(u_j) = \sigma_u^2$. 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 $\widehat{\underline{\mu}}$, where $\widehat{\mu}_j = \mu_j + e_j$ 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 $Var(w_c) = \sigma_w^2$ and $Var(u_j) = \sigma_u^2$ yield our overall estimate of the standard deviation of observational mortality, according to Equation (15).

"Shrunk" empirical Bayes posteriors of observational mortality are given by Equation (15) and our estimates of $Var(w_c) = \sigma_w^2$, $Var(u_j) = \sigma_u^2$, and $Var(u_j) = \sigma_e^2$. Formulas for these posteriors are derived from the regression of $\underline{\mu}$ on $\widehat{\underline{\mu}}$ and give the best linear unbiased prediction of $\underline{\mu}$ from $\widehat{\underline{\mu}}$ 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

$$\begin{aligned}\underline{\mu}^* &= Cov([\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C], [\mu_A, \mu_B, \mu_C]')Var([\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C]')^{-1}[\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C]' \\ &= V(V + diag([\sigma_{e,A}^2, \sigma_{e,B}^2, \sigma_{e,C}^2])^{-1}[\hat{\mu}_A, \hat{\mu}_B, \hat{\mu}_C]',\end{aligned} \tag{16}$$

[42] Roughly equal-sized bins are given by the following cutoffs on plan size: $\leq$ 1,000, 1,000-2,500, 2,500-5,000, 5,000-15,000, 15,000-25,000, 25,000-75,000, 75,000-200,000, and $\geq$ 200,000.

where

$$V = Var(\underline{\mu}) = \begin{bmatrix} \sigma_w^2 + \sigma_u^2 & \sigma_w^2 & 0 \\ \sigma_w^2 & \sigma_w^2 + \sigma_u^2 & 0 \\ 0 & 0 & \sigma_w^2 + \sigma_u^2. \end{bmatrix} \tag{17}$$

This shows that the posterior for the third plan C is

$$\mu_C^* = \frac{\sigma_w^2 + \sigma_u^2}{\sigma_w^2 + \sigma_u^2 + \sigma_{e,C}^2} \hat{\mu}_C = \frac{Var(\mu_C)}{Var(\mu_C) + Var(e_C)} \hat{\mu}_C, \tag{18}$$

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 (18). 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 $Var(e_j)$ is very small), making $\underline{\mu}^*$ close to $\widehat{\mu}$. This fact is summarized in Appendix Figure A2, which shows the distribution of a "pseudo shrinkage coefficient," $\widehat{\mu}_j / \mu_j^*$ given our estimates of the variance parameters in Equation (15). The median coefficient is one, with nearly all coefficients found to be larger than 0.75.

C.2 Plan-Level Exclusion Restriction

This Appendix shows how our exclusion restriction (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, $Cov(\tilde{Z}_{it}, \varepsilon_{it}) = 0$ if and only if $\phi_Z = 0$ in the beneficiary-level regression of

$$\varepsilon_{it} = \phi_Z Z_{it} + X_{it}' \phi_X + e_{it}. \tag{19}$$

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 $Pr(D_{ij,t-1} = 1)$-weighted plan-level regression of

$$\bar{\varepsilon}_{jt} = \phi_Z \mu_j T_{j,t-1} + X_{j,t-1}' \phi_X + e_{jt}, \tag{20}$$

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

C.3 Discrete Choice Microfoundation

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, that consumers in non-terminated plans are fully inertial, and that consumers in terminated plans make an unrestricted choice to maximize a latent utility of

$$U_{ijt} = \alpha_{it}' W_{jt} + \xi_j + u_{ijt}. \tag{21}$$

Here α_{it} captures potentially heterogeneous preferences over observed plan characteristics W_{jt} and ξ_j denotes a fixed plan unobservable. We follow the standard convention in such models (e.g. Berry et al. (1995)) of treating residual utility u_{ijt} as an independent *iid* shock, though we do not require any parametric assumptions on the distribution of α_{it} or u_{ijt}. 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 (21) 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 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

$$\Pi_{k\to j}(X_{k,t-1}) \equiv Pr(D_{ijt} = 1 \mid D_{ik,t-1} = 1, X_{k,t-1}) = Pr(D_{ijt} = 1 \mid X_{k,t-1}) \equiv \pi_j(X_{k,t-1}), \tag{22}$$

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 (22) holds when fallback choice probabilities $\Pi_{k\to j}(X_{k,t-1})$ 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 (22) 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 average observational mortality to zero in this subsample: i.e. $\sum_{j:X_{j,t-1}=x} \mu_j = 0$ for the conditioning value x. The forecast residual η_j is furthermore defined to be mean-zero and uncorrelated with μ_j in this subsample. This means the the instrument and fallback

residual are conditionally uncorrelated among beneficiaries in non-terminated plans:

$$Cov(Z_{it}, \eta_{it} \mid T_{i,t-1} = 0, X_{it} = x) = \sum_{k:X_{k,t-1}=x} \mu_k \eta_k = 0, \tag{23}$$

where we use the as-good-as-random assignment of plan terminations and the fact that beneficiaries in non-terminated plans are fully inertial. Furthermore, when Equation (22) holds, we have the same relationship among beneficiaries in terminated plans:

$$Cov(Z_{it}, \eta_{it} \mid T_{i,t-1} = 1, X_{it} = x) = \sum_{k:X_{k,t-1}=x} \mu_k \left(\sum_j \eta_j \Pi_{k \to j}(x) \right) = \sum_j \eta_j \pi_j(x) \left(\sum_{k:X_{k,t-1}=x} \mu_k \right) = 0. \tag{24}$$

This logic extends to feasible IV regressions which control flexibly for the lagged plan characteristics that make Equation (22) hold.

It remains to be shown that the discrete choice model (21) admits a set of $X_{j,t-1}$ satisfying Equation (22). We show this by building up to Equation (21) 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. 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 (22) holds provided the lagged plan characteristics that consumers exhibit heterogeneous and persistent preferences over, in $W_{k,t-1}$, 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 variation in fallback choice probabilities appears very general. Consider, for example, a version of Equation (21) which allows for random coefficients on plan unobservables:

$$U_{ijt} = \alpha_{it}' W_{jt} + \nu_{it} \xi_j + u_{ijt}, \tag{25}$$

with ν_i potentially correlated with ε_i, along with α_{it} and u_{ijt}. If ξ_j were observed, one could control for it in $X_{i,t}$ to account for any persistent unobserved heterogeneity due to ν_{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 ν_{it} is almost-surely positive, for example, the market share functions given by (25) 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 (25).

Second, we note that one may extend the model to allow for imperfect inertia among non-terminated beneficiaries. Suppose, for example, that an exogenous proportion of such beneficiaries are free to make an active choice each year and maximize utility as if their plan had terminated. Then the conditional fallback choice probabilities of beneficiaries in terminated and non-terminated plans will again be the same, satisfying Equation (22) and thus the fallback condition.

C.4 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 $\widehat{\mu}_{it} = \sum_j \widehat{\mu}_j D_{ijt}$, instrumented by some $\tilde{Z}_{it}$ satisfying our first-stage, exclusion, and fallback conditions, write $\widehat{\mu}_j = \mu_j + e_j$ for idiosyncratic estimation error satisfying $E[e_j] = 0$, $Var(e_j) = \sigma^2_{e,j}$, and $Cov(e_j, \mu_j) = Cov(e_j, \eta_j) = 0$. Suppose for simplicity that $E[\beta_j] = E[\mu_j] = 0$ and $Var(\mu_j) = \sigma^2_\mu$. Then, under Assumptions 1-3 this IV regression identifies an attenuated forecast coefficient, given by

$$\tilde{\lambda} = \frac{\frac{1}{J}\sum_j Cov(\beta_j, \widehat{\mu}_j)}{\frac{1}{J}\sum_j Var(\widehat{\mu}_j)} = \frac{\frac{1}{J}\sum_j Cov\left(\lambda\mu_j + \eta_j, \mu_j + e_j\right)}{\frac{1}{J}\sum_j Var(\mu_j + e_j)} = \lambda \frac{\sigma^2_\mu}{\sigma^2_\mu + \frac{1}{J}\sum_j \sigma^2_{ej}}. \tag{26}$$

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

As with classic attenuation bias, this attenuation bias can be addressed by replacing $\widehat{\mu}_j$ with a posterior mean μ^*_j like those considered in Appendix C.1. To see this simply suppose $\mu^*_j = \frac{\sigma^2_\mu}{\sigma^2_{ej}+\sigma^2_\mu}\widehat{\mu}_j$ 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

$$\begin{aligned}\frac{\frac{1}{J}\sum_j Cov(\beta_j, \mu^*_j)}{\frac{1}{J}\sum_j Var(\mu^*_j)} &= \frac{\frac{1}{J}\sum_j Cov\left(\lambda\mu_j + \eta_j, \frac{\sigma^2_\mu}{\sigma^2_{ej}+\sigma^2_\mu}(\mu_j + e_j)\right)}{\frac{1}{J}\sum_j Var\left(\frac{\sigma^2_\mu}{\sigma^2_{ej}+\sigma^2_\mu}(\mu_j + e_j)\right)} \\ &= \lambda \frac{\frac{1}{J}\sum_j \frac{\sigma^4_\mu}{\sigma^2_{ej}+\sigma^2_\mu}}{\frac{1}{J}\sum_j \frac{\sigma^4_\mu}{\left(\sigma^2_{ej}+\sigma^2_\mu\right)^2} Var\left(\mu_j + e_j\right)} = \lambda.\end{aligned} \tag{27}$$

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.5 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

$$Y_{ijt} = \beta_j + X_{it}'\gamma + \varepsilon_{it} + \zeta_{ijt}, \tag{28}$$

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 yields a second-stage equation of

$$\begin{aligned} Y_{it} = \sum_j Y_{ijt} D_{ijt} &= \sum_j \beta_j D_{ijt} + X_{it}'\gamma + \varepsilon_{it} + \zeta_{it} \\ &= \lambda \mu_{it} + X_{it}'\gamma + \varepsilon_{it} + \eta_{it} + \zeta_{it}, \end{aligned} \tag{29}$$

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, exclusion, and fallback conditions continue to be necessary for estimation of Equation (29) 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

$$\bar{\zeta}_{jt} = \phi_Z \mu_j T_{j,t-1} + X_{j,t-1}'\phi_X + e_{jt}, \tag{30}$$

where $\bar{\zeta}_{jt} = E[\zeta_{it} \mid D_{ij,t-1} = 1]$ 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 $Cov(\tilde{Z}_{it}, \xi_{it}) = 0$, 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.3 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; we discuss such an exercise in Section 4.3.