Mortality Effects and Choice Across Private Health Insurance Plans

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

January 2020

Abstract

For competition in health insurance markets to foster quality improvements, consumers should attend to meaningful differences in plan health effects. We study whether they do in Medicare Advantage (MA) markets, where beneficiaries choose from an array of private managed care plans subsidized by the government. Observational "mortality-added" estimates suggest substantial variation in the quality of different MA plans operating in the same county. To gauge the extent of selection bias in these estimates, we develop an instrumental variables procedure that asks whether observational mortality-added accurately predicts the mortality effects of quasi-experimental plan terminations. We find that consumers who are forced to switch from high- (low-)mortality plans to more typical alternatives face a reduced (increased) mortality risk, of the same magnitude predicted by the observational measure. We use these validated mortality-added estimates to investigate the underlying drivers of MA plan effects and assess whether consumer demand accounts for this important dimension of plan heterogeneity.

PRELIMINARY AND INCOMPLETE; DO NOT DISTRIBUTE OR CITE.

*Abaluck: jason.abaluck@yale.edu, Yale University School of Management and NBER; Caceres: mauricio.caceresbravo@yale.edu, 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 ...

1 Introduction

In most markets, firms choose not only the prices of products but also product quality. A large theoretical literature shows how competitive pressures can lead to quality improvements that are valued by consumers. But if consumers cannot observe or do not value product quality, competition alone may fail to foster quality. Muted incentives for quality provision are especially possible in health care markets, where imperfect or asymmetric information has long been viewed as a source of inefficiency (Arrow, 1963). The welfare consequences of low or varied quality health care provision may also be especially large.

To leverage competitive pressures in health care markets, policymakers have increasingly relied on public quality measures of providers, such as doctors and hospitals. Recent research has shown both that these measures predict true provider health effects and that patients tend to seek out higher-ranked providers. At the same time, the past decade has seen an increased reliance on competitive health insurance markets, with an extensive literature investigating whether consumers are financially prudent in selecting between different insurers or plans. To date, no empirical evidence exists for whether consumers can distinguish between plans of different quality, and thus the scope for health insurance competition to improve consumer health.

This paper estimates the effects of health plan choice on beneficiary mortality, investigates why some plans have better causal mortality effects, and assesses whether consumer demand responds to this dimension of plan heterogeneity. Measuring quality is difficult: observational estimates may reflect unobserved consumer sorting and quasi-experimental designs are likely under-powered to detect health effects. The existing literature develops tools to combine observational and quasi-experimental techniques to develop efficient and unbiased estimates of *provider* quality. We extend these tools and clarify the underlying theoretical assumptions necessary to estimate the causal effect of alternative health plans. Additionally, we relate our estimates to provider-level estimates using hand-collected data on plan networks.

Our setting is the Medicare Advantage (MA) market, in which individuals choose from a broad array of private managed care plans partly subsidized by the government. The MA program is large and growing, covering one in three Medicare beneficiaries in 2017. Its model of subsidized private provision is also highly influential, serving, for example, as the forerunner of the Affordable Care Act exchanges.

We begin by documenting large differences in one-year MA plan mortality rates, adjusting for observable differences in plan enrollees. Specifically, these observational *mortality-added* models suggest that a one standard deviation decrease in plan mortality-added is associated with a 0.7 percentage point change in beneficiary mortality, on a baseline one-year mortality rate of 3.8%. This represents a 19% reduction in mortality, comparable to observational estimates of institutional

quality dispersion in other healthcare settings (e.g. Hull [2018]). Plans with lower observational mortality-added also tend to be highly ranked by other quality measures, such as Medicare Star Ratings.

Observational mortality-added estimates may reflect unobserved beneficiary sorting as well as true plan health effects. To gauge the extent of such selection bias in our setting, we leverage quasi-experimental variation in choice sets arising from plausibly exogenous MA plan terminations. Intuitively, when plans with high or low observational mortality exit from a market ("terminate"), their enrollees tend to move into new plans that are more typical in terms of the observed mortality rate. Beneficiaries in non-terminated plans, in contrast, tend to be highly inertial and thus remain in high- or low-mortality plans. If the observational estimates reflect true mortality-added, we would thus expect cohort mortality to decline (rise) when high- (low-)mortality plans exogenously exit the market, relative to similar markets where no terminations occur.

We formalize this intuitive approach to validating observational mortality-added with a novel instrumental variables (IV) framework. Our main parameter of interest is the mortality-added *forecast coefficient*, defined by the infeasible plan-level regression of true plan health effects on observational mortality-added. We show how a feasible beneficiary-level IV regression of mortality on enrolled plan mortality-added, instrumented by the interaction of the mortality-added of one's plan in a previous year and an indicator for whether that plan terminated, identifies this forecast coefficient under two conditions. First, we require that terminations are as-good-as-randomly assigned to plans and that terminations of high- and low-mortality plans have at most a comparable direct effect on mortality, along with the indirect effect of changing subsequent plan enrollment. This standard *exclusion restriction* is sufficient for testing the null hypothesis of no selection bias – that observational mortality-added perfectly captures true plan mortality-added – but not for estimating the plan forecast coefficient when such bias is present. For the latter, we derive a new *fallback condition.* Intuitively, this condition states that the "fallback" plans which beneficiaries would switch into, when forced by plan termination to make an active choice, are similar to those they chose initially in terms of the unforecastable component of plan mortality-added. We derive microfoundations for this condition in a discrete choice model and develop empirical validations based on observable differences in fallback plans.

IV estimates in the MA setting reveal that observational mortality-added is a highly reliable predictor of the true effect of MA plan choice. Namely, across a variety of different specifications – including those leveraging quasi-experimental policy variation in the termination of Private Fee-For-Service plans in (Pelech, 2018) – we find reduced-form effects of terminations on beneficiary mortality which are nearly perfectly predicted by strong first-stage effects on plan mortality-added. This generates forecast coefficient estimates that are statistically indistinguishable from one, suggesting that on average a consumer selecting a one percentage point lower mortality-added plan

reduces her expected one-year mortality by one percentage point. At the same time, in straightforward extensions of our IV approach we find significant residual variation in the projection of true plan health effects on observational mortality-added. We use these estimates to construct improved empirical Bayes predictions of MA plan mortality-added, combining the observational and quasi-experimental variation in plan choice that optimally trade off precision and selection bias.

We use our hybrid estimates of plan health effects to investigate a number of policy-relevant questions. First, why do some plans make beneficiaries healthier than others? To understand mechanisms, we leverage both administrative Medicare data and hand collected network data to construct "inclusive" measures of plan mortality-added derived from mortality-added estimates of hospitals in each plan's network. Hospital estimates are obtained from samples of representative fee-for-service beneficiaries (for whom there is no network variation), using conventional risk-adjustment methods that Hull (2018) and Doyle et al. (forthcoming) find are reliable measures of hospital quality. We then aggregate hospital mortality-added estimates to the network level, using market share weights. Correlating these inclusive measures with our plan mortality-added posteriors shows that hospital networks drive much, but not all, of the variation in plan mortality-added. This suggests that insurers are both informed buyers of health care services and good agents for consumers.

We then turn to the question of whether consumers indeed value plan mortality-added differences. Given conventional estimates of the value of a statistical life year (VSLY), our estimates suggest consumers should be willing to pay tens or even hundreds of thousands of dollars per year to access better plans. Yet in both hedonic regressions and a formal plan choice model we find consumers are much less sensitive, with an estimated VSLY falling well below the typical range. However, we find that publicly available quality measures are coarse, but weakly correlated with plan mortality-added.

In assessing both the distribution of MA plan health effects and its implications for policy, our paper relates to several distinct literatures. Most broadly, we study effects of managed care, as in Curto et al. [2015] and Dranove et al. [2017], though this literature typically focuses on utilization and costs rather than beneficiary health effects. An important exception is Duggan et al. [2015], who find that MA plan terminations in counties with only a single MA plan lead to increased hospitalizations, but lack power to detect consequential impacts on consumer mortality and do not study effects of plan choice within the MA marketplace. Further afield, Geruso et al. [2017] study random assignment of low-income beneficiaries to alternative Medicaid Managed Care plans. They find large differences in spending for the same beneficiaries and more switching from low spending to high spending plans than vice-versa, but again lack power to detect even large mortality differences.

In studying consumer attentiveness to health effects, we add to a large literature on behavioral

biases in insurance plan choice (Abaluck and Gruber 2016, 2011; Ericson and Starc 2016, Handel 2013, Handel and Kolstad 2015). These results further indicate that potentially suboptimal consumer choices impact the characteristics of plans offered (Starc and Town 2018). Again, however, this literature has to date largely focused on the financial aspects of insurance plans, leaving unstudied the critical question of whether some plans make consumers healthier than others, and whether consumer choice at least partially accounts for these differences.[1]

Lastly, we add to a recent literature using quasi-experimental methods to estimate the health effects of choosing different hospitals, doctors, nurses, and medical service areas (Hull 2018, Fletcher et al. 2014, Yakusheva et al. 2014, Finkelstein et al. 2017). This work builds on methodological advances in mortality-added estimation in other fields such as education (Chetty et al., 2014b, Angrist et al. 2016, 2017), and we are the first to apply these methods to measure the mortality-added of individual insurance plans. Several papers in this literature estimate IV regressions of individual outcomes on observational predictions of institutional quality to test for forecast bias (Chetty et al. 2014a; Doyle et al. 2017). A methodological contribution of our paper is to clarify the assumptions under which such regressions may recover the structural relationship between observational and true quality (and thus measure the extent of forecast bias, when non-zero), and how such forecast coefficient estimates can be used to combine observational and quasi-experimental variation, following Angrist et al. [2017], Hull [2018], and Chetty and Hendren [2017].

The remainder of this paper is organized as follows. The next section describes the institutional setting and data, documents large variation in observational mortality-added across MA plans, and motivates our quasi-experimental validation with plan terminations. Section 3 then develops a formal econometric framework for IV identification of plan forecast coefficients and related parameters. Section 4 presents our main estimates, and Section 6 concludes.

2 Setting and Data

2.1 Medicare Advantage

The Medicare program, established in 1965, primarily covers Americans 65 and older. Parts A and B – typically referred to as "traditional Medicare" (TM) – cover hospitalizations and physician services, respectively. The Centers for Medicare and Medicaid Services (CMS) administers TM. While most beneficiaries receive coverage through TM, private insurance options are also available. The parallel private program has gone by a variety of names over time (see McGuire et al.

[1] Gaynor et al. (2016) finds that hospitals improve quality of care when they face competitive pressures to do so, while Chandra et al. (2014) document evidence that patients respond to public hospital quality rankings.

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

Private MA insurers covers medical spending and receive a capitated monthly payment. MA plans must provide all the insurance benefits of TM. Competitive plans also offer lower cost sharing and supplemental benefits to attract consumers. For MA enrollees, there is typically a trade-off. MA plans restrict access to providers by forming networks, similar to commercial HMOs, but offer generous financial coverage. MA's popularity generally coincides with the level of federal reimbursement and it has grown over time. As of 2018, 34% of Medicare beneficiaries enrolled in a Medicare Advantage plan . Enrollment rates have continued to grow post-Affordable Care Act (ACA). There is also significant geographic heterogeneity in the popularity of MA plans, though the typical beneficiary chooses between TM and (often, many) Medicare Advantage plans available in their county. Across consumers within a market, MA may be more attractive to middle class retirees or consumers with lower risk.

Most plans in the MA program are managed care plans, in which insurers employ supply side controls and restrict access to providers. However, we also observe private-fee-for-service (PFFS) plans, which traditionally offer broader provider access than other MA plans. For some of our results, we specifically focus on terminations in PFFS plans arising from a plausibly-exogenous policy change. PFFS plans grew in the mid-2000s as a result of generous reimbursement. In 2008, when PFFS plans represented 28% of the MA market, the Medicare Improvements for Patients and Providers Act required all PFFS plans to form new provider networks, significantly increasing their fixed costs. Pelech (2018) documents significant plan termination is the year following the policy; by 2011, PFFS plans represented only 11% of the market.

The MA program has always been controversial. Cherry-picking by MA plans could lead to over payment by the federal government or skew benefit design to attract favorable risks (Brown et al. 2003). By contrast, there is the potential for better medical management under MA (Starc and Town 2019). Despite potential efficiency gains, a substantial portion of the private (financial) gains from the MA program likely accrue to insurers (see Cabral et al. (2018) and Duggan et al. (2016)).

2.2 Data and Summary Statistics

Our primary data source is plan enrollment and mortality for the universe of Medicare beneficiaries. We also observe geographic information and demographics (age, sex and race). For TM enrollees, we also observe inpatient claims. For 20% of all enrollees (MA or TM), we observe prescription drug claims. In our main analysis sample, we restrict to beneficiaries aged 65 and over who were ever enrolled in an MA plan in the 50 US states or DC from 2006 to 2011 , have non-missing enrollment information during those years, and did not change zipcodes. The unit of observation

is a beneficiary-year.

Summary statistics are presented in Table 1. During our time period, 21% of beneficiaries enrolled in a Medicare Advantage plan. Medicare Advantage plans differ in their observational mortality rates, as discussed below. Table 1 shows that enrollees in higher mortality plans tend to be older and are more likely to be white and female on average, consistent with selection on observables. We also report the percentage of enrollees in terminated plans. We define a plan termination as the last year a plan appears in the landscape file if it was not consolidated. Interestingly, lower observational mortality plans are somewhat more likely to terminate.

Table 1: Summary Statistics

% in MA			21.2		
	TM	MA Q1	MA Q2	MA Q3	MA Q4
Age	78.9	77.0	77.6	78.2	79.5
% Female	59.1	56.0	57.3	59.5	62.7
% White	87.2	80.1	79.9	83.2	84.3
% in Terminated Plan	0.0	4.4	1.7	3.5	2.9
Observations	18,769,565	1,450,424	1,446,808	1,450,114	1,444,767

Notes: Table presents summary statistics describing mean enrollee demographics, coverage, and consumption. The unit of observation is the enrollee-year.

We leverage additional data to investigate the mechanisms by which some plans make enrollees healthier. Specifically, we combine the MA data with hand-collected data on plan networks. We use Medicare claims data for TM beneficiaries to measure the mortality-added of hospitals and insurance plans for different types of beneficiaries. These auxiliary data sources allow us to characterize how inputs such as plan expenditures or physician and hospital networks relate to plan mortality-added.

2.3 Observational mortality-added

In this section, we describe variation in one-year mortaility rates across MA plans. Simply put, we observe substantial variation even after conditioning on demographics and county-year fixed effects: the beneficiary-weighted standard deviation of mortality across plans is 1.7 percentage points, or 31% of the mean.Of course, the raw mortality measure could reflect numerous factors.

First, the variation could reflect statistical noise arising from plans with few beneficiaries. Second, the variation could reflect differential selection by consumers of different latent mortality. Third, the variation could reflect differences in the causal effect of plan networks or other plan features on consumer health.

To account for statistical noise, we apply conventional empirical Bayes methods that "shrink" all plan-level estimates of mortality-added towards their contract and county-level means, as described in the Appendix.[2] However, differences in observational mortality could still reflect the fact that some plans are chosen by sicker beneficiaries than others. We first control for observable beneficiary characteristics to partially address this concern. Let Y_i denote individual i's mortality outcome and let $D_{ij} = 1$ if individual i is observed in plan j and 0 otherwise. We estimate:

$$Y_i = \sum_{j=1} \mu_j D_{ij} + X_i \psi + \varepsilon_{i,} \tag{1}$$

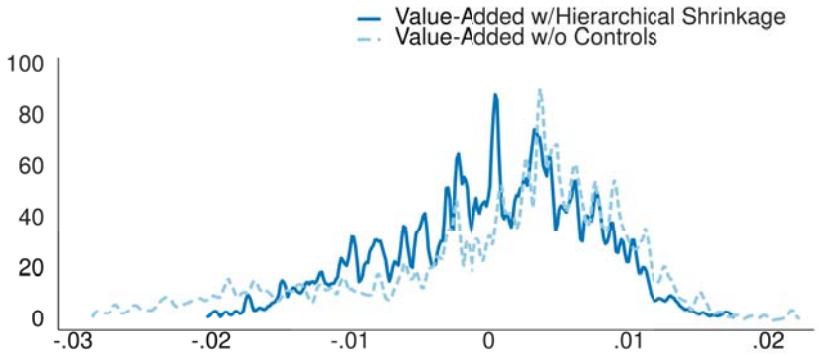
where the control vector X_i includes controls for age (in years), race, sex, year and county (location) fixed effects. When we use the phrase "observational mortality," we are referring to our shrunken estimates of μ_j in equation 1.

Figure 1 plots the distribution of observational mortality. We demean mortaility at the county-year level; as a result, we are describing variation within a given choice set. The solid line in Figure 1 shows this distribution for Medicare Advantage plans. The variation across plans in shrunken mortality rates remains substantial even after applying empirical Bayes shrinkage: a one standard deviation reduction in mortality rates is 1.0 percentage points, or an 18% reduction in annual mortality. We find evidence of selection on observables: adding age, race, and sex controls reduces the standard deviation of shrunken mortality rates by 30%, from 1.0 percentage points to 0.7 percentage points. Nonetheless, substantial variation remains.[3] However, there is obviously additional scope for selection due to differences in health not captured by drug expenditures. In the following subsection, we provide suggestive evidence that $\mu_j$reflects the causal impact of plan enrollment.

[2]Specifically, we estimate a hierarchical linear model which allows for contract-level clustering of plan estimates.

[3]Even when we add flexible controls for prescription drug consumption, the standard deviation falls only slightly. This evidence is suggestive that the differences across plans may be causal.

Figure 1: Observational Mortality

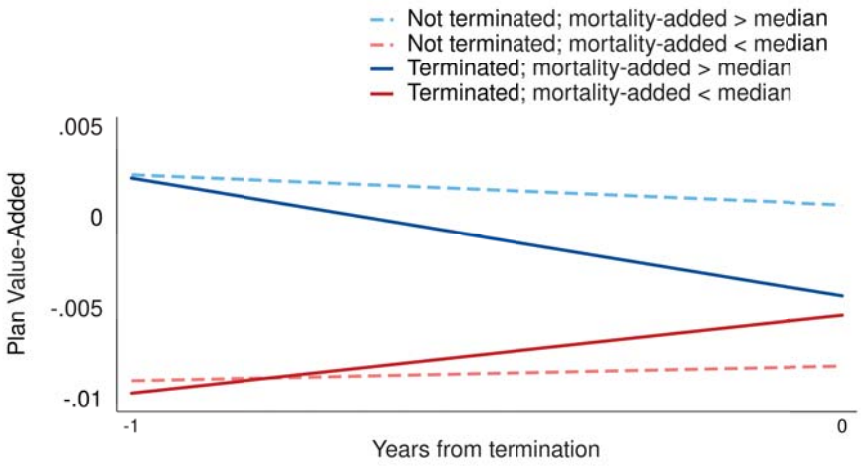


2.4 Plan Terminations

The observational mortality estimates only address selection on observerables. However, we can validate our observational measures using quasi-experimental variation by leveraging variation in choice sets over time. Intuitively, if a "bad" (high mortality-added) plan exits a market, enrollees in that plan will likely move to better plans and see their mortality fall. Conversely, if a "good" plan exits a market, enrollees in that plan should see their mortality rise.

Figure 2 suggests that plan terminations steer enrollees to plans with very different observational mortality rates. The figure plots the change in the observational mortality of the plans in which individuals are enrolled for various cohorts. The dotted lines plot the change in plan observational mortality-added for beneficiaries enrolled in plans which are above-median (dashed blue) and below-median (dashed red) in year t-1. Most beneficiaries are inertial, meaning that they remain in the same plans and thus are enrolled in a plan with the same observational mortality (defined timelessly). The solid lines plot plan observational mortality for beneficiaries enrolled in plans which terminated. Beneficiaries enrolled in plans that terminate (necessarily) re-sort. Beneficiaries in above-median observational mortality (solid blue) plans move into average (lower mortality) plans following termination (and vice versa). Following terminations, everyone enrolls in observationally similar plans regardless of previous enrollment . Relative to other enrollees in above-median plans in year t-1, beneficiaries in terminated plans are enrolled in lower mortality plans. Relative to other enrollees in below-median plans in year t-1, beneficiaries in terminated plans are enrolled in higher mortality plans.

Figure 2: Plan Terminations and Subsequent Plan mortality-added

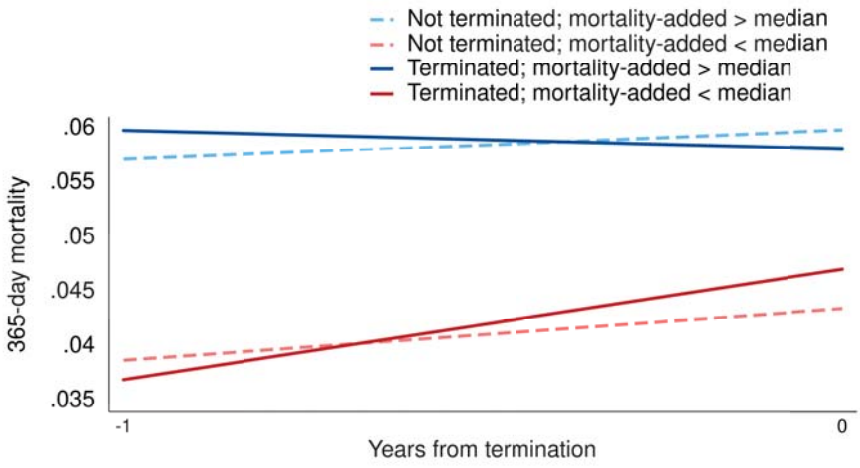


Does this change in the observational mortality of the plan in which an individual is enrolled lead to a change in realized mortality? Figure 3 suggests that it does. In this figure, the dotted lines show realized mortality rates in each year for the cohort of beneficiaries enrolled in each type of plan in period t-1. However, beneficiaries enrolled in plans which do not terminate in year t-1 have higher mortality in year t: this is to be expected, as they are one year older. Beneficiaries enrolled in an above median observational mortality plan that terminates see a *decrease* in mortality.[4] By contrast, beneficiaries enrolled in below median observational mortality plans see mortality increase more than expected due to aging. The results suggest that being reassigned to plans with different observational mortalities from one's current plan has a causal effect on mortality, and in the direction we would predict given the measured observational mortality. In the next section, we formalize the interpretation of this slope in terms of the reliability of observational mortality-added in predicting true plan health effects.[5]

[4]These beneficiaries are also one year older. However, the effect of being reassigned to an average plan (rather than a "bad" plan) outweighs the effect of aging.

[5]Figure 9 shows the relationship between Figure 2 and Figure 3 for additional categories of mortality-added (deciles and vigntiles). On the x-axis, we plot the predicted change in observational mortality given the year t-1 observational mortality of the chosen plan (this is large if year t-1 mortality is small and small if year t-1 mortality is large). On the y-axis, we plot the realized change in mortality for beneficiaries in each category. Beneficiaries who are predicted to experience a larger change in mortality based on observational mortality rates do in fact see a larger change in mortality. Interestingly, the points in Figure 9 follow fairly closely a line through the origin with a slope of around one, quite similar to the ratio of reduced form and first stage effects in Figures 2 and 3.

Figure 3: Plan Terminations and Subsequent Mortality



3 Econometric Framework

We now formalize how plan terminations can recover the predictive relationship between observational mortality-added and causal mortality-added in an IV framework. We first outline the setting and target forecast coefficient parameter, before providing two conditions (one novel) under which this parameter is identified by a particular IV regression.

3.1 Plan Health Effects

Let $Y_{ijm} \in \{0,1\}$ denote the potential mortality outcome if individual i in market m were to enroll in plan j, and let

$$\beta_{jm} = E[Y_{ijm}] \tag{2}$$

denote average potential plan mortality. That is, β_{jm} captures the expected mortality if all beneficiaries in individual $i's$ market (a particular county) were to enroll in plan j. In practice, we measure β_{jm} relative to the county mean; for notational simplicity, we supress market subscripts and write demeaned mortality added as β_j. To further simplify, we for now abstract away from treatment effect heterogeneity and assume the unobservable component of potential health outcomes is additively separable into a component due to plans and a component due to patients: $Y_{ij} = X_i'\gamma + \beta_j + \varepsilon_i$, where X_i is a vector of observables (which includes a constant) with $E[X_i\varepsilon_i] = 0$ by definition of

γ.[6]

Consumers choose among the set of available plans in their local market, with $D_{ij} = 1$ indicating that individual i enrolls in plan j.[7] Individual mortality is then given by$Y_i = \sum_j Y_{ij} D_{ij}$. Substituting in the previous expression for Y_{ij}, observed outcomes can be written

$$Y_i = \sum_j \beta_j D_{ij} + X_i'\gamma + \varepsilon_i. \tag{3}$$

Equation (3) is a causal model linking beneficiary plan choice D_{ij} to mortality Y_i.

Nonrandom consumer selection creates a fundamental econometric challenge: to the extent that plan j attracts consumers with poor (good) unobserved health, its observed mortality rate will be an upward- (downward-)biased estimate of β_j. For this reason, the OLS parameters μ_j need not equal the causal parameters β_j. We next discuss our approach to plan mortality-added estimation in light of this challenge.

3.2 The Forecast Coefficient

In principle, we could estimate the full set of coefficients in (3) with a set of instrumental variables that affect plan choice D_{ij} while remaining uncorrelated with the unobserved determinants of mortality ε_i. In practice, however, we are unlikely to find enough instruments for such an IV procedure. To overcome this challenge, we instead attempt to identify the relationship between observational mortality-added μ_j and true plan mortality-added β_j in the population of plans.

Our principal object of interest is the plan *forecast coefficient* λ, defined by the projection of β_j on μ_j:

$$\beta_j = \gamma_0 + \lambda \mu_j + \eta_j, \tag{4}$$

where $E[\eta_j] = E[\mu_j \eta_j] = 0$ by definition of λ. Here, λ is the coefficient we would recover if we observed the true β_j (we do not) and regressed it on mortality-added predictions μ_j. In a linear model, this object allows us to best use all of the information in observational mortality to predict true treatment effects. Given λ, we can adjust our mortality-added estimates for selection and conduct counterfactual simulations of mortality if beneficiaries were assigned to different plans.

We refer to the residual from this regression η_j as the plan *forecast error*, reflecting the fact that for a given observational mortality α_j some plans increase mortality by more or less than expected.

[6] In practice, the vector X_i can be different from the one described in equation 1. It is straightforward to conduct our analysis separately for individuals with different subsets of observable characteristics in order to recover a separate β_j. Below, we extend our model to allow for unobservable heterogeneity in treatment effects.

[7] We suppress time subscripts here except when describing consumer choice over time.

Substituting equation (4), into equation (3), we obtain

$$Y_i = \lambda \mu_j + X_i \gamma + \varepsilon_i + \eta_i, \tag{5}$$

where $\mu_i = \sum_j \mu_j D_{ij}$ denotes the predicted mortality of beneficiary i given her plan choice and $\eta_i = \sum_j \eta_j D_{ij}$ is the corresponding component of plan mortality-added that cannot be forecasted.[8] To estimate λ, we will use instrumental variables constructed from plan terminations.

If μ_j is an unbiased predictor of β_j (that is, if $\lambda = 1$), then when plan terminations induce beneficiaries to change plans, mortality should change by exactly the amount we would predict given μ_j. In other words, suppose plans A and B have observational mortality rates of 5% (conditional on observables). Suppose A terminates, and all of its beneficiaries move to plan C, with an observational mortality rate of 3%. If λ =1, we would then predict that annual mortality would fall by 2% for the cohort from plan A relative to those in plan B. If $\lambda = 1/2$, mortality would instead fall by 1%, as the observed difference between the plans would be partly due to selection. We next formalize this intuition.

3.3 Identification

As described above, our primary instrument is constructed as the product of lagged plan termination indicators T_i^{t-1} and lagged plan mortality-added predictions μ_i^{t-1}:

$$Z_i \quad = \quad \mu_i^{t-1} T_i^{t-1}. \tag{6}$$

We first derive conditions for this IV procedure to identify λ in a simplified setting without treatment effect heterogeneity and controlling only for a constant, lagged observational mortality-added μ_i^{t-1}, and a lagged termination indicator T_i^{t-1}.[9] Let $\tilde{Z}_i$ denote Z_i after partialling out observable controls.

The IV coefficient $\hat{\lambda}$ consistently estimates λ given a first stage (that $\tilde{Z}_i$ and μ_i are asymptotically correlated) and exclusion restriction (that $\tilde{Z}_i$ and $\varepsilon_i + \eta_i$ are asymptotically uncorrelated). The first-stage condition is highly intuitive. When plan choice is inertial, we expect individuals to largely remain in their previous year's plan absent a termination. However, individuals forced to make an active choice by a termination tend to revert to more typical plans, so that terminations conditional on lagged observational mortality-added predict current observational mortality-added. Under this regression to the mean, which we previously documented in Figure 2, the interaction

[8] As noted above, the X_i described in equation 5 can, for example, include lagged planned characteristics. However, if the vector is the same in both equations, OLS estimation of equation of equation (5) will mechanically give a coefficient of 1 rather than recovering λ.

[9] We extend the model to allow for heterogeneous treatment effects in Section 3.4.

coefficient is likely to be negative in large samples; we more formally motivate this condition with a behavioral model of plan choice below. Inspection of equation 5 suggests that the usual exclusion restriction now consists of two parts: $Cov(\tilde{Z}_i, \varepsilon_i) = 0$ and $Cov(\tilde{Z}_i, \eta_i) = 0$. We discuss each of these in turn.

3.3.1 Quasi-random Assignment of Plan Terminations: $Cov(\tilde{Z}_i, \varepsilon_i) = 0$

Ass-good-as-random assignment of terminations to plans would suffice for $Cov(\tilde{Z}_i, \varepsilon_i) = 0$. However, this is a stronger assumption than we strictly require. In Appendix E, we show that the condition $Cov(\tilde{Z}_i, \varepsilon_i) = 0$ is equivalent to an (infeasible) plan level difference-in-differences regression:

$$\bar{\varepsilon}_j \quad = \kappa + \omega T_j + \phi \mu_j + \psi \mu_j T_j + e_j, \tag{7}$$

where $\bar{\varepsilon}_j = \frac{\sum_i D_{ij}^{t-1} \varepsilon_i}{\sum_i D_{ij}^{t-1}}$ denotes the average ε_i among individuals formerly enrolled in plan j, and estimation of equation (7) is weighted by lagged market share $\frac{1}{N} \sum_i D_{ij}^{t-1}$. We note that $Cov(\tilde{Z}_i, \varepsilon_i) = 0$ if and only if $\psi = 0$ in this regression.

When terminations are as-good-as-randomly assigned, ψ will be asymptotically zero: for any lagged mortality, terminated and non-terminated plans will have similar enrollees. Even if terminations are not randomly assigned, ψ may still be zero. In this case, the model requires the difference in unobservables between terminated and non-terminated plans to not vary with lagged observational mortality-added. That is, if terminated plans are different, they are not differentially different across ex ante high and low observational mortality-added plans.[10]

We can indirectly test the exclusion restriction empirically in several ways: we can test for balance in observable consumer characteristics across terminated and non-terminated plans. We can also test the necessary conditions more directly by estimating equation (5), replacing mortality with predicted mortality given pre-determined consumer characteristics such as age and gender. Finally, the intuition above suggests that $Cov(Z_i, \varepsilon_i)$ can be tested by constructing pre-trends.[11] We conduct all of these tests below.

[10] Suppose, for example, that among low observational mortality plans, terminations occur because population health appears to be systematically worsening while among high observational mortality plans terminations occur due to exogenous financial shocks. We might wrongly conclude that a relative decline in health among cohorts in terminated, low mortality plans was due to those individuals being reassigned to medium mortality plans rather than the fact that health was worsening among that population anyway.

[11] In our setting, this means graphing mortality over time for cohorts of beneficiaries enrolled in terminated and non-terminated plans several years before terminations.

3.3.2 The Fallback Condition: $Cov(\tilde{Z}_i, \eta_i) = 0$

Recall that η_i is an attribute of the plans that individuals choose following terminations; thus, random assignment of plan terminations is not sufficient for $Cov(\tilde{Z}_i, \eta_i)$ because consumers in terminated plans may subsequently make systematically different choices. We thus require a second condition which states that these choices are "typical" in a sense we make precise below. We refer to this novel component of IV exclusion, $Cov(Z_i, \eta_i) = 0$, as the *fallback condition*. We consider this condition under a mild generalization of the previous discussion in which we flexibly control for a vector of lagged plan characteristics $W_i^{t-1} = \sum_j D_{ij}^{t-1} W_j$ in the IV regression producing $\hat{\lambda}$ along with μ_i^{t-1} and T_i^{t-1}. Appendix F shows that in this more general model the fallback condition is satisfied if, given an appropriate law of large numbers,

$$E\left[\mu_j \left(\sum_k \eta_k \left(p^1_{j\to k} - p^0_{j\to k}\right)\right) | W_j\right] \quad = \quad 0, \tag{8}$$

where $p^t_{j\to k}$ denotes the probability of choosing plan k given that you were enrolled in plan j with a termination status of $t \in \{0,1\}$. This condition says that we cannot predict a consumer's change in forecast error η_j given the observational mortality of the chosen plan and termination status. The condition could be violated if, for example, people in terminated plans learned what makes plans effective at reducing mortality and chose accordingly. In this case, conditional on choosing plans with high observational mortality-added, consumers may disproportionately choose plans with low η_j, leading to low actual mortality.

The formalization of the fallback condition is novel to our paper and warrants further interpretation. Specifically, η_j is not a structural objects. These are the error term in a regression of (unobserved) mortality-added on observational mortality, which itself depends on consumer choices. Thus, it is not obvious what conditions on underlying choices would lead the fallback condition to be satisfied.

To clarify these issues, we provide a sufficient discrete choice micro-foundation. To build intuition, we assume that consumers in non-terminated plans are "fully inertial" and remain in their current plan – an assumption we relax in Appendix F. Consumers have preferences over true plan mortality-added β_j and other (observed and unobserved) plan characteristics W_{jt} and ξ_{jt}. Consumer $i's$ utility for plan j at time t is given by:

$$U_{ijt} = \alpha_i h(X_{it}, W_{jt}) + f_i(\beta_j) + \nu_i \xi_j + u_{ijt}, \tag{9}$$

where ξ_j (unobserved) is the time-specific deviation from the mean valuation, ν_i is an individual-

specific unobservable (potentially correlated with ε_i), and u_{ijt} is an additive error.[12] The model allows for heterogeneous preferences over mortality-added β_j and other plan characteristics, X_{it} and W_{jt}.[13] Tt follows immediately from equation (8) that the fallback condition will be satisfied with no persistent unobservable heterogeneity in plan choices. In this case (i.e. if $\nu_i = 0$ and either $\alpha_i = \bar{\alpha}$or h is not a function of W_{jt}), $p_{j\to k}(X) = p_k$ and the term in parentheses is not a function of j. More subtlety, the same argument accommodates heterogeneous preferences over observable plan attributes (i.e. allowing for heterogeneous α_i and allowing h to be a function of W_{jt}). In this case, we need to condition on the observable plan attributes W_{jt}.

Importantly, we can also accommodate monotone persistent unobservables. That is, we can allow for the possibility that heterogeneous consumers place different weight on the unobservable plan attribute ξ_j. With this utility function, observable-adjusted mortality rates will give biased estimates of β_j because unobservably sicker (or healthier) beneficiaries will prefer certain types of plans (parametrized by ξ_j). Appendix F shows that the fallback condition will nonetheless be satisfied provided we condition on lagged market shares. In this case, the fallback parameter $\lambda = \frac{Cov(\beta_j, \beta_j + b_j)}{Var(\beta_j + b_j)} = \frac{\sigma_\beta^2 + \sigma_{\beta b}}{\sigma_\beta^2 + \sigma_b^2 + 2\sigma_{\beta b}}$. In terms of utility parameters, the forecast coefficient will be close to 1 if most of the variation in μ_j comes from variation in underlying treatment effects β_j, it will be zero if most of the variance comes from variation in bias (parametrized by ξ_j), and it will also depend on the covariance of β_j and ξ_j.

Appendix F also provides additional examples of choice models for which the fallback condition is satisfied (and associated proofs). Intuitively, these models also describe the consumer behavior that drives selection bias. Given unobserved heterogeneity in preferences over unobservable plan characteristics, the correlation between ε_i and v_i is likely non-zero. The relationship between λ and this correlation is complex (and not necessarily monotone). Therefore, we estimate λ directly.

In addition to micro-founding the fallback condition, we can investigate it empirically. The fallback condition asks whether the unobservable forecast error is correlated with our instrument. While we do not observe this unobservable forecast error from regressing β_j on μ_j, we can construct the observable analogue by regressing observed plan characteristics on μ_j, computing the residual and then assessing whether that residual is uncorrelated with our instrument. We conduct this test below.

[12] The idiosyncratic error term needs to be distributed the same across plan choices j but not necessarily consumers. In fact, this model is quite flexible relative to what is normally considered in the discrete choice literature (and applied examples). For example, we need make distributional assumptions on the error term.

[13] The model requires that a random coefficient on mortality-added be "single-signed" – all consumers must weakly prefer lower (or higher) mortality plans.

3.4 Extensions

Before applying the above theoretical framework, we note three straightforward extensions. First, we can relax the assumption of no direct disruption effects from terminations by including a termination main effect, T_i^{t-1} in the control vector W_i. Intuitively, as in the difference-in-differences Figures 2 and 3, the IV regression compares termination effects for individuals previously enrolled in plans of different observational mortality, allowing for a common direct effect. We can further interact the termination main effect by individual demographics, such as age, gender, or pre-termination utilization, to allow this effect to vary by these observables.

Second, we note that while the above exposition is made clearer from assuming terminations are as-good-as-randomly assigned across different plans, in practice the IV and control construction allow the termination probability to depend on a plan's observational mortality μ_j. By including other lagged plan characteristics as controls, we can further weaken the quasi-experimental assignment assumption to hold only within observably-similar plans.

Finally, in the Appendix, we show how the framework can accommodate unobservable selection on treatment effects. Our argument proceeds similarly, but the fallback condition becomes slightly stronger, now requiring that any difference in the degree of selection on treatment effects between terminated and non-terminated plans is unrelated to plan mortality-added.

4 Results

In this section, we apply the results developed above in order to estimate the relationship between observational mortality and true mortality-added. Intuitively, we will assess the degree to which observational measures predict the change in mortality observed following plan terminations.

4.1 First Stage and Reduced Form Estimates

We first show that changes to consumers' choice sets lead to changes in the observational mortality of the plan in which consumers are enrolled. We predict contemporaneous observational mortality-added μ_i as a function of a termination dummy interacted with lagged observational mortality-added μ_i^{t-1}. Following the logic of Figure 2, formally, we estimate:

$$\mu_i = \pi_Z Z_i + g\left(\mu_i^{t-1}\right) + \pi_T T_i^{t-1} + X_i \pi_X + \varepsilon_i + \eta_i. \tag{10}$$

The excluded instruments $Z_i = g\left(\mu_i^{t-1}\right) T_i^{t-1}$ are a termination dummy interacted with a function of lagged observational mortality-added μ_i^{t-1}. We control for both the direct effect of termination $\pi_T T_i^{t-1}$ and lagged mortality-added, $g\left(\mu_i^{t-1}\right)$; we also control for demographics and lagged plan

enrollment in $X_i\pi_X$.

We parameterize the excluded instrument in a number of ways. First, we parameterize $g\left(\mu_i^{t-1}\right)$ linearly, such that the excluded instrument is the termination dummy interacted with lagged observational mortality-added μ_i^{t-1}. Second, we parameterize $g\left(\mu_i^{t-1}\right)$ a dummy for above or below median lagged mortality-added (corresponding exactly to our plots). Third, we include deciles of lagged observational mortality-added. The results are in Table 2. Columns (1) and (3) of 2 present linear specifications, while in columns (2) and (4), the excluded instrument is the termination dummy interacted with a dummy for above median lagged mortality-added. In Appendix Table 4, we present a specification in which the excluded instrument is the termination dummy interacted with a dummies for each decile of lagged mortality-added. In all specifications, we also include county-year fixed effects. The inclusion of these fixed effects implies that all comparisons are within markets, rather than across markets. Columns (3) and (4) include additional demographic controls.

The first stage coefficient estimates are consistent with regression to the mean (as hoped): beneficiaries enrolled in high mortality-added plans in year $t-1$ tend to choose plans which are more typical in terms of observational mortality following terminations. When a high mortality plan disappears, individuals move to a plan which on average has lower mortality rates and mortality rates fall for the cohort induced to move by the termination. To see this in column (3), note that terminations reduce the mortality-added of the enrolled plan by 0.37 percentage points for each additional percentage point of lagged mortality-added. Column (4) shows that the termination of an above median mortality-added plan in year $t-1$ reduces the mortality-added of the enrolled plan by 0.005 percentage points. The reduced form coefficient in the second row describes how terminations in year $t-1$ affect contemporaneous mortality. As we would predict, the termination of an above median observational mortality-added plan reduces contemporaneous mortality-added by 0.004 percentage points.

Table 2: First Stage

	(1)	(2)	(3)	(4)
First Stage	-0.364	-0.005	-0.365	-0.005
	[0.028]	[0.001]	[0.028]	[0.001]
Reduced Form	-0.376	-0.004	-0.341	-0.004
	[0.099]	[0.001]	[0.091]	[0.001]
Specification	Lagged μ_i^{t-1}	Lagged Median	Lagged μ_i^{t-1}	Lagged Median
Demographic Controls	No	No	Yes	Yes
First Stage F-Stat	164.8	110.2	167.0	112.4
N	16,135,957			

Notes: The first row presents results from the first stage regression of contemporaneous μ_i on the instrument. The second row reports the analagous reduced-form regression. Columns (1) and (2) do not include demographic controls. Columns (3) and (4) flexibly control for age and include race and sex dummies.

So far, we have shown that following plan terminations, beneficiaries in high observational mortality plans end up in plans with lower observational mortality and beneficiaries in low observational mortality plans end up in plans with higher observational mortality compared to their inertial counterparts. The key question then becomes: does mortality for those cohorts change in the manner we would predict given observational mortality? If so, we would conclude that observational mortality does not merely vary due to selection and is useful in predicting how mortality would change in counterfactuals where beneficiaries were assigned to different plans. Appendix Figure 9 plots our first-stage estimates of the predicted change in observational mortality against the actual change in mortality for each decile of lagged observational mortality. The first-stage coefficient is on the x-axis and the reduced-form coefficient is on the y-axis. If observational mortality-added is predictive of causal plan effects, we expect this relationship to be (approximately) linear with a slope of one. Naturally, the slope of this graph will correspond to the forecast coefficient we estimate in Section 4.3.

4.2 Balance Tests

Before reporting our estimates of the forecast coefficient, we report tests of the conditions we derived in Section 3 for the validity of our identification strategy.

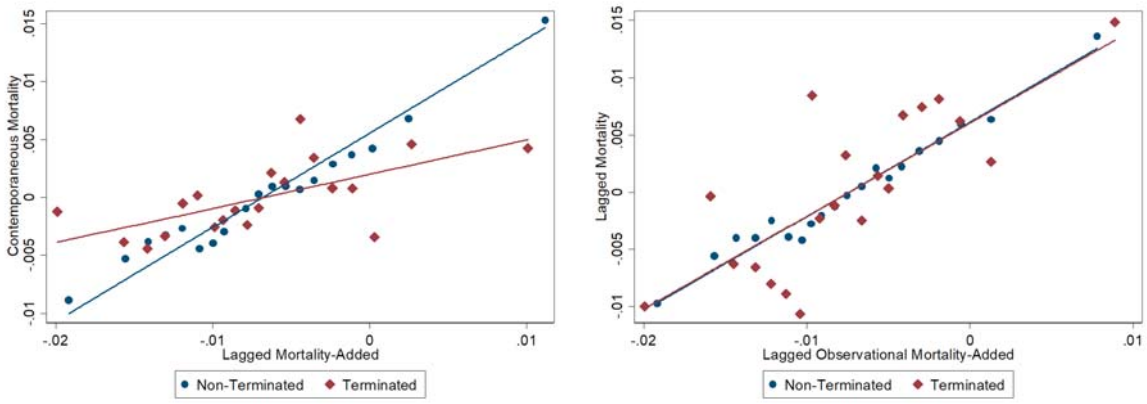
A sufficient condition for the exclusion restriction to hold is that terminations are randomly assigned to plans. The main concern in our setting would be that "bad" plans may be more likely to strategically exit the market. To address this concern, we show that beneficiaries enrolled in terminated plans in period $t-1$ are observationally similar to those enrolled in non-terminated

plans. To show this, we regress predicted mortality given enrollee characteristics on a full set of county and year fixed effects and a termination dummy. The results are in Panel A of Table 5. The coefficients statistically significant (unsurprising, given our large sample) but are small in magnitude – an order of magnitude smaller than the standard deviation of mortality across plans.

Our empirical strategy does not strictly require that terminations be random (though it is a natural sufficient condition). Rather, we need the relationship between the unobservable determinants of mortality and lagged observational mortality to be the same between terminated and non-terminated plans. We can test this directly by placing either lagged mortality or predicted mortality on the left-hand side of our empirical specification in lieu of actual mortality. Figure 4 shows the variation driving our main result, the relationship between mortality and lagged observational mortality by decile of lagged observational mortality. Among non-terminated plans, we see a steep positive relationship: cohorts in high-mortality plans last year have high mortality this year. However, the relationship is much flatter among terminated plans. Cohorts in high mortality terminated plans last year have lower than expected mortality this year and cohorts in low mortality terminated plans have higher than expected mortality this year.

Our balance tests ask whether this is also true of lagged or predicted mortality. In fact, it is not. The first panel of Figure 4 shows that the relationship between predicted mortality and lagged observational mortality is quite similar for terminated and non-terminated plans. One might have worried, for example, that the results in Figure 3 were driven by terminated low observational mortality plans having older enrollees and terminated high observational mortality plans having younger enrollees. The second panel of Figure 4 shows that this is not the case.

Figure 4: Lagged and contemporaneous mortality and mortality-added



A second way of testing whether our exclusion restriction holds is to consider pre-trends. If the difference between terminated and non-terminated high and low observational mortality plans arises in the post period only due to the changes in enrollment induced by re-assignment, then

we should see no differences in the pre-period. This is indeed the case. Appendix Figure 10 considers our usual four categories of terminated and non-terminated plans with above and below median observational mortality, but considers cohorts enrolled in each type of plan two years prior to termination Cohorts are on parallel trends until terminations induce reassignment.

4.3 Forecast Coefficient

In this section, we present second stage estimates that correspond to equation 5. The specification allows us to leverage all of the variation in the data to identify the object of interest λ.We ask whether the changes in mortality predicted given the choices beneficiaries make following terminations and observational mortality measures match the changes in mortality we observe in the data.

Panel B of Figure 4 allows us to highlight the variation driving the reduced form estimates. Most consumers in non-terminated plans remain in the same plan year after year, leading to a strong relationship between *contemporaneous* mortality and *lagged* μ_i^{t-1}. If termination leads consumers to choose more "typical" plans, the relationship between *contemporaneous* mortality and *lagged* μ_i^{t-1} will be muted for consumers in terminated plans in period $t-1$. That is, consumers in lower mortality terminated plans will re-sort into observationally worse plans, leading to higher subsequent mortality. Consumers in high mortality terminated plans will re-sort into observationally better plans, leading to lower subsequent mortality. This is exactly what we see in Panel B of Figure 4.

Table 3 reports estimates of λ from IV estimation. In all specifications, we include market (county-by-year) fixed effects and cluster the standard errors at the county level. We also flexibly control for lagged market share and μ_i^{t-1}, interacted with plan type. Column (1) presents a specification in which we predict contemporaneous μ_i as a function of a termination dummy interacted with lagged μ_i^{t-1} linearly: $Z_i = \mu_i^{t-1} T_i^{t-1}$. The forecast coefficient is 1.034, indicating that the quasi-experimental variation reliably predicts the observational effect.

One concern is that functional form drives our estimates of λ. However in columns (2) and (3) we present specifications in which we instrument with the termination dummy interact with above/below median μ_i^{t-1} and deciles of μ_i^{t-1}, respectively. Across all specifications, we cannot reject that λ is different from 1. A second natural concern – despite the evidence above – is that there is additional scope for selection driving our results. Following the logic of Altonji et al. [2005] and Oster [2019], we show that are results are robust to the inclusion of demographic controls (our preferred specification). Comparing columns (1) and (4) (similary, (2) and (5) or (3) and (6)), we see that the coefficient remains relatively unchanged. Across all specifications, λ ranges from 0.83 to 1.2.

Table 3: Forecast Coefficient

	Dep Var: 365 Day Mortality					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full Sample						
λ	1.034	0.881	1.208	0.936	0.830	1.055
	[0.263]	[0.235]	[0.205]	[0.232]	[0.182]	[0.156]
Specification	Lagged μ_i^{t-1}	Median	Decile	Lagged μ_i^{t-1}	Median	Decile
Demographic Controls	No	No	No	Yes	Yes	Yes
N	16,135,957					

Notes: Table reports regression estimates of the forecast coefficient from IV estimation with linear, median, and decile specifications. Columns (1), (2), and (3) do not include demographic controls. Columns (4)-(6) flexibly control for age and include race and sex dummies.

Plan average mortality-added predicts mortality remarkably well. This result may initially seem surprising. However, the literature, summarized by Cutler, Finkelstein, and McGarry (2006), documents a complex relationship between plan characteristics and consumer sorting. For example, some studies find little evidence of asymmetric information (Cardon and Hendel 2001). By contrast, heterogeneity in risk preferences can lead to advantageous selection (Fang, Keane, and Silverman 2008). Furthermore, extensive margin selection into Medicare Advantage is not inconsistent with our estimates of λ.[14] Finally, consumers may not be aware of or place high value on mortality-added. To explore the extent to which consumers sort into high quality plans, we next document the relationship between mortality-added and publicly available quality measures.

4.4 Correlates of Mortality-Added

We have documented substantial variation in observational mortality-added across insurance plans. The results above indicate that these differences are due, in large part, causal. The relative health effects of different MA plans are likely to swamp relative cost savings, an aspect of plan heterogeneity which has historically been more intensively studied. Put differently, in a well-functioning market, consumers would be willing-to-pay substantially higher premiums for low mortality plans, who would, in turn, capture larger market share over time.

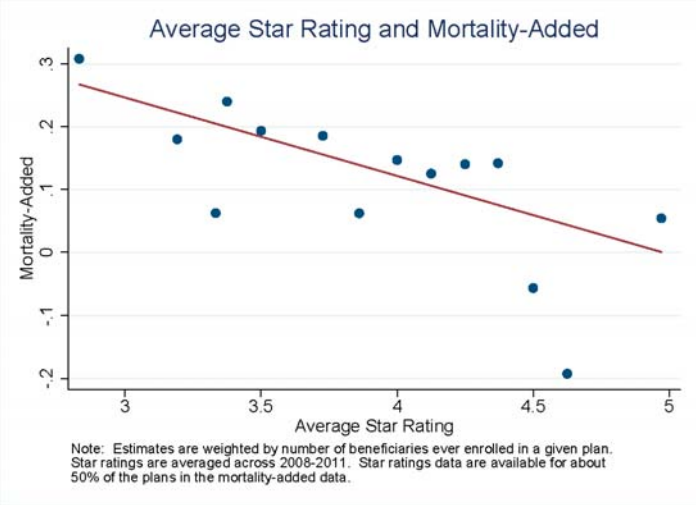
However, there are reasons to be skeptical of this view, as institutional facts may dull the impact of insurer competition. First, consumers may not have access to granular quality information. To help beneficiaries select plans, CMS rates plans on a 1–5 scale, with 5-stars indicating the highest quality. The measures consider both clinical quality and patient experience, discussed below. However, the star ratings have been publicly available since 2007. Second, consumers may not respond to publicly available measures such as star ratings (Dafny and Dranove 2003, Darden and McCarthy 2015). Third, insurers may game these metrics (Decarolis and Guglielmo 2017). A natural first question is whether the publicly available quality information is a good predictor of true mortality-added.

Figure 5 plots a binscatter where plans are binned by star rating, which is depicted on the x-axis. Average mortality-added within each bin is plotted on the y-axis. The line represents the linear best fit. Publicly available quality measures predict our "true" estimates. In fact, the results show that there is a striking negative correlation between average star rating and mortality-added: five star plans have lower mortality-added than three star plans by almost 0.2 percentage points,

[14] In our setting, risk adjustment attempts to equalize insurer profitability across beneficiaries by increasing subsidies for sicker enrollees. Despite this, there may still be selection conditional on the risk adjustment (Brown et al. (2014), Newhouse et al. (2015), Cabral et al. (2018), Duggan et al. (2016), Lustig (2010)). However, many of the patterns of selection documented in this market are fairly subtle (Polyakova (2016), Carey (2017), Han and Lavetti (2017), Lavetti and Simon (2018), Starc and Town (2019)).

about equal to the change in mortality from aging by 1 year at age 70.One standard deviation increase in star ratings is associated with a .07 percentage point decrease in mortality-added.

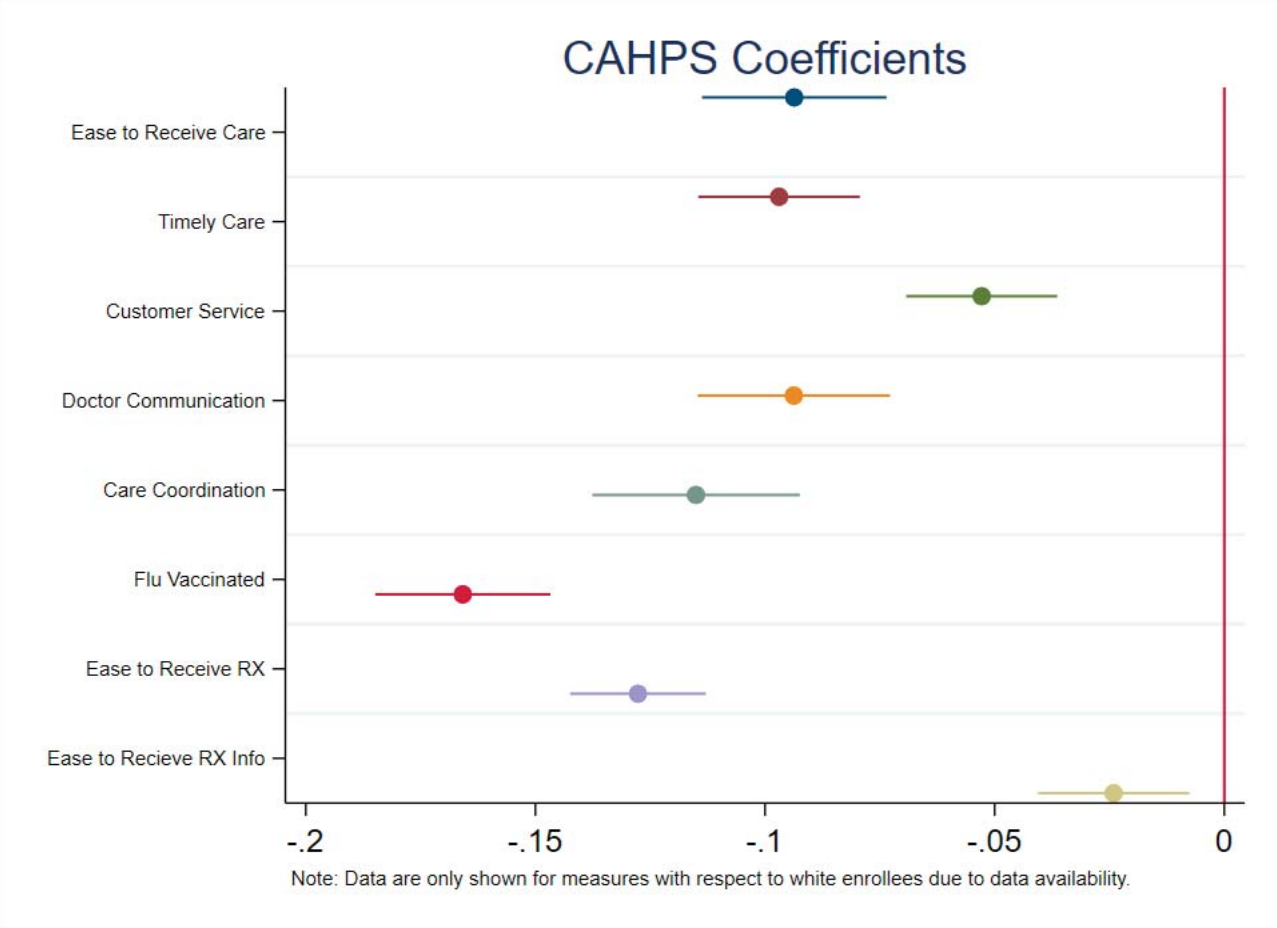
Figure 5: Star Ratings and mortality-added



Our results indicate that publicly available quality predicts mortality-added. This raises two important questions. First, what drives both mortality-added μ_i and the star ratings? Second, what if consumers were more attentive to star ratings or mortality-added μ_i? To answer the first question, we explore heterogeneity in benefit design across several dimensions: publicly available quality scores, financial outcomes, and provider networks. We present detailed definitions, data sources, and summary statistics in Appendix C and focus on observable plan characteristics that could explain the patterns we observe in the data.

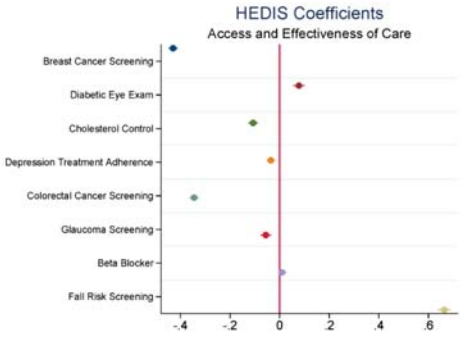
Star ratings are a function of HEDIS scores – which measure clinical quality – and CAHPS scores – which measure consumer satisfaction. To better understand the variation driving the relationship between star ratings and mortality-added, Figure 6 summarizes the correlates of the estimated plan mortality-added. Each row represents a different proxy variable, and coefficients from separate bivariate OLS regressions in which we control for county fixed effects and weigh by enrollment.

Figure 6: Star Ratings and mortality-added



The relationship between mortality-added and CAHPS scores is extremely consistent across categories. Higher consumer satisfaction scores are consistently correlated with lower mortality-added. This is true across categories associated with the plan itself (e.g., customer service, ease to receive prescription information) and with providers with whom the plan contracts. The right panel of Figure 6 explores the relationship between a subset of HEDIS scores and mortality-added. Here, the relationship is less clear. For example, higher cancer screening seems to be consistently correlated with lower mortality-added. However, fall risk screening is positively correlated with mortality-added. Commonly used metrics, such as the use of beta blockers, are also uncorrelated with mortality-added.

Figure 7: Star Ratings and mortality-added

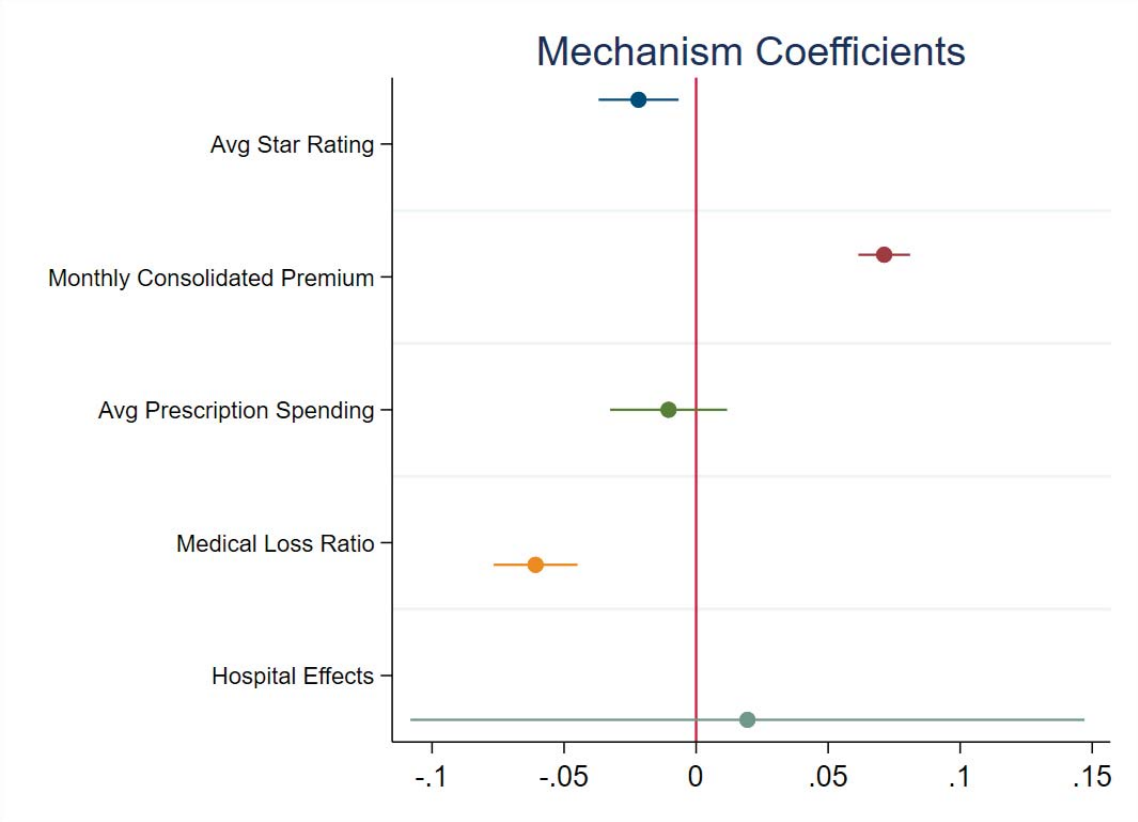


Second, we explore the relationship between measured plan mortality-added and the financial features of plans: premiums and observed medical loss ratios. Medical loss ratios are the percentage of premiums paid out in claims; higher loss ratios are associated with lower insurer margins. We also explore the relationship between measured plan mortality-added and proxies for utilization. Because we do not observe all medical claims directly, we focus on prescription drug claims. Previous research has shown that MA plans offer generous coverage of valuable drugs (Starc and Town 2019), which is associated with higher utilization. To explore this relationship, we leverage detailed prescription drug claims. While medical claims for MA enrollees are not systematically available, we observe drug claims for a 20% sample of MA enrollees.

Third, we explore the relationship between measured plan mortality-added and plan networks. We measure plan networks along two dimensions: hospitals and physicians. Previous research documents the health effects of choosing different hospitals, doctors, nurses, and medical service areas (Hull 2018, Fletcher et al. 2014, Yakusheva et al. 2014, Finkelstein et al. 2017). To measure hospital networks, we gather discharge data by payer for California, Massachusetts, and Maryland. We combine this data with measures of hospital quality from Hull (2018). By averaging across discharges within a payer, we create weighted average hospital quality. To measure physician networks, we follow the procedure outlined in Frakt et al. (2019). This procedure allows us to characterize payers as having "broad" or "narrow" physician networks.

Figure 8 summarizes the correlates of the estimated plan mortality-added. Each row represents a different proxy variable. The points in the left panel are coefficients from separate OLS regressions in which we control for county fixed effects and weigh by enrollment. The points in the right panel are coefficients from post-Lasso multivariate OLS. All quality measures and covariates are normalized to standard deviation units for ease of interpretation.

Figure 8: Financial Features, Network Breadth, and mortality-added



The results show that high quality plans are systematically different in terms of financial performance. High mortality-added plans tend to have higher premiums. However, plans with lower MLRs also tend to have higher mortality-added. That is, high quality plans tend to pay out a higher fraction of premiums in claims. They also tend to spend more on prescription drugs, though we cannot reject no correlation. Finally, our measures of hospital network quality are extremely noisy and exist for only three states. Perhaps as a result, the relationship between hospital quality effect and mortality-added is extremely noisy.

5 Conclusions

We document substantial variation in mortality rates across Medicare Advantage plans which persist after controlling for demographics and prescription drug consumption. Leveraging a quasi-experiment generated by plan terminations, we show that this variation reflects the causal effect of plan enrollment on mortality. Additionally, we find that consumers are insensitive to variation in mortality-added when they choose plans. Given conventional estimates of the value of a statistical life, health effects of the magnitude we document should dwarf all other factors relevant for plan

choice.

Our analysis has several policy implications. Naturally, it raises the question of why some plans are better than others and whether worse-performing plans can reduce mortality by imitating better performing ones, or whether better plans direct patients to rival resources (such as better doctors) which cannot be concurrently utilized by all plans. We are investigating this question in ongoing work.

Our analysis also raises the question of why consumers are not more responsive to variation in plan mortality-effects. The obvious answer is that they are not aware of these differences. While consumers can use rough heuristics such as whether plans include prestige hospitals in their network, plan mortality rates are not published. This suggests there may be large gains from helping to direct consumers to higher-quality plans, both in a partial equilibrium sense and even moreso in a general equilibrium environment where such policies might induce plans to invest more in quality improvements. A note of caution is in order however: publishing unadjusted plan mortality rates might induce plans to invest in selecting healthier beneficiaries rather than improving beneficiary health. Ideally, less gameable measures of plan mortality-added should be used to help beneficiaries sort into better plans.

Our results highlight both the shortcomings and the promise of the multipayer system currently in vogue in the US. In the current system, because consumers are not informed about plan mortality-added, plans have too little incentive to invest in making consumers healthier. Nonetheless, some plans are extremely high performing. A better designed system could ensure that the benefits of these high performing plans were more widely shared as well as accelerating the adoption of technologies which provide higher-quality care at lower cost.

References

J. Abaluck and J. Gruber. Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review*, 2016.

Jason Abaluck and Jonathan Gruber. Heterogeneity in choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program. *American Economic Review*, 101 (3):377–381, 2011.

Joseph Altonji, Todd Elder, and Christopher Taber. Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1): 151–184., 2005.

Joshua Angrist, Peter Hull, Parag Pathak, and Christopher Walters. Interpreting tests of school vam validity. *American Economic Review: Papers & Proceedings*, 106(5):388–392, 2016.

Joshua Angrist, Peter Hull, Parag Pathak, and Christopher Walters. Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 132(2):871–919, 2017.

R. Chetty, J.N. Friedman, and J.E. Rockoff. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632, 2014a.

Raj Chetty and Nathaniel Hendren. The impacts of neighborhoods on intergenerational mobility: County-level estimates. NBER Working Paper No. 23002, 2017.

Raj Chetty, John Friedman, and Jonah Rockoff. Measuring the impact of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679, 2014b.

Vilsa Curto, Liran Einav, Jonathan Levin, and Jay Bhattacharya. Can health insurance competition work? evidence from medicare advantage. NBER Working Paper No. 20818, 2015.

Joseph Doyle, John Graves, and Jonathan Gruber. Evaluating measures of hospital quality. NBER Working Paper No. 14607, 2017.

David Dranove, Christopher Ody, and Amanda Starc. A dose of managed care: Controlling drug spending in medicaid. Working paper, October 2017.

Mark Duggan, Amanda Starc, and Boris Vabson. Who benefits when government pays more? pass-through in the medicare advantage program. NBER Working Paper No. 19989, 2015.

Amy Finkelstein, Matthew Gentzkow, Peter Hull, and Heidi Williams. Adjusting risk adjustment – accounting for variation in diagnostic intensity. *New England Journal of Medicine*, 376(7): 608–610, 2017.

Michael Geruso, Timothy J. Layton, and Daniel Prinz. Screening in contract design: Evidence from the aca health insurance exchanges. Forthcoming, American Economic Journal: Economic Policy, 2017.

B. R. Handel and J. T. Kolstad. Getting the most from marketplaces: Smart policies on health insurance choice. Brookings Hamilton Project Discussion Paper 2015-08, 2015.

Benjamin Handel. Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, (103):2643–2682, 2013.

Peter Hull. Estimating Hospital Quality with Quasi-Experimental Data. SSRN Scholarly Paper ID 3118358, Social Science Research Network, Rochester, NY, February 2018. URL https://papers.ssrn.com/abstract=3118358.

Emily Oster. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2):187–204, 2019. doi: 10.1080/07350015.2016.1227711. URL https://doi.org/10.1080/07350015.2016.1227711.

K. Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2009.