NBER WORKING PAPER SERIES

THE HEALTH COSTS OF COST-SHARING

Amitabh Chandra
Evan Flack
Ziad Obermeyer

Working Paper 28439
http://www.nber.org/papers/w28439

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2021

This work was funded by grant P01AG005842 from the National Institute on Aging. We thank, without implicating, Amy Finkelstein, Ben Handel, Jon Kolstad, Jon Gruber, Sendhil Mullainathan, Julian Reif, Jon Skinner, and Josh Schwartzstein, for many comments that have improved our paper. We thank Mohan Ramanujan for technical assistance with the Medicare data. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w28439.ack

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Amitabh Chandra, Evan Flack, and Ziad Obermeyer. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Health Costs of Cost-Sharing
Amitabh Chandra, Evan Flack, and Ziad Obermeyer
NBER Working Paper No. 28439
February 2021
JEL No. I12,I13

ABSTRACT

We use the design of Medicare's prescription drug benefit program to demonstrate three facts about the health consequences of cost-sharing. First, we show that an as-if-random increase of 33.6% in out-of-pocket price (11.0 percentage points (p.p.) change in coinsurance, or $10.40 per drug) causes a 22.6% drop in total drug consumption ($61.20), and a 32.7% increase in monthly mortality (0.048 p.p.). Second, we trace this mortality effect to cutbacks in life-saving medicines like statins and antihypertensives, for which clinical trials show large mortality benefits. We find no indication that these reductions in demand affect only 'low-value' drugs; on the contrary, those at the highest risk of heart attack and stroke, who would benefit the most from statins and antihypertensives, cut back more on these drugs than lower risk patients. Similar patterns exist for other drug–disease pairs, and irrespective of socioeconomic circumstance. Finally, we document that when faced with complex, high-dimensional choice problems, patients respond in simple, perverse ways. Specifically, price increases cause 18.0% more patients (2.8 p.p.) to fill no drugs, regardless of how many drugs they had been on previously, or their health risks. This decision mechanically results in larger absolute reductions in utilization for those on many drugs. We conclude that cost-sharing schemes should be evaluated based on their overall impact on welfare, which can be very different from the price elasticity of demand.

Amitabh Chandra
John F. Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
amitabh_chandra@harvard.edu

Ziad Obermeyer
School of Public Health
University of California at Berkeley
2121 Berkeley Way
Berkeley, CA 94704
zobermeyer@berkeley.edu

Evan Flack
Harvard Kennedy School
evanjflack@gmail.com

A data appendix is available at http://www.nber.org/data-appendix/w28439

1 Introduction

As insurers place more emphasis on cost-sharing via deductibles, coinsurance, and copayments, patients pay more out-of-pocket for health care. This can induce more efficient care if patients can balance these costs against benefits. But this calculus is not easy. How should a patient rank-order her 5 prescribed medicines in terms of value (120 options), and where should she draw the line (6 options)? A nascent literature in economics and medicine suggests that patients struggle with this task, and respond to increases in out-of-pocket prices by cutting back indiscriminately on low- and high-value health care alike (Brot-Goldberg et al., 2017; Chandra et al., 2010; Choudhry et al., 2011). This fact is often taken as evidence that patients respond to prices in ways that systematically deviate from the cost-benefit calculus, an idea termed 'behavioral hazard,' with far-reaching implications (Baicker et al., 2015). Could the cost-sharing measures put in place to reduce over-use of care also reduce health (Chandra et al., 2019)?

Despite its potential importance, direct evidence for this phenomenon is scarce. Cutbacks in care are an indirect measure of patient welfare, and might not translate into more direct measures, like mortality. Reductions in demand are welfare-enhancing if there is over-consumption of care, whether driven by moral hazard (Cutler and Zeckhauser, 2000; Einav and Finkelstein, 2018) or physician beliefs (Cutler et al., 2019; Chandra and Staiger, 2020). Since the RAND Health Insurance Experiment (Newhouse and Group, 1993), the first study to note broad cutbacks due to cost-sharing, the literature has sought to pinpoint over-consumption by relying on expert categorizations of care: 'low-value' or 'high-value.' But these labels are blunt, and ignore large heterogeneity in benefits across patients (Chandra and Skinner, 2012; Chandra and Staiger, 2020). Care that is low-value on average can be very high-value for some patients (and reciprocally). Stress testing, a prime example of low-value care, is high-value for a patient with heart attack (Mullainathan and Obermeyer, 2019). So an outside observer might be a poor judge of what is high-value care for a patient, particularly given the patient's information advantage: cutbacks may be maximizing other dimensions of utility the observer does not see, such as side effects, risk preference, or other consumption (Einav et al., 2015). In other words, welfare cannot be inferred from demand responses alone (Baicker et al., 2015). When a patient reduces care in response to price changes, what matters for welfare is not whether an expert believes that care to be high- or low-value, but whether the reduction led to

worse outcomes for that patient.

Direct harms to health resulting from cost-sharing, though, have been difficult to detect. This is not for lack of study. A central challenge is that the prices patients face are not random, and typically depend on prior utilization, creating spurious correlations between prices and health. In some settings, this can be solved via experimental or quasi-experimental shocks to prices. But most settings with clean identification also have data constraints. Mortality is not always measured, and when present, it is rare. To measure a 10% increase over a 1% baseline risk of mortality at age 65, for example, would mean randomizing over 325,000 patients. No rigorous studies of cost-sharing even come close (e.g., in Finkelstein et al. (2012), the Oregon Health Insurance Experiment treatment group was 9,000 patients, and the RAND experiment randomized approximately 5,500 patients in total (Newhouse and Group, 1993)). So it is perhaps unsurprising that studies have largely found null effects (Choudhry et al., 2011), or effects only on proxies for health outcomes, largely utilization of hospital or emergency care, or self-reported health (Chandra et al., 2010; Finkelstein et al., 2012; Geruso et al., 2020).

Here we explore the impact of cost-sharing on mortality, using data from Medicare's prescription drug program. To simulate random assignment of drug prices, we use a strategy introduced by Aron-Dine et al. (2015) and Kaplan and Zhang (2017): variation in drug prices as a function of a beneficiary's month of birth. This variation flows from a quirk in Medicare's drug benefit structure, and specifically the annual spending thresholds that shift the price of drugs. Every January, beneficiaries start the year paying 25% out-of-pocket for drugs; but when they reach approximately $2500 of total drug consumption, they pay 100% out-of-pocket for the next drug. By themselves of course, these price changes are not exogenous: they depend on prior utilization. But critically, plan thresholds are not pro-rated in beneficiaries' first calendar year of enrollment, and eligibility for enrollment begins in the month beneficiaries turn 65. So those born in later months of the year enroll in later months of the year, and in turn have less time to reach thresholds, meaning they face lower prices on average. Thus birth month creates exogenous variation in prices, by influencing enrollment month in the program.[1] We focus on end-of-year prices and outcomes, specifically the

[1]To illustrate, a patient turns 65 in February and enrolls in Part D. If she spends more than ∼$2500 over the next 11 months, she will go from paying 25% of her drug costs to 100%. An otherwise identical September enrollee, however, has only 4 months to spend the same amounts, meaning she is less likely to exceed ∼$2500: she will pay only 25% of her drug costs.

month of December following Einav et al. (2015) and Einav et al. (2018). This strengthens our instrumental variables strategy: the later in the year, the more differences in measured spot price across enrollment months correspond to differences in full future price of a drug, before prices reset for all enrollment months on January 1.

So far we have relied heavily on the existing literature for our identification strategy, but have not addressed a major unsolved problem: used alone, the enrollment month instrument lacks the precision needed to detect health effects. This is because while enrollment month shifts December prices, it is a blunt tool. Within a given month, there is still wide variation in year-end prices, as a function of a beneficiary's particular spending trajectory. For example, consider two February enrollees. One spends $500 per year, meaning despite her early-year enrollment, she remains far from spending thresholds, and pays only 25% of her drug costs at year-end. Another spends $5,000 per year, meaning she enters the 'donut-hole' coverage gap, and pays the full 100% of her drug costs at year-end. Instrumenting for price with enrollment month alone will assign both the same (average) year-end price, over-estimating the price for the first, and under-estimating it for the second. The resulting imprecision in the first stage creates problems for detecting rare health outcomes in the second. This could be avoided if we could condition on spending trajectories: ideally, we would like to compare beneficiaries on similar trajectories, who face different prices solely because they enroll in different months, instead of comparing large groups with heterogeneous prices. But naturally, we cannot use realized year-end spending, which is endogenous to enrollment month-driven variation in cost-sharing.

Our strategy is to estimate, for each beneficiary, what their total 12-month spending *would* be in the absence of cost-sharing, and control for this prediction in our analysis. To do so, we draw on data from a separate sample of Medicare beneficiaries to generate predictions on spending. The sample is similar to the one we study, but unaffected by enrollment month or cost-sharing: 'dual-eligible' 65-year old Medicare enrollees, on Medicaid or other low income subsidies, who have the same enrollment criteria for Medicare Part D but face minimal cost-sharing. With these data in place, the task of predicting 12-month spending in the absence of cost-sharing is a straightforward 'prediction problem' (Kleinberg et al., 2015). We use machine learning tools to fit a function in the dual-eligible sample, and apply it to generate 'counterfactual' predictions in our main sample: how much would they have spent in 12 months, without cost-sharing? Interacting these predictions with

enrollment month in our first stage allows us to generate highly accurate instrumented estimates of year-end prices, that vary widely—but still exogenously—within enrollment month.

Our analysis reveals three facts. First, we document a large mortality burden attributable to cost-sharing. We focus on the top 30% of our sample in terms of predicted spending, who are close enough to plan thresholds for their prices to vary with enrollment month. We estimate a 33.6% (11.0 percentage points, p.p, change in coinsurance, or $10.40 per fill) median price change, and a 32.7% (0.048 p.p.) increase in monthly mortality. We conduct a range of falsification checks that argue against violations of the identifying assumptions (i.e., some enrollment months being systematically sicker than others), and rule out 'harvesting', small changes in the timing of death for people who would have died soon after. We also note that, because of the rarity of mortality, even these large effects on mortality would not be detected by the typical physician: detecting a 35% increase in mortality from a baseline of 1% would require a physician with perfect recall and 30,000 patients with randomly assigned prices (vs. a typical physician panel size of 1500-2000 patients (Raffoul et al., 2016)). Our estimates capture mortality effects in one setting: December mortality for non-disabled 65-year-olds. But patients in Medicare—and those in other insurance plans—face a range of similar price increases throughout the year, and over their life-span. Thus the total mortality impact of cost-sharing is likely to be far larger than the specific setting we study.

Second, we explore the behaviors underlying this effect. We examine the particular drugs patients drop, and find large cutbacks in medicines with known mortality effects from clinical trials. For each p.p. increase in the coinsurance rate, patients make 5.6% to 18.9% fewer fills for drugs that lower cholesterol (statins), blood pressure (e.g., ACE inhibitors, beta blockers), blood sugar (oral hypoglycemics); and drugs that treat acute exacerbations of emphysema and asthma (inhalers). Reductions in these life-saving classes account for 42.1% of the overall demand response, and simply multiplying by their average mortality effects from clinical trials, we can account for 25.8% of the overall mortality increase we observe. Of course, just because these drugs appear 'high-value' on average does not mean they are high-value to the patient who drops them. But if anything, individual patients with the most to gain from these drugs—those at the highest risk of precisely the outcomes the drugs prevent or treat—have equal or even higher reductions in demand vs. lower-risk patients. Using a machine learning model to predict adverse cardiovascular events like heart attack and stroke, we find that the riskiest one-third are 280.6% *more* likely to drop cardiovascular

drugs (e.g., statins) than the bottom two-thirds. We find similar results for those at high risk of diabetic and pulmonary complications. And while we do not observe individual-level income, we demonstrate that these effects are similar across high- and low-income zip codes alike, making it less likely, but not impossible, that these effects are driven by liquidity and socioeconomic forces.

Third, we document crude, but deadly, empirical regularities in patient behavior. Simply inspecting the number of drugs filled at year-end reveals a substantial mass of people who, when faced with higher prices, choose to fill *no* drugs—no matter how many drugs they were on prior to the price shock, or their individual health risks. To illustrate the non-trivial scale of this behavior transparently, we compare February to September enrollees predicted to be in the coverage gap. In response to a 14.9 p.p. mean increase in price, 2.8 p.p (18.0%) more choose to fill no drugs at year-end. Among patients on 4 or more drugs at baseline, 6.2 p.p. (72.6%) more fill no drugs.[2] Mechanically, this behavior results in larger absolute reductions in drug use for those on more drugs to begin with, and is correlated with higher marginal mortality rates. This suggests that price-induced cutbacks in high- and low-value care, long known to exist on average in populations, also exist *within individuals*, when they cut back on all their medications.

These findings provide evidence that the price elasticity of demand is an insufficient statistic for welfare, a possibility noted by both Baicker et al. (2015) and Einav and Finkelstein (2018). For the decision-making underlying the mortality effect to be optimal, using conventional notions of optimality and the usual life-year valuation of $100,000 per year, a 65-year old in our sample would have to believe that she had at most 1.28 years left to live. Alternatively, based on average life expectancy at 65 (19.2 years), her implied life-year valuation would be $6,628 per year. These extremely low life-year valuations could be viewed as an opportunity: improving the design of prescription drug insurance offers policy makers the opportunity to purchase large gains in health at extremely low cost per life-year. They also raise a variety of other important research and policy questions, many of which hinge critically on the cognitive underpinnings of patient decisions. How do patients actually balance benefit against cost, and why do those on k drugs chose to drop k drugs—instead of m, where $(k - m)$ is optimal? And how should policy-makers design insurance that is responsive to these mental processes, and helps patients make the best possible decisions?

[2] This does not appear to reflect a floor effect, i.e. zeros are not more common simply due to left-censoring: it is more common, not less, in those with more pre-shock drug fills, and more than we would predict in a simulation model where each drug has a certain likelihood of being dropped.

For example, many current initiatives aim to make prices more visible and transparent to encourage 'shopping.' These well-intentioned efforts could aggravate several potential mechanisms of error in decision making: cost salience (Bordalo et al., 2013), loss aversion (Kahneman and Tversky, 1979), present bias (Laibson, 1997; O'Donoghue and Rabin, 1999), and inattention (Handel and Schwartzstein, 2018; Gabaix, 2019), to name a few. The total welfare impact of such policies—which may be quite different from their cost impact alone—deserves careful scrutiny.

2 Empirical Strategy

2.1 Price changes

Since 2006, Medicare Part D has offered prescription drug coverage to seniors and disabled individuals in the US. The benefit's non-linear structure with respect to out-of-pocket costs is illustrated in Panel A of Figure 1 (Einav et al., 2015). Using the 2008 details to describe the plan, coverage begins with a deductible phase in which the beneficiary pays the entire cost of all drugs until she has spent $275. She then faces a 25% coinsurance rate that lasts until her total spending has reached the initial threshold of $2,510 (called the initial coverage limit, or ICL). After this, the beneficiary falls into the "coverage gap" or "donut hole" and again pays the total cost of all drugs. The beneficiary is then insured again by the "catastrophic coverage" after reaching $5,726 of total spending (catastrophic coverage limit, or CCL). In the catastrophic arm she either has a 5% coinsurance rate or copays of $2.25 for generic or preferred drugs, and $5.60 for other drugs (Figure 1).[3] We will refer to these coinsurance rates as prices for simplicity, but note they are out-of-pocket prices, and should not be confused with the full price of the drug—which patients are not paying and do not routinely observe. The cutoff points for each coverage arm change slightly from year to year but the basic structure remains the same. Starting in 2011, the coverage gap began to close as a result of policy changes (it fully closed in 2020), with coinsurance rates for generic and branded drugs in

[3]The overall non-linear structure was largely the result of a political compromise balancing the desire to cover very sick beneficiaries (the catastrophic phase) with reducing the total cost of the program (the coverage gap); a review is found in Oliver et al. (2004). Insurers may offer coverage that is "actuarially equivalent", or "enhanced" compared to the standard benefit. One common deviation from the standard design is to replace the deductible phase with uniform coinsurance (∼31%) until the first threshold is reached. Additionally, most plans do not base cost-sharing on coinsurance rates but rather copays based on drug tiers for each coverage arm. In practice, copays equate to roughly the same level of cost-sharing in each arm as the coinsurance rates specified by the standard benefit (Einav et al., 2015).

the coverage gap falling from 100 percent to 50 and 93 percent respectively. Coverage arm spending thresholds and coinsurance rates by year are shown in Appendix Table A.1.

2.2 Simple Model

With these policy details in mind, we first consider a simple OLS approach where mortality is regressed on price at the end of the year (EOY):

$$\mathsf{Mortality}_i^{\mathsf{EOY}} = \beta_0 + \beta_1 \mathsf{Price}_i^{\mathsf{EOY}} + \mathsf{Year}_i \beta_2 + X_i \beta_3 + \epsilon_i \tag{1}$$

Here, $\mathsf{Mortality}_i^{\mathsf{EOY}}$ is an indicator for patient i's mortality in December, $\mathsf{Price}_i^{\mathsf{EOY}}$ is the average coinsurance rate (also in December), Year_i is the calendar year, and X_i is a vector of demographic controls (race and sex); the year controls allow for changes in the Part D plan design over time and capture secular changes in prices and mortality.

We focus on the end of the calendar year following Einav et al. (2015) and Einav et al. (2018). The closer we get to the end of the year, the more spot prices approach future prices. This is important because future prices earlier in the year are difficult to pin down: the decision to fill a drug at a given spot price early in the year changes the future price in complex ways, because of the non-linear price schedule (i.e., price increases in the coverage gap, and decreases in the catastrophic coverage). As a result, there is no single well-defined price elasticity of demand: demand responses earlier in the year combine responses to spot prices at a given point in time, with responses to expectations about prices later in the year—all of which are nonlinear because of the thresholds. Later in the year, however, these complexities are mitigated: there is less time to reach the next spending threshold before the January reset, meaning spot prices and future prices equalize. So concretely, we restrict ourselves to the month of December, when we are able to measure the "short-run" demand and health responses to increases in the end-of-year spot price. These are well-defined economic objects, and likely the parameters of most interest to economists and policy-makers.

It is easy to see why estimates of (1) produce incorrect estimates of the health response to cost-sharing, for $\mathsf{Price}_i^{\mathsf{EOY}}$ is not exogenous. Patients select into an end-of-year price as a function of their prescription drug spending over the course of the year. Their spending is in turn determined by both their underlying health, and thus need for drugs, as well as their own choices regarding which drugs

to fill. Sicker patients fill more drugs, spend more, and fall into the coverage gap (higher price)—unless they cut back on filling drugs, spend less, and avoid the gap; healthier patients consume fewer drugs, spend less, and remain in initial coverage (low price). Confounding, by baseline health and patients' choices, makes it impossible to identify a causal link between prices and health, and motivates the need to find a setting where higher or lower prices are randomly assigned. This need motivates our instrumental variables (IV) approach, that uses enrollment month to generate variation in end-of-year prices.

2.3 Instrumental Variables Model

As noted by Aron-Dine et al. (2015) and Kaplan and Zhang (2017), the spending thresholds that define coinsurance are not pro-rated in the first year of enrollment. As a result, a person who enrolls in February, say, is more likely to be exposed to higher prices from the coverage gap by the end of the year than a person who enrolls in September. Aron-Dine et al. (2015) and Kaplan and Zhang (2017) also note that because individuals are eligible to enroll in Medicare Part D on the first day of their 65th birth month, enrollment at age 65 is primarily driven by birth month. So birth month drives enrollment month, which drives price variation. If variation in enrollment month is as-good-as random, an assumption we test—via a range of falsification tests below, and additionally in Appendix Table A.2—then so is the resulting variation in end-of-year price. We follow Aron-Dine et al. (2015) and use enrollment month as the instrument in our main specifications, as opposed to birth month, because it lets us instrument for year-end prices more precisely; we show in Table 4 that that results are similar—if somewhat less precise—if we use birth month instead.[4]

Thus our basic instrumental variables model uses enrollment month to instrument for end-of-year price (coinsurance rate). The assumption is that enrollment month changes the likelihood that a beneficiary reaches a given plan coverage arm by end-of-year, which changes the price they would have to pay for a drug and ultimately drug consumption.[5]

[4]Unlike Medicare Part A, enrollment in Part D is not automatic. Beneficiaries can enroll during a 7-month long initial enrollment period that runs from three months before to three months after their 65th birth month (though those enrolling before their birth month do not begin coverage until their birth month). If an individual chooses to enroll later, she faces a penalty of higher premiums for the remainder of her tenure on Part D. Coverage starts on the first day of the month after the individual enrolls, but not before the first day of the enrollee's birth month. The vast majority of beneficiaries enroll in their birth month (64%) or in the IEP (87%). For those who do not, we verify that they appear to select into enrollment months similarly across birth months (Appendix Figure A.1).

[5]Below we explore the possibility that drug price variation acts on health via mechanisms other than drug consumption, e.g., by reducing other health consumption like physician visits or hospitalizations. While we find little

This yields the following first stage and reduced form equations:

$$\mathsf{Price}_i^{\mathsf{EOY}} = \pi_0 + \pi_1 \mathsf{EnrollMonth}_i + \mathsf{Year}_i \pi_2 + X_i \pi_3 + u_i \tag{2}$$

$$\mathsf{Mortality}_i^{\mathsf{EOY}} = \theta_0 + \theta_1 \mathsf{EnrollMonth}_i + \mathsf{Year}_i \theta_2 + X_i \theta_3 + \epsilon_i \tag{3}$$

2.3.1 Refining the Instrument

Using enrollment month alone, however, would result in poor estimates of instrumented year-end prices in the first stage. As written, the instrument in equation (2) estimates similar year-end prices for all beneficiaries enrolling in the same month, and ignores a large amount of heterogeneity in prices within enrollment month. On average, our instrument is valid because prices do decrease in enrollment month: a February enrollee is more likely to enter the coverage gap's higher prices (a coinsurance rate of 100%) than a September enrollee who will likely stay in the initial coverage's lower prices (a coinsurance rate of 25%). But if we also knew that a February enrollee would only spend around $500 by year-end, for example, we would know that—unlike another February enrollee who would spend $5000 by year-end—she will have lower prices: even with nearly the whole year to accumulate spending, she will not reach $2500 or enter the coverage gap. Some large fraction of the sample is effectively made up of these 'non-compliers,' whose spending is too low to be affected by the spending thresholds. For these patients, enrollment month will not affect their price—but equation (2) will nonetheless estimate a higher price for them, based on the average effect of enrollment month driven by compliers.

Within the group of compliers, whose spending is high enough to approach any spending threshold and shift prices, there is another kind of heterogeneity. Among patients whose spending is very close to the first coverage gap threshold, price variations due to enrollment month will be less extreme than those whose spending would place them squarely in the middle of the gap: the latter are sure to fall into the higher prices of the coverage gap if born earlier in the year, whereas the former may or may not. This creates larger vs. smaller price gradients by enrollment month, respectively. A more extreme example of heterogeneity affects the highest-spending beneficiaries.

evidence for this, we emphasize that this would not violate the exclusion restriction. By contrast, using drug consumption as the endogenous variable would be harder to reconcile with the possibility that these other channels produce health effects.

As noted above, earlier enrollment months expose roughly 30% of patients to a higher likelihood of being in the coverage gap (higher year-end price: 100% coinsurance), meaning prices go up on average. But the highest 2-3% of the sample (as we show directly below) is—after being exposed to these higher prices of the coverage gap—subsequently exposed to a higher likelihood of being in the catastrophic coverage (lower year-end price: 7% coinsurance), meaning prices go down by year-end. While this non-monotonicity affects only a small fraction of our population, it does muddy the interpretation of local average treatment effects.[6]

Throughout this discussion of heterogeneity in year-end prices within enrollment month, one key, but unobserved, variable emerges as critical: how much would a beneficiary spend, absent enrollment month effects and plan thresholds? If we knew a beneficiary's latent spending 'type'—their level of year-end spending with respect to Part D thresholds–we could estimate the effect of enrollment month precisely and without monotonicity concerns: the effect of enrollment month would be strictly monotonic within a spending type. This implies a specific empirical challenge: to estimate this latent variable with observed variables. Once such a proxy for (counterfactual) spending is estimated, we could then condition on it in our first stage, estimating enrollment month effects flexibly within bins of the proxy. While the effect of enrollment month would vary across bins of spending—no effect for low spenders, and effects of varying magnitudes and even signs for those approaching and above spending thresholds—the effects would be homogeneous and highly precise within sufficiently fine-grained bins.

2.3.2 Predicting Counterfactual Spending

We now turn to the task of creating a proxy measure for counterfactual year-end spending. Of course, we cannot simply use realized year-end spending, because it is endogenous: it incorporates the many downstream consequences of enrollment month. For example, we would not want to compare a February enrollee who spent $1000 by year-end to a September enrollee who spent $1000 by year-end. Both end the year facing the low prices of the initial coverage phase. But the former (who spent roughly $100 per month to reach $1000) is likely to have far better health outcomes at baseline than the latter (who spent $300 per month, and would have spent $3000 by year end had

[6]The technical requirement for monotonicity in our setting is that the instrument (later enrollment month) decreases the likelihood of facing higher prices in December, but for a fraction of the sample the instrument increases the likelihood of facing higher prices because of catastrophic coverage.

she been born in February). So we would not want to group these two beneficiaries together in the same bin when estimating the effect of enrollment month: the effect of enrollment month would be confounded by baseline health.

The example above suggests that utilization early in a beneficiary's Part D enrollment—specifically, before cost-sharing starts to affect utilization—might hold clues about latent end-of-year spending level, and thus the price they would face in the absence of enrollment month variation. We suggest that this is a 'prediction problem' (Kleinberg et al., 2015) and thus amenable to machine learning predictions: the goal is to make an accurate predictions on how much a beneficiary would spend (not, e.g., to understand the drivers of this spending). If we had such predictions on 'counterfactual' spending, it would be straightforward to control for them in an improved instrumental variables setup, that flexibly estimates how enrollment month affects year-end prices within bins of end-of-year spending.

To form these predictions, we require two data elements. First, we need some window into what beneficiaries' spending would be if uncontaminated by enrollment month (and by future spending thresholds and price expectations). We choose to specify our prediction target as the spending level a beneficiary would reach in 12 months of enrollment, if she faced minimal cost-sharing (and thus no enrollment month variation in cost-sharing). We do so because we have data on this exact quantity from a separate sample of beneficiaries, who are very similar to the ones we study, but face near-zero cost-sharing: 65-year old Medicare enrollees eligible for Medicaid and other low income subsidies (LIS), the so-called 'dual-eligibles.' These patients have the same enrollment criteria for Medicare Part D as others, but their coinsurance rate is not a function of drug spending, and there is no coverage gap. These beneficiaries will form the sample in which we train a predictive model of spending as a function of observable characteristics.

With this sample in place, the second data element we require is a set of predictors that are formed in the same way in both the prediction sample and our main sample. Most importantly, the predictors themselves must not be affected by enrollment month, which affects utilization differently between the sample in which the prediction is formed (dual-eligibles) and the sample in which it is applied (our main analytic sample). In other words, the covariates must be 'pre-treatment,' where the treatment is enrollment month-driven cost-sharing. Demographics are an obvious candidate set of variables, since they are unaffected by enrollment month—but empirically, they do not have

enough predictive power to generate good predictions. Guided by the prior literature on behavioral responses to cost-sharing thresholds (Einav et al., 2015), we thus consider data measured early in beneficiaries' Part D enrollment, in our case the first 90 days. It has previously been noted that drug filling behavior in this period does not seem to respond to future price changes from the coverage gap. We verify this, in Appendix Table A.2, where we show that 30- 60-, and 90-day Part D spending are all similarly indistinguishable across enrollment months, arguing against any early gradient in spending induced by cost-sharing. Nor does plan choice appear to vary with enrollment month, as we show that Part D premiums are balanced across months in Table A.2. Below, in Table 3, we show that our results on mortality are similar, albeit somewhat less precise, if we use shorter periods of initial claims to form spending predictions (30 and 60 days).

With these two data elements in place, we then form predictions on how much dual-eligibles would spend in a 12-month period. We start with a randomly-selected training set of 88,854 dual-eligible enrollees, 70% of the full population. As predictors, we use demographics and Part D claims to form a set of 1,770 variables that include sex, race, zip code, drugs filled, and spending in the first 90 days of enrollment. As our outcome, we form a measure of 12-month spending from enrollment (e.g., for an enrollee who joined on September 1, we are predicting spending from September 1 to August 31 of the following year as a function of these features). Our machine learning model consists of an ensemble of two predictors, LASSO (ℓ_1-regularized regression) and gradient boosted trees (a combination of multiple tree-based models, each fit to the residual of the last). Both are fit on our training set, and used to generate predictions in a separate 10% sample, where we perform no-intercept OLS of 12-month spending on the two predictions to create a weighted average, the final output of the model. We use this ensemble to generate predictions in the remaining 20% hold-out set of dual-eligibles that the model has never seen (i.e., a random sample of patients who do not appear in either the training or the ensembling set) to verify accuracy out-of-sample (Figure 1 Panel B).

We then apply this same model to generate predictions for our main sample of (non-dual-eligible) beneficiaries. First, we form the same 1,770 features, demographics and claims from the first 90 days of enrollment, and apply the estimators formed in the dual-eligible sample to generate predictions of one-year total spending. We interpret these fitted values as predictions on each beneficiary's annual (12 month)spending, had she faced no cost-sharing throughout the entire year.

Of course, we cannot verify these predictions—after all, they are counterfactual—but we can test the rank order correlation between predicted and actual one-year spending, and test for an effect of enrollment month on actual spending conditional on predicted spending. In our main specifications, we convert predicted spending into within-enrollment month percentiles. We do so both because this only requires the assumption that within-enrollment month *ranking* of predicted counterfactual spending correlates with year-end prices (as we show below), rather than needing to pin down the absolute amount, a weaker assumption; and because empirically, we found that results were more consistent this way, perhaps because of year-to-year changes in absolute spending limits.

2.3.3 Final Instrumental Variables Model

These predictions are used to effectively control for annual spending in our IV setup. Concretely, we interact enrollment month with dummies for the predicted counterfactual spending percentile ($\widehat{\mathsf{Spending}}_i$) and calendar year ($\mathsf{Year}_i$), to capture changes in plan design (e.g., the exact coinsurance rates and spending thresholds). We also include all direct effects and two-way interactions of enrollment month, predicted spending, and calendar year, which we summarize in the equation below as $f(\mathsf{EnrollMonth}_i, \widehat{\mathsf{Spending}}_i, \mathsf{Year}_i)$. Finally, we include demographic controls for race and sex (X_i), as well as plan fixed effects (Plan_i). In the second stage, we use instrumented end-of-year prices ($\widehat{\mathsf{Price}}_i^{\mathsf{EOY}}$) to estimate the causal effect of price on December mortality ($\mathsf{Mortality}_i^{\mathsf{EOY}}$), using the same controls as in equation (4).

$$\begin{aligned}\mathsf{Price}_i^{\mathsf{EOY}} = \pi_0 + \pi_1 \mathsf{EnrollMonth}_i \times \widehat{\mathsf{Spending}}_i \times \mathsf{Year}_i + \\ f(\mathsf{EnrollMonth}_i, \widehat{\mathsf{Spending}}_i, \mathsf{Year}_i) + X_i\pi_2 + \mathsf{Plan}_i\pi_3 + u_i\end{aligned} \tag{4}$$

$$\mathsf{Mortality}_i^{\mathsf{EOY}} = \beta_0 + \beta_1 \widehat{\mathsf{Price}}_i^{\mathsf{EOY}} + f(\widehat{\mathsf{Spending}}_i, \mathsf{Year}_i) + X_i\beta_2 + \mathsf{Plan}_i\beta_3 + \epsilon_i \tag{5}$$

This machine-learning approach allows us to estimate the model with a stronger first stage and without non-monotonicity concerns.

2.4 Data

Our main sample consists of a 20 percent random sample of all Medicare Part D enrollees from 2007 to 2012, restricted (as noted above) to those who enroll in Part D between February and September in the year they turn 65. In order to calculate mortality rates over the month of December, we also exclude those who die before December 1 of the same year.

We make some additional sample restrictions. First, we subset to all beneficiaries who become eligible for Medicare because they turn 65, under the Old Age and Survivors Insurance (OASI).[7] This leaves us with 730,279 observations. We then remove all individuals dually-eligible for Medicaid or other low income subsidies (but use them for prediction and falsification), as they face low prices that do not change as a function of yearly spending, which leaves us with 594,672 observations. We keep only individuals in stand-alone prescription drug plans (PDPs) and standard Medicare Advantage (MA) plans for which we can observe plan characteristics.[8] This reduces our sample to 580,236 observations. We also remove all individuals that enroll in a deductible plan as their initial claims vary with enrollment month due to the future price effects (Einav et al., 2015), bringing the sample to 384,538. A series of other minor subsets brings our analytic sample to 358,706 beneficiaries.[9] Roughly half (51.7%) are in PDPs while the remaining individuals are in MA plans.

We make two additional exclusions in our estimation sample, with respect to enrollment month. First, we drop those who enroll in October and later. We do so primarily because as Aron-Dine et al. (2015) note, these beneficiaries are still ramping up their drug utilization, whether because of new coverage or transitioning from a previous insurer. As a result, their December utilization is spuriously low compared to beneficiaries enrolling earlier in the year, who have reached steady state in terms of consumption by December (we show this empirically in Appendix Figure A.2.) In addition, those born in October and later are legally allowed to enroll in January without penalty, because January is in their 4-month initial enrollment period (IEP). Empirically, January enrollment appears to be an outlier in terms of volume of patients enrolling, and January enrollees are observably different from all other enrollment months. So we follow Aron-Dine et al. (2015) (once again!) and

[7] We exclude those who enroll in Medicare before age-64, for disability or end-stage renal disease.

[8] The included MA plan types are HMO, HMO POS, Local PPO, Private FFS, and Regional PPO).

[9] We exclude individuals in special needs plans, those with non-standard ICL locations, and those not residing in the US 50 states or Washington, DC.

exclude them from our sample.[10]

Our measure of price is calculated by taking the average coinsurance rate a beneficiary faces in a given plan, year, and coverage arm. Because some plans are rare in 65 year olds, particularly in the handful of beneficiaries who enter the catastrophic coverage, we take this average across all fills in a plan in the entire Part D universe, not just realized fills made by 65 year olds in their first year. To calculate the coinsurance rate, we divide the sum of all out-of-pocket spending by the sum of all total spending in a given plan, year, and coverage arm.[11] Our measure of mortality is the beneficiary date of death reported to Medicare as required by law, from Social Security and health care providers.

Our analyses of specific drug filling decisions use the Medicare Part D drug claims made by beneficiaries in our sample including fill date, total cost, out of pocket (OOP) cost, and 11 digit National Drug Code (NDC) identifiers. To classify drugs into clinically meaningful categories, we use the RxNorm and RxClass APIs to link NDCs to their corresponding Anatomical Therapeutic Chemical (ATC) codes, a hierarchical system for drug classification. This allows us to map, for example, a claim for Lipitor to the drug class of statins (HMG-CoA reductase inhibitors), within the 'lipid-modifying agents' category. To measure medical diagnoses, procedures, and other kinds of health care utilization besides drugs, we use the subsample of individuals that are enrolled in standalone PDPs (non-MA), for whom we observe all Medicare Parts A and B claims, including diagnoses, procedures, and admit/release dates.

Summary statistics for both the main (non-dual) and dual prediction samples are show in Table 1. Most of the sample is white (89.6%) and female (59.2%). One-year mortality is 0.9%, which is lower than in the entire US population, 1.25% (Arias, 2014), but not surprisingly so given the demographic composition of the sample. While we do not observe diagnoses in MA beneficiaries, the fee-for-service subsample has medical risk factors typical of this age group: looking at the most commonly assigned diagnosis, we find high cholesterol (36.6%), high blood pressure (35.7%), prior heart disease (17.7%), and diabetes (15.9%).

[10] We find evidence that those born in January are less likely to delay enrollment, and that those born in other months (e.g. November) are more likely to delay enrollment to January as opposed to other months.

[11] We also tried calculating the coinsurance rate of each fill, and then averaging across this measure. Empirically, this alternative method over-weights low-cost fills with 100% coinsurance, leading to inflated estimates of plan-level rates (e.g. 50% in initial coverage).

3 Effect of Cost-Sharing on Heath Outcomes

3.1 Spending Prediction

Figure 1, Panel B shows the performance of our spending prediction model. We plot realized 12-month spending against predicted 12-month spending (in $500 bins): first in the dual-eligible holdout set, to check the basic accuracy of our model, and then in our main analytic sample, to verify its correlation with spending in our population of interest. In the dual-eligible holdout, predicted spending closely tracks actual spending throughout the entire spending distribution, as seen by its close proximity to the dashed 45 degree line. In the main sample, we use realized spending over the 12 month period starting with January of calendar year two to evaluate the utility of model predictions. We see that spending is uniformly below that of the dual-eligible sample—unsurprising because the dual-population is sicker, and does not face cost-sharing—but monotonically increasing in predicted spending. We take this as evidence that the model trained in the dual sample rank-orders those in the main sample well, which is our goal (Spearman rank order correlation coefficient of 0.70). Additionally, the model's performance in the main sample does not vary by enrollment month: in Appendix Table B.1, we regress realized on predicted spending interacted with enrollment month (and each individually) and find no significant effects on enrollment month or the interaction term.

3.2 Reduced Form Results

Figure 2 graphically shows our identification strategy. Panel A shows that beneficiaries who enroll earlier in the year are more likely to find themselves in both the coverage gap and catastrophic coverage (e.g., February: 11.8% and 1.7%, respectively) than those who enroll later in the year (e.g., September: 1.5% and 0.2%, respectively). In Panels B and C, we group beneficiaries into three categories of cost-sharing based on their predicted 12-month spending without cost-sharing. We choose these bins in a way that corresponds to plan spending thresholds across years. The 1-70th percentiles of predicted spending, $\leq$ $1,811) are likely to stay in the initial coverage phase, where cost-sharing is low; the 71-97th percentiles, $1,811to $6,393 are likely to approach or enter the coverage gap, where cost-sharing is high; and the 98-100th percentiles, $>$ $6,393, are likely to enter the catastrophic coverage phase, where cost-sharing is once again low.

Panel B illustrates the relationship between enrollment month and December price. Those with low predicted spending, whatever their enrollment month, go on to face the *same* year-end price, because they do not spend enough to approach any spending thresholds, even when enrolling early in the year. Those with intermediate predicted spending enter the coverage gap if they enroll early in the year; but if they are born later in the year, they face *lower* December prices, because they never reach the coverage gap. Finally, those with high predicted spending enter the catastrophic phase if they are born later in the year; but if they are born earlier in the year, they face exogenously *higher* December prices, because they remain in the coverage gap rather than reaching the catastrophic phase. This compactly illustrates the countervailing effects of enrollment month on price, mediated by the opposite sign of the coverage gap vs. the catastrophic coverage, that together violate the monotonicity assumption. While binning beneficiaries into these three categories does not completely solve our monotonicity problem—within a bin, particularly near the thresholds, the effects of enrollment month might be complex—they do show intuitively how predicted spending increases precision in our instrumented estimates of price.

Panel C shows our reduced form result of mortality on enrollment month. As expected, there is no relationship between December mortality (measured between December 1 and 31) and enrollment month in the lowest spending bin, where there is no price variation by enrollment month. Among beneficiaries predicted to be in or near the coverage gap, however, earlier-month enrollees—facing higher end-of-year prices—have higher December mortality. Among those predicted to be in or near catastrophic coverage, by contrast, earlier-month enrollees—facing lower end-of-year prices—have *lower* mortality. We report the regression version of Figure 2 in Panel A of Table 2. For enrollees predicted to be in the coverage gap, each additional enrollment month reduces the December coinsurance rate by 2.27 p.p., while mortality decreases by 0.0113 pp (on a base of 0.132 p.p.).[12] For those predicted to be in catastrophic coverage, each additional enrollment month raises the December coinsurance rate by 1.52 p.p. and mortality by 0.0465 p.p. (on a base of 0.279 p.p.).[13]

[12] For those predicted to be in the initial coverage phase, we do find a very small negative effect, driven by less than perfect predictions of which beneficiaries will end up in the initial vs coverage gap phases. For example, if we remove those in the 51-70th predicted spending percentiles, who are most likely to instead be in the coverage gap, there is no significant effect of enrollment month on end-of-year price in the predicted initial coverage bin. However, because the bins correspond (roughly) to actual plan spending thresholds, we chose not to modify them to avoid arbitrariness.

[13] Because monotonicity is not satisfied within-category, rather than dividing the reduced form mortality effects from column (3) by the first-stage coinsurance effects from column (2), we defer to our 2SLS specification below.

3.3 Two-Stage Least Squares Results

We now turn to estimating the health effects of price differences using a 2SLS framework. This is important because the three predicted spending bins we have used so far (Figure 2 and Table 2, Panels A) are very coarse ways to measure the actual year-end price faced by different beneficiaries. For example, in Panels B and C of Figure 2 the linear coefficients on enrollment month are estimated within very broad spending bins: they combine everyone from the same enrollment month, whether they are predicted to be just past a spending threshold, or very far past it. In other words, wide predicted spending bins combine people who might face (or anticipate) very different actual year-end prices. As noted above, this mis-measurement decreases precision of the first-stage and reduced-form estimates; for the top-spending few percentiles, it also risks running afoul of non-monotonicity.

The very fine-grained predictions on spending from our machine learning model allow us to improve on this considerably. As Figure 3 shows, the relationship between year-end coinsurance rate and enrollment month varies considerably by percentile of predicted spending. So there is no reason to restrict ourselves to three bins. The intuition here is the same as the three-category plot in Figure 2: the linear trend shown in each panel is represented as a coefficient and graphed on the y-axis, for each percentile of predicted spending. A coefficient of zero means that there is no first stage, and a positive (negative) coefficient measures the increase (decrease) in the December coinsurance rate from enrolling one month later.[14] The advantage of this approach is clear, for example, in Figure 3: in Figure 2 above, we grouped all beneficiaries from the 71st to 97th percentiles of predicted spending together, as those predicted to enter the coverage gap. But Figure 3 shows that the year-prices these beneficiaries face respond to enrollment month very differently, depending on their fine bin of predicted spending. It also shows precisely where the sign of the instrument's effect on price changes sharply, as the highest-spending beneficiaries approach the catastrophic phase.

This motivates us to estimate the first stage of the 2SLS model in equation (4) using percentiles of predicted spending, and allowing the effect of enrollment-month to vary in these very granular bins. We then use the result to estimate equation (5). Our first stage F-statistic is 210.7. In

[14] As noted above, we use percentiles of predicted spending rather than absolute levels because percentiles performed better for predicting year-end prices across multiple years of data. Spending levels at or below the 70th percentile are grouped into deciles because at lower levels of predicted spending, there is not enough variation to define unique within enrollment month percentiles. It is clear however, that there is no heterogeneity in the association between price and enrollment month in this low predicted spending group.

Panel B of Table 2, we show a large and positive effect of higher prices on mortality: for every p.p. increase in coinsurance rate we estimate a 0.0043 p.p. increase in mortality. This estimate is smaller than what we had obtained in the three-bin analysis (Panel A) but is more accurate because it reflects a tighter alignment of the first-stage with the second. For example, in Panel B, we allow for the possibility that enrollees with larger first-stages (e.g., larger cutbacks) are not the same ones experiencing a larger second stage (i.e., larger mortality), which could inflate the IV estimates in Panel A.

To convey the scale of our results, we translate the two-stage estimate into a concrete counterfactual: the total mortality impact of high cost-sharing under the standard non-linear Medicare Part D design, relative to a flat, low-cost-sharing design. We can approximate this within the constraints of our data by comparing mortality between February and September enrollees. February enrollees experience something quite close to the standard Part D design (i.e., after the first year of enrollment), under which all beneficiaries have 12 months to reach plan-defined spending thresholds; of course, February beneficiaries face one less month than usual. Conversely, September enrollees largely experience the simple, low-cost-sharing design of initial coverage phase. To estimate this counterfactual, we calculate the median difference in the (instrumented) coinsurance between February and September enrollees: 11.0 p.p.. Multiplied by the 2SLS estimate, translates to a 0.048 p.p. mortality difference, on a base monthly mortality rate (in December) of 0.147 points—a 32.7% increase. These estimates are presented in Panel C of Table 2.

3.4 Falsification Tests and Robustness Checks

There are some reasons to worry that the variation in mortality we document might be driven by birth (or enrollment) month via channels other than the price of prescription drugs. Small but statistically significant variation in long-term health outcomes by birth month, whether driven by seasonal disease patterns or selection, has occasionally been documented in rich countries, though most of the evidence comes from post-natal outcomes in developing countries (Currie and Schwandt, 2013; Buckles and Hungerman, 2013; Zhang et al., 2019). Our approach provides several convenient falsification tests of the exclusion restriction, all arguing against the existence of an independent source of variation in mortality due to birth or enrollment month (as well as differential selection into enrollment month conditional on birth month, as discussed above).

Two facts from our reduced form results in Figure 2, Panel C are already preliminarily reassuring. First, among those predicted to have low spending and thus to be unaffected by spending thresholds, we see no relationship between enrollment month and mortality. Second, we see that mortality *increases* in enrollment month for those predicted to be in the coverage gap, but *decreases* in enrollment month for those predicted to be in the catastrophic phase. If early-year enrollees were systematically sicker (less sick) than late-year enrollees, we would expect a negative (positive) relationship between mortality and enrollment both in *both* groups of patients; whether they are predicted to spend more or less than plan thresholds should not matter.

Next, in Table 3 we examine the relationship between enrollment month and December mortality in two other settings lacking the specific policy quirks we exploit in our main specifications, where we expect enrollment month to have minimal effect on outcomes. First, we test for effects of enrollment month on December in 65-year old dual-eligible enrollees, who face no cost-sharing and thus low prices throughout the year (Panel A). Second, we test for differences in mortality in our main sample, during the 'ramp-up' period: we use data from the first 60 days of enrollment to predict spending (instead of 90 days in our main analysis), and test for mortality differences by enrollment month in the third month of enrollment. If enrollment month is correlated with mortality by mechanisms other than price, we would have expected to see differences before December as well as during December.[15] In neither of these settings do we find significant reduced-form relationship in any predicted spending grouping.

In Table 4, we also explore the robustness of our 2SLS estimates to a number of key design choices: either changing the sample that we use for estimation, or altering the specification of our first stage equation (4). Our 2SLS estimates are robust to not including plan fixed effects, using shorter (30 or 60 day) periods of initial claims to determine the the spending prediction, using birth month (as opposed to enrollment month) as the instrument, and restricting to only those that enroll in the month they turn 65. We also show significant effects using longer outcome periods.[16] Additional robustness checks are presented in Appendix Figure C.1. Together, these results give us

[15] This sample includes our main sample along with an additional 4131 beneficiaries that either died or lost Part D coverage before December. We do note however, that while spot-prices may be equivalent over this period (61-90 days since enrollment), there could be future-price effects because the early-year enrollees are more-likely to reach the coverage gap and catastrophic coverage.

[16] Estimates using outcomes past December of year 1 are more difficult to interpret because annual-spending levels (that determine price) reset on January 1, so in year 2, prices are equal across enrollment months.

confidence in the stability of our 2SLS estimates.

4 Understanding the Demand Response

In this section we examine the underlying demand response at a granular level: what are the specific decisions patients make in response to price shocks? We are particularly interested in whether the large mortality effect we document above can be plausibly be tied to the nature and quantity of the drugs patients choose to drop.

As a first step, we characterize changes in several aggregate measures of utilization in response to price increases, like the probability of filling any prescription, the number of fills, and total spending on prescription drugs. We replace December mortality as the outcome of interest in the second stage equation (5) with these measures, all in the same year-end period that we measure mortality. The estimates are reported in Table 6. We find that a 1 p.p increase in coinsurance rate leads to a 0.20 p.p. decrease in the likelihood of filling any medication, 0.031 fewer claims, and a $5.54 reduction in total spending.

4.1 Decomposition of Mortality Effect by Individual Drug Classes

Next, we examine the specific drugs patients drop, focusing on potentially life-saving medicines that might underlie changes in mortality. We study eight common "high value" drug classes, for which the clinical literature has established known, large (average) mortality effects. To estimate this class-specific demand response we again replace December mortality in Equation (5) with the number of December fills made in each class. We report these 2SLS estimates in Table 6 Panel B, along with the percentage of beneficiaries on a class and the average number of fills made in December (both for predicted compliers). Patients cut back substantially on all 8 classes: in response to a 1 p.p. increase in coinsurance, patients make fewer fills for cholesterol-lowering drugs (statins—0.004 fills), antihypertensives (ACE inhibitors—0.0009, β-blockers—0.0011, thiazide diuretics—0.0010, calcium channel blockers—0.0008, angiotensin receptor blockers—0.0013), diabetes drugs (non-insulin blood glucose lowering drugs—0.0023) and respiratory drugs (inhalants—0.0015).

Knowing the degree of cutbacks in each class allows us to perform a simple validation exercise where we estimate the share of the 32.7% increase in mortality that can be accounted for by

cutbacks in drugs that have RCT estimates of mortality. Using statins as an example, we calculate the reduction in statin usage by compliers (those in the top 30% of predicted spending), and arrive at an 11.7% reduction in statin fills (column 4), which we multiply by the clinical trial mortality effect of -26.7% (column 5).[17] Statins cutbacks predict a mortality increase of 3.1%—but only if everyone in our sample were eligible for a statin and cut back. To scale this to the relevant population, we need to know who is eligible for a statin (which is not simply the realized December rate of filling statins, which is endogenous). To estimate a lower bound on eligibility, we use the fraction of our sample who filled a statin at any point in the first 90 days of enrollment, and scale the mortality effect by this fraction, 0.584, for statins. This implies that the decrease in statin utilization we observe should translate into an increase in mortality of 1.82%. Repeating with the other 7 drug classes, and summing across each drug's individual mortality contribution we find that cutbacks in these 8 classes account for a 8.44% increase in mortality, which is 25.8% of the overall mortality increase of 32.7%.

We chose these classes solely because we could identify estimates of their effect on mortality from placebo-controlled randomized trials, which are closest to our quantity of interest: the effect on mortality of taking (dropping) a drug. But these classes only account for 42.1% of the overall reduction in fills, and they are by no means the only drugs to have large effects on mortality. Antibiotics, insulin, corticosteroids (for acute exacerbations of emphysema), antidepressants, and many other life-saving drugs are in widespread use but, for ethical or historical reasons, have rarely been tested in placebo-controlled trials. In addition, the above calculation assume a simple additive mortality benefit of drug initiation. But the benefits of taking multiple drugs (e.g. a statin and hypertensive) may be greater than the sum of their parts, and the harmful effects of suddenly stopping a medication (as in our data) may be asymmetrically greater than the beneficial effects of starting it (in a trial). Finally, as we explore below, the highest-risk patients may be more likely—not less—than lower-risk patients to forgo their drugs in response to price increases.

While larger, the magnitude of the mortality effect we detect is of the same order as a growing literature on the effect of health insurance on mortality. Most closely related to our study is Huh

[17] We convert intent-to-treat (ITT) estimates reported in clinical trials to treatment-on-the-treated (ToT) by dividing by compliance in the trial, to be comparable to our 2SLS estimates. More information on the clinical trial estimates can be found in Appendix Table D.1. To estimate the decrease in statins we multiplied column 2 by the median change in price from the first stage (11.0p.p.), and divided by the average number of December fills (column 1).

and Reif (2017), who estimate effect of introduction of Medicare Part D on mortality by comparing eligible 66 year-olds to ineligible 64 year-olds, and find annual mortality reductions among compliers of 8.2%. Other studies by Miller et al. (2019) (9.4%) and Goldin et al. (2020), (10.1%) exploring the effects of any insurance and Medicaid respectively, likewise find effects of a similar magnitude. Our estimates benefit greatly from the addition of machine learning predictions, which increase the precision of our first stage, and also allow us to hone in on a very specific population in whom effects are larger. We show this by estimating the basic 2SLS model (equations (2), (3)), where enrollment month is *not* interacted with bin of predicted spending. While the coefficient (0.00401 p.p.) is similar to our main 2SLS coefficient, the standard error (0.00329) is considerably larger.

4.2 Channels Besides Drugs: Physician and Hospital Visits

This discussion has emphasized the role of drug consumption in reducing mortality. But in response to increased drug prices, patients may choose to spend more out-of-pocket for drugs and cut back on other preventative care such as physician visits. We tested for such cutbacks by replacing December mortality with measures of hospital (Part A) and physician visit (Part B) spending in equation 5.[18] Results are presented in Appendix Table D.3. We find no discernible differences in either spending measure which argues against other channels besides medications in impacting patient health.

4.3 Implications for 'Polypharmacy'

These findings have important implications for the widely-held concept of 'polypharmacy' in the medical literature: the idea that being on many drugs has negative health effects, due to adverse drug-drug interaction effects. Several prior studies (reviewed in Hajjar et al. (2007)) use the cross-sectional, positive, relationship between the number of drugs and subsequent mortality to argue that more drugs are harmful for patients. In response to concerns regarding confounding, they note that the relationship persists even after adjusting for many measured potential confounders.

By contrast, our causal estimates provide evidence that being on higher numbers of drugs reduces mortality. This is in line with what we might expect based on clinical trials, but contrary to the results we would get from a cross-sectional analysis. In Table 5 we contrast OLS and 2SLS estimates for the effect of the number of drugs a patient fills, and the total amount she spends, on

[18] These models are estimated using the subsample of non-MA enrollees for whom we observe Parts A and B claims.

mortality. OLS estimates indicate that each additional December drug fill and each additional dollar of total spending increases mortality by 0.018 p.p and 0.00011 p.p., respectively The positive OLS association persisting even when controlling for demographic characteristics and machine learning predictions on a patient's number of medicines, which in this setting mimics the usual approach of a propensity score in the medical literature.[19]

2SLS estimates, on the other hand, find that each additional drug fill and each additional dollar of total spending *decreases* mortality by -0.099 p.p. and -0.00079 p.p., respectively. This pattern holds true even if we define the endogenous variable as an indicator for being on five or more drugs, which is a common cutoff for polypharmacy. We view these 2SLS estimates as suggestive evidence for the intuition that the positive relationship between fills and mortality found in OLS regressions is due to confounding. Practically, using the number of fills as the endogenous variable in the first stage, rather than the coinsurance rate, likely violates the exclusion restriction: while we find no evidence of this (as noted above), other healthcare seeking behaviors may be affected by price changes induced by enrollment months and in turn affect mortality. Using spending as the endogenous variable may be better in this regard, since it is more closely tied to coinsurance rate, though similar caveats apply.[20]

The positive cross-sectional relationship between number of drugs and mortality may provide a clue to why the large mortality effects we find have not previously been noted. Mortality due to cutbacks on medications occur in patients who consume more drugs (concretely, in our sample, we consider those in the top 30% of predicted drug spending). These patients have many other competing causes of mortality; after all, that is why they are on many medications to begin with. It may simply be difficult, for practicing physicians and for researchers, to causally tease out the effect of cutting back from the effect of the underlying diseases for which the medications are prescribed. In addition, our 2SLS estimates of the mortality effect of cost-sharing is highly unlikely to be noticed by a physician with a typical panel size of 1500-2000 patients (Raffoul et al., 2016). Detecting a mortality increase of the size we detect vs. a baseline rate of 1% would require a panel of almost

[19]We use the same machine learning setup as described above to predict total spending, but instead use the number of fills a patient has in 12 months as the response variable. The prediction is accurate, with a Spearman rank-order correlation coefficient (with number of claims in December of year 2) of 0.47.

[20]For the OLS regressions on the number of fills, we use mortality in January of year 2 as the outcome because those that die in December will mechanically have fewer fills because they are alive for less days in the month. In the 2SLS regressions we use December mortality, as in our main model from Table 2.

30,000 patients, in addition to perfect recall and randomly-assigned prices.

4.4 Heterogeneity in the Demand Response by Patient Risk and Income

Of course, just because drugs appear high-value on average, based on treatment effects from clinical trials, does not mean they are high-value to the patient who drops them. Treatment benefit is highly heterogeneous across individuals (Chandra and Skinner, 2012; Chandra and Staiger, 2020), and the simplicity of the 'high-value' and 'low-value' labels belies the fact that very few medical interventions can be sensibly characterized in this manner (Mullainathan and Obermeyer, 2019). So it is premature to say that this behavior implies that patients are making mistakes, in the sense of exhibiting behavioral hazard. Indeed, neoclassical models, such as a Roy model of patient decision making with heterogeneous treatment effects, would predict that cutbacks would be concentrated among individuals in whom the drug is less clinically valuable. The corollary of this is that those with the highest potential benefit should be willing to pay more for a drug, and thus less likely to cut back when the price increases. To distinguish between these views, it would be useful to have a measure of the (health) benefit an individual patient might get from a given treatment.

A key insight comes from the medical literature, which documents that in some settings a patient's benefit from preventive medicines is proportional to her baseline risk of the outcomes the drug prevents. There is strong evidence to support this assumption, as well as clinical guidelines guiding physician behavior, for cardiovascular drugs to prevent atherosclerotic cardiovascular disease (e.g., statins, 58.4% of compliers in our sample, and several antihypertensives, e.g., ACE inhibitors, 29.8%).[21] There is less strong evidence, but medical consensus and biological plausibility, for two other common drug classes: diabetes drugs that lower blood sugar (e.g., insulins and oral medications—26.3%) and prevent complications of diabetes; and respiratory medicines for such as inhalers, to counter exacerbations of obstructive lung disease (22.1%). We emphasize that we do not need to assume this model is optimal, in the sense that it captures 'true' treatment heterogeneity—but to the extent that patients or doctors believe the basic principle of high risk to high benefit is

[21] This is often simply assumed by doctors and incorporated into risk scores, like the American College of Cardiology's 10-year risk calculator that is used to allocate treatments for cardiovascular disease. But there is also growing evidence that this treatment heterogeneity is real, from clinical trials. For example, the JUPITER, HOPE-3, CARDS, and ASCOT trials (reviewed by Bibbins-Domingo et al. (2016)) show 30-50% larger absolute risk reductions from statins in groups with higher predicted risk of heart disease, whether defined by age, diagnosed risk factors (e.g., diabetes), or biomarkers (e.g., LDL, CRP). Studies of polygenic risk scores show similar heterogeneity, with higher-risk participants getting nearly three times the absolute risk reduction (Natarajan et al., 2017).

correct, our results should identify patients who believe they would benefit from a given treatment.

For each of these classes, we compile a list of observable adverse outcomes, Y_G, one for each source of preventable risk G: heart attack and stroke for cardiovascular medicines, diabetic complications (e.g., foot amputation) for diabetes medicines, and pulmonary collapse requiring mechanical assistance for respiratory medicines.[22] We then form separate predictive models for each outcome, in a similar fashion to the one-year spending prediction described above, using all dual-eligible enrollees. We use initial claims (again, within 90 days of the first day of a beneficiary's birth month) to predict the likelihood of acute events/complications, in a training sample restricted to those who are *not treated* with the particular medications in question (e.g., when predicting risk of heart attack, we exclude patients on statins, etc.). We do this to form a prediction on the risk of complications if untreated (i.e., in potential outcomes notation, we wish to estimate $E[Y_{G0}|X]$, not $E[Y_{G1}|X]$), to more closely answer the question, without a given medication, what would my risk of complication be?[23]

To model the class-specific demand response by risk, we estimate the following first and second stage equations, similar to equations (4) and (5), but with both the instrument and instrumented price interacted with an indicator for high predicted risk (top one-third of the sample, where risk begins to increase; see appendix Figure B.1).

$$\begin{aligned}\mathsf{Price}_i^{\mathsf{EOY}} = \pi_0 + \pi_1 \mathsf{EnrollMonth}_i \times \widehat{\mathsf{Spending}}_i \times \mathsf{Year}_i \times \widehat{\mathsf{Risk}}_{i,g} + \\ f(\mathsf{EnrollMonth}_i, \widehat{\mathsf{Spending}}_i, \mathsf{Year}_i, \widehat{\mathsf{Risk}}_{i,g}) + X_i\pi_2 + \mathsf{Plan}_i\pi_3 + u_i\end{aligned} \tag{6}$$

$$\begin{aligned}\mathsf{Fills}_{i,g}^{\mathsf{EOY}} = \beta_0 + \beta_1 \widehat{\mathsf{Price}}_i^{\mathsf{EOY}} + \beta_2 \widehat{\mathsf{Risk}}_{i,g} + \beta_3 \widehat{\mathsf{Price}}_i^{\mathsf{EOY}} \times \widehat{\mathsf{Risk}}_{i,g} + \\ f(\widehat{\mathsf{Spending}}_i, \mathsf{Year}_i, \widehat{\mathsf{Risk}}_{i,g}) + X_i\beta_3 + \mathsf{Plan}_i\beta_4 + \epsilon_i\end{aligned} \tag{7}$$

Here $\mathsf{Fills}_{i,g}^{\mathsf{EOY}}$ is the number of December fills for disease class g by person i.

Table 7 presents these results. We find that the highest risk beneficiaries cut back as much, if not more, in all three major drug–risk pairs. This trend is especially pronounced for cardiovascular

[22] Specifically, we use ICD9 codes 410-411 and 433-435 for cardiovascular events, and 5188, 7991, 9604, and 9607 for respiratory events. For diabetes complications, we use all codes in the Diabetes Complications Severity Index (Young et al., 2008)

[23] Naturally this choice of prediction target also induces selection bias: we form predictions on Y_{G0} in patients selected into treatment status $T = 0$, but then wish to generate predictions on in patients with arbitrary treatment status. In our setting, our predictions are likely to underestimate risk on average, because doctors select patients into treatment $T = 1$ on the basis of higher risk.

drugs: for each p.p. increase in coinsurance rate, low-risk patients make 0.003 fewer fills of cardiovascular drug, while high-risk patients make an additional 0.009 p.p. fewer fills. This finding is incompatible with a moral hazard model of behavior, which would predict that patients at high risk of a cardiovascular event, who would benefit the most, should also be willing to pay the most for treatment. We find similar trends for diabetes and respiratory drugs, we find that those in the top one-third of the risk distribution cut back roughly the same (0.002 and 0.002 fills per 1 p.p coinsurance rate, respectively) as those in the bottom two-thirds (0.003 and 0.002).[24]

Our framework also provides an opportunity to examine heterogeneity in the demand response by income, albeit less precisely than by risk. In column (3) of Table 7 we replace the indicator for being in the highest third of risk with an indicator for being in the highest third of zip code income, a summary measure of SES that may be better correlated with permanent income and lifetime wealth than transitory income. Here, we find that the response of Medicare enrollees from the highest income zip codes is no different than that of those from lower income zip codes, suggesting that socio-economic circumstance, as proxied by zip code income, is not a mediator of the demand responses we observe. Of course, there is variation in income within zip codes, often quite a bit, so this does not by any means rule out income effects or liquidity constraints. However, to the extent that we see similar behaviors in rich and poor areas alike, it forms some upper bound on how important these effects can be on average.

5 Observations on Patient Behavior

For a patient at the pharmacy desk, who has just learned that her basket of medications has a new and higher price, a standard view of the optimal choice calculus is: for each of k prescribed medicines in her basket, she should compare the marginal benefit of each to its marginal cost, and fill only the $m : m \leq k$ above a certain value threshold, subject to some total budget constraint. But this exercise raises a number of questions. First, what are these k medications? The literature suggests that even such simple-appearing questions can be surprisingly challenging. In some settings, as

[24] While in the main text and Table 7 we present demand response estimates in terms of number of fills, we are also able to calculate the percentage reduction for compliers by dividing the 2SLS estimate by the mean number of fills (by risk type) in December. For cardiovascular drugs, we find that each p.p. increase in the coinsurance rate leads to a 0.46% and 0.82% reduction for low- and high-risk patients respectively. For diabetes (respiratory) drugs, low risk patients cutback 0.96% (1.7%) while high risk patients cutback 1.1% (1.6%).

few as one-third of elderly patients can correctly recall what their medications are, let alone what their purpose might be (Evans and Crawford, 1999). More detailed questions about the marginal benefit of a given medication, while fundamental to the choice calculus, would challenge even a sophisticated expert: for example, what is the cumulative health hazard of failing to fill a drug until prices reset on January 1? And what is the dollar value of that health loss? Does the decision for one medication change the cost-benefit calculus for other medications? What is the airspeed velocity of an unladen swallow?

5.1 Is this Behavioral Hazard?

Given the complexity of the choice calculus, and the specific possibility of errors in the estimation of the marginal benefits, we now turn to the question of interpreting whether our results are broadly consistent with behavioral hazard (Baicker et al., 2015). As we outline above, there are two ways to approach this determination. The first and most common strategy in the literature is to compare patients' choices to expert opinion. Behavioral hazard is declared if patients appear to be making ill-considered choices, typically cutbacks in care that appears to be high-value like preventive screenings and medications (Brot-Goldberg et al., 2017; Chandra et al., 2010; Geruso et al., 2020). But cutbacks in care utilization—even if the care appears high-value—are an indirect measure of welfare (Baicker et al., 2015), and need not represent errors (see Chandra et al. (2019)). The second strategy, which we pursue here, is to examine whether the health offsets caused by higher prices have magnitudes that can are reconcilable with reasonable estimates of the value of statistical life. We emphasize that it is not possible to identify behavioral hazard without an assumption about the value of life. If patients were rational, in the sense of being able to trade off higher prices against lower life, the size of the mortality effect should should reflect the value of life associated with a small price increase.

A simple calculation illustrates why this is unlikely to be the case. Average life expectancy, based on the age and sex composition of our sample, is 19.2 years (17.7 years for men and 20.3 years for women, Arias (2014)). Consider a rational patient who is contemplating a risk of dying in December of her 65th year. We can view this patient as contemplating a gamble where she trades off her remaining life years (all of which she loses if she dies) with probability ($\beta_1^{Mort} = 0.000043$, from Table 2), in exchange for a certain amount of money. If we knew how much money a patient would save, we could divide the savings by the forgone expected life-years to learn what

value patients implicitly place on their life-years. To calculate what patients save, we replaced the mortality outcome in the 2SLS model (equation (5)) with total drug spending in December.[25] We find that a 1 p.p. increase in coinsurance decreases total drug spending β_1^{Spend} = -$5.54. So in our setting, enrollees act as though they value life at $\beta_1^{Spend}/(\beta_1^{Mort} * 19.2)$ = $6,628 dollars per life-year. This is an order of magnitude below usual valuations of around $100,000 to $150,000 for a life-year. Approaching the calculation in reverse, at a $100,000 valuation, patients would need to believe they had only 1.28 years left to live. Naturally, it is possible that there are other unobserved dimensions of utility that we do not observe. But the size of these would have to be extremely large to offset the implied values of life that we do observe.[26]

One possible explanation for these low valuations is the 'harvesting' hypothesis. The deaths we observe might concentrated in a small subset of high-risk beneficiaries, whose life expectancy was sufficiently low to make this tradeoff reasonable. Under this hypothesis, we would expect to see that subsequent mortality for remaining beneficiaries would be *lower* than average (since deaths that happened in December would not happen in the later months). Appendix Table C.1 shows that there is no indication that mortality reverses over the second calendar year of coverage: two stage estimates of monthly mortality risk as a function of instrumented December spending are entirely flat over the next year.

5.2 Zeroes and Ones

Given that mortality increases well beyond what we would expect from the dollar value of cutbacks, our results suggest that patients on average make errors in deciding which drugs to fill. But these facts give us scant indication of the nature of these mistakes. We now present some simple descriptive results on behaviors that we observe in our setting, that may provide clues to the underpinnings of errors in complex decisions.

A striking finding is a substantial mass of people who, when faced with higher prices at year-

[25] Here we are conservatively assuming that patients would have to pay the full total cost of the drug out-of-pocket. Because on average across plan arms they only pay some fraction, then the patient's life-year valuation is in fact lower—so we view $6,628, is an upper bound for this estimate.

[26] We might be concerned that average life expectancy would be lower for the compliers in our sample, who despite being only 65, are on more medicines and sicker. While we do not have specific estimates applicable to our population, we do know that life expectancy in those with atherosclerotic cardiovascular disease at age 70, for example, is still 8.79 and 11.0 years for men and women (Peeters et al., 2002), respectively. Even using these values, the implied life-year valuations are far lower than any commonly cited in the literature.

end, choose to fill *no* drugs. We show this simply by inspecting the raw number of drugs filled by patients predicted to be in the coverage gap. Figure 4, Panel A, compares the number of December fills m for February vs. September enrollees (who face full cost-sharing, vs. the lowest cost-sharing in our sample, respectively). The left panel shows data for the full sample, while the right panel restricts to those on a larger number of drugs at baseline, using the number of claims filled in the first 30 days of enrollment which we denote $\hat{k}$.[27] In both graphs, the largest difference between February and September distributions is at $m = 0$—in other words, patients who appear to drop *all* their medications. The difference is particularly striking in those with higher values of $\hat{k}$. Under a rational model of consumption we would have expected to see these patients dropping their $(k - m)$ marginal drugs: for example, someone on 5 drugs might have filled the 3 or 4 drugs that now fall within their willingness to pay based on b/c. Instead, they fill none—and this tendency to fill zero drugs textitincreases in $\hat{k}$.[28]

Panel B of Figure 4 builds on the descriptive results from Panel A, again for patients predicted to be in the coverage gap. It presents reduced form estimates of the effect of enrollment month on the likelihood of filling a certain number of drugs. The heatmap presents estimates from a set of regressions where the outcome variable is an indicator for having m claims in December (rows: m = 0-1, 2-3, ...). This is regressed on enrollment month interacted with bins of claims in the first 30 days (columns: $\hat{k}$ = 0-1, 2-3, ...). Each line thus corresponds to a separate reduced form regression, and each column shows the coefficient on enrollment month interacted with k_i (how many drugs a beneficiary should be on, estimated in their first month of claims). Negative estimates (shaded in red, by magnitude) correspond to filling behaviors that are more common among those who face high prices in December, while positive estimates (shaded in blue) correspond to behaviors of

[27] The true number of baseline drugs k a patient is prescribed may be higher, but is unmeasured: the steady-state number of drug fills later in the year is endogenous to enrollment month. So we use $\hat{k}$ as a proxy for k; our analyses assume only that $\hat{k}$ and k are positively correlated. That said, the bottom panel of Figure 4 shows that patients facing low prices in December fill $\hat{k} \approx k$ drugs, suggesting it is a reasonably good proxy.

[28] The fact that those on more drugs at baseline are *more* likely to end up on zero drugs, not less, intuitively argues against the possibility that the excess zeros we observe is due to a floor effect (i.e., each drug is dropped with some constant probability, and zeros are more common simply due to left-censoring). We formalize this intuition by simulating a distribution of December claims by September enrollees under this model. Starting with the September distribution, we drop each drug filled by each beneficiary with $p = 0.137$, the empirical difference between February and September filling likelihood. This simulates a scenario where beneficiaries have some likelihood of dropping each drug, with no specific preference for dropping all their drugs. We can then compare this scenario to the observed scenario in terms of the number of zeros and ones. We find that under the actual distribution there are significantly more February enrollees with zero or one claim than under the simulated one; this is not the case for any other number of claims (see appendix Table D.4).

enrollees who face lower prices in December. Two visual patterns stand out: first, the blue diagonal line indicates that those facing low prices in December tend to fill the same number of drugs we estimate they are prescribed: $m \approx \hat{k}$. Second, the red-horizontal line shows that, irrespective of how many drugs they are prescribed, those facing high prices in December are substantially more likely to simply fill 0 or 1 claims in December: $m \leq 1$. This pattern is much less pronounced for other numbers of December claims. To put these numbers into the same scale as our previous results, we estimate a 2SLS model, similar to equation (5), where the second stage outcome is an indicator for having less than, or equal to, a specific number of December claims. We find that a one p.p. increase in coinsurance rate makes an individual 0.20 or 0.17 p.p. more likely to have zero or one fills, respectively. We report the results of these 2SLS regressions, along with indicators for higher numbers of claims in Appendix Table D.5.

Finally, we explore whether patients who cutback to zero drug fills appear are among those most likely to experience higher mortality on the margin. To assess this, we estimate an additional 2SLS model, where the outcome is December mortality and the endogenous variable in the first stage is an indicator for having specific number of drug claims in December (i.e., instead of coinsurance rate, we estimate separate models for $m = \{0, 1, ..., \geq 5\}$). To be clear, we do not present this as a formal causal estimate in any way: clearly the exclusion restriction is violated, given the many other behaviors that result from price increases besides the decision to fill m drugs. Rather, we wish to explore whether the marginal likelihood of reducing consumption to m drugs is correlated with the marginal likelihood of mortality; we simply choose a 2SLS coefficient, of mortality on instrumented likelihood of filling m medications, as a convenient measure of this correlation. With these caveats in mind, we present two pieces of suggestive evidence in Appendix Table D.5. First, we confirm that higher prices increase the likelihood of filling zero or one drug in December, and to a lesser extent, two drugs; and reduces the probability of filling five or more drugs. Second, marginal mortality increases do appear to be most correlated with the marginal likelihood of filling no drugs, relative to other numbers of drug fills. This implicates, but does not establish, the decision to fill zero drugs as one potential source of increases in December mortality.

5.3 Relationship to Known Behavioral Models

There are several classes of behavioral models that could explain why patients might drop more drugs than is optimal. We broadly distinguish between them based on how each might affect the patient's engagement with the cost-benefit calculus. First, the weights for this calculus may be distorted such that costs are over-weighted relative to their true value. For example, if the patient arrives at the pharmacy counter to find that her drug basket has shot up in cost relative to her expectations, it could cause her to overweight costs relative to benefits, as predicted by salience theory (Bordalo et al., 2013, 2020). If such deviations from previously set reference points are viewed as losses, rather than costs, the effect could be similar (Kahneman and Tversky, 1979). Present bias (Laibson, 1997; O'Donoghue and Rabin, 1999) could likewise cause patients to over-weight present costs over future benefits. Whether context-dependent (e.g., salience) or not (e.g., present bias), such mis-weightings could easily distort the calculus in the direction of dropping too many drugs. Second, patients may use heuristics, effectively substituting simpler problems for the more difficult full calculation of marginal costs and benefits (Tversky and Kahneman, 1974). For example, they may use a simple decision rule like filling *the* most important 1 prescription; they may drop the most or least expensive drugs, or the latest drug that had been added, or an earlier one in the regimen. Third, patients could disengage from the cost-benefit calculus altogether, under a variety of models. Inattention (Handel and Schwartzstein, 2018; Gabaix, 2019) could cause patients to fail to process all the information required to know true benefits and costs. Choice fatigue (Augenblick and Nicholson, 2016; Iyengar and Kamenica, 2010) or failures of 'meta-reasoning,' where patients judge a problem as 'unsolvable' and simply give up (Ackerman and Thompson, 2017), could work similarly. Of note, meta-reasoning necessarily involves heuristics, since the only way to truly know whether or not a problem is solvable is to solve it.

While our observational data prevent us from adjudicating between these accounts, we can inspect the data for specific patterns of utilization that might suggest heuristics or inattention. For example, under these models, we might expect patients to stick with some 'default' action, instead of making a considered decision in response to new information on prices. This raises the question of what that default might be, in the case of a patient who is used to filling k drugs, then sees her basket price increase. Perhaps the most natural candidate is to stick with the same

basket of drugs; but of course this does not fit with the the large changes in consumption we find when prices increase. One possibility is that some patients view being on no drugs as a default of sorts. Another potential default could be choosing to spend the same amount out-of-pocket when prices increase, which is of course different from choosing to be on the same basket of drugs. But in fact, patients change their out-of-pocket spending amounts considerably in response to price increases. On average, patients *increase* out-of-pocket spending, by \$14.6 in response to a 11.0 p.p. coinsurance increase. Naturally, because drug consumption decreases on average, this increase in spending is not enough to keep total drug consumption at its previous levels: consumption decreases by \$61.20 in response to the same price shock.[29] A final possibility in this vein is choice fatigue. In several real-world and experimental settings, having more choices pushes decision makers towards simpler options, even when they are more risky (Iyengar and Kamenica, 2010). At the extreme, the proliferation of choice can cause people to avoid decisions entirely, as in Augenblick and Nicholson (2016) where voters are more likely to abstain after having faced more prior votes. Choice fatigue may be particularly relevant for patients on many drugs who choose to fill none of them when prices increase, although our setting does not offer a clean way to distinguish variation in number of drugs from other influences on decision-making.

We also tested for the presence of other related behaviors that might represent heuristics, by calculating how the demand response varies with some potentially salient characteristics of drugs. Table 8 shows, for example, that a patient's most expensive drug is 45.9% more likely to be dropped than their least expensive drug. The first medication they ever filled is 37.2% more likely to be dropped than their last (i.e., first-in, first-out).[30]

It is also worth noting that, while some of the mental processes underlying patients' decision making appear to be errors, we also find evidence of the same kinds of sophisticated consumer behavior noted in previous studies (Einav and Finkelstein, 2018). For example, patients are 80.3% more likely to drop branded medications than generic ones when prices increase. They also seem to optimize dimensions of utility that we observe only imperfectly. Using a novel index of the side effect likelihood of certain medications (O'Connell et al., 2018), we find that drugs that commonly

[29] We estimate this by replacing December mortality with December out-of-pocket spending in equation (5).

[30] All rank-orderings of drugs (most vs. least expensive, first vs. last filled) are calculated based on data from the first 90 days of enrollment. We generate a "bundle" of fills from this period, defining fills at the NDC-9 level, and then measure the outcome as an indicator for whether or not that NDC-9 was filled in December (as in equation (5)).

cause side effects are 90.4% more likely to be dropped in response to price increases than others. Finally, and most relevant to our own empirical setup and motivated by (Einav et al., 2015), we find evidence of forward-looking behavior and cross-year substitution. Beneficiaries who face the highest prices in December consume more medicines in January, and fill their medications earlier in the month, indicating they are quite aware that prices will reset. All of this indicates that there is no contradiction between error and behavioral hazard in some decisions (e.g., involving long-term health outcomes), and sophistication and moral hazard in others (e.g., involving prices and plan structures). These forces coexist within populations, and perhaps even within patients.

6 Conclusions

We find that small increases in cost cause patients to cut back on drugs with large benefits, ultimately causing their death. Cutbacks are widespread, but most striking are those seen in patients with the greatest treatable health risks, in whom they are likely to be particularly destructive. It is difficult to affirmatively establish that we have identified behavioral hazard, in the precise sense of a systematic failure to balance the cost with the benefit of care. But we emphasize that the size of the mortality increase cannot be reconciled with any current understanding of the value patients place on life.

We emphasize that our results do not capture the total impact of cost-sharing on health. We estimate only mortality, not morbidity, and only how December price changes affect 65-year-olds' December mortality: a very specific setting, and a very short time period. But patients face cost-sharing throughout the year, and the life-span. If they respond with cutbacks similar to the ones we observe here, they would experience similar increases in mortality in many other settings and over longer time periods. While these effects are as-yet undetected, there is no reason to think that they are not present and equally large. Indeed, because our estimates are formed on largely healthy 65-year olds, effects in the larger (older) Medicare population may be quite different, and potentially larger, if the benefit of drugs is increasing in the underlying mortality hazard (e.g., older patients, nursing home patients, dementia patients), and if drug benefits cumulate over time horizons longer than one month. Understanding the range of health consequences of cost-sharing, and developing new policies to limit harms, is an urgent need. Large-scale randomized experiments,

in the tradition of RAND or Oregon, would be useful to assess the size of behavioral hazard in different populations. Even with such rigorous designs, machine learning could play an important role in reducing the sample sizes required to detect mortality effects, by oversampling patients at high risk of behavioral hazard.

Despite the magnitude of the mortality effect we uncover, unambiguous evidence linking price increases to health effects has been lacking, in large part due to the rarity of mortality. Machine learning predictions played a key role in our empirical strategy. First, we used them to control for patients' counterfactual spending, in order to more precisely estimate the effect of enrollment month on price. A different set of predictions were also key to creating clinically meaningful indices of patient vulnerability to drug-treatable risks. In both cases, we used 'out-of-the-box' algorithms to generate predictions, then incorporated them into a traditional instrumental variables setup. This illustrates the utility of machine learning predictions, which despite not being causal can be valuable inputs into estimating a causal parameter (Mullainathan and Spiess, 2017).

Implicated in these results is a behavioral pattern we cannot explain: a number of patients choosing to fill a very small number of drugs—zero or one—when prices increase, irrespective of how many drugs they were on to begin with. While not conclusive, this behavior is more common in the same patients who suffer the largest mortality effects from price shocks on the margin. It could be consistent with a large number of behavioral mechanisms that distort or scrap the rational calculation of marginal costs and benefits. Adjudicating between the many possible explanations for this anomaly—salience of costs (Bordalo et al., 2013), present bias or memory failures (Laibson, 1997; O'Donoghue and Rabin, 1999), inattention (Handel and Schwartzstein, 2018; Gabaix, 2019), or failures of 'meta-reasoning' (Ackerman and Thompson, 2017)—is a fruitful area for behavioral scientists interested in research with direct implications for life and death.

Our findings should be interpreted alongside a sizable literature on moral hazard in health care, which has documented that patients' can also respond to price changes for prescription drugs in basically rational ways (Einav and Finkelstein, 2018). In the same setting as this literature (price changes from the Medicare Part D non-linear contract design), we replicate many of its key findings, and generally find evidence of considerable sophistication existing alongside deeply flawed decision making. However, given the presence and large magnitude of the coexisting errors, the price elasticity of demand for medical care is not a sufficient statistic for welfare: seemingly rational

price responses are outweighed by 'sub-optimal' health response (Baicker et al., 2015; Einav and Finkelstein, 2018). Our results thus contribute to a growing literature in economics documenting the mortality consequences of plan choice (Abaluck et al., 2020), receipt of health insurance (Goldin et al., 2020; Huh and Reif, 2017; Miller et al., 2019), income (Gross et al., 2020), all of which point to taking a more complete view of welfare, beyond demand elasticities. The magnitudes of both the utilization *and* health responses are needed to understand welfare and design insurance contracts.

Our results argue that health insurance should address both behavioral hazard and moral hazard, proportional to their importance for welfare. The large mortality consequences of behavioral hazard are not currently factored into the design of cost-sharing—but should be. The extremely low life-year valuations we document can be seen as an opportunity for policy-makers to purchase large gains in health at extremely low cost, by investing in intelligent redesign of cost-sharing policies. One way to do so would be via value-based insurance design (VBID), where proven treatments (e.g. anti-hypertensives) are given zero (or even negative) copayments, while treatment with ambiguous benefit (e.g. proton pump inhibitors) are given high copayments (Chernew et al., 2007). Such models, whether focused on drugs or more broadly, have shown promise in a variety of settings, for increasing adherence and reducing disparities. Machine learning could extend this idea by allowing for individualized formularies, where each patient faces a particular formulary based on their specific health risks (Chernew et al., 2008; Lewey et al., 2013; Choudhry et al., 2014; Gaffney et al., 2020). While our results could be interpreted as encouraging for these efforts, they also suggest two risks. First, we do not fully understand patients' choice calculus, and so should be wary of extrapolating from existing, simple models. For example, if (positive) copayments distort the cost-benefit calculus and cause patients to drop *all* their medicines—even the ones with no copayments—then value-based schemes would not improve welfare. The task would therefore require understanding and optimizing the patient-level demand response as a whole, as opposed to optimizing decisions about the marginal drug. Second, market forces—not policy makers—will determine whether patients would even take-up such plans if offered. If insurance companies offered a value-based plan, but patients were unaware of their behavioral biases, they would not want to pay higher premiums for more generous coverage. This suggests a role for regulation, which could enforce some aspects of value-based pricing. It may also be possible to take aim at the behavioral biases underlying errors directly, through some set of nudges or coaching. These are important areas for research.

One conclusion remains clear: patient cost-sharing introduces large and deadly distortions into the cost-benefit calculus. Payers should evaluate the merits of these policies in light of their impact on health, not just on health care costs.

References

Abaluck, J., Bravo, M. M. C., Hull, P., and Starc, A. (2020). Mortality Effects and Choice Across Private Health Insurance Plans. Technical Report w27578, National Bureau of Economic Research.

Ackerman, R. and Thompson, V. A. (2017). Meta-Reasoning: Monitoring and Control of Thinking and Reasoning. *Trends in Cognitive Sciences*, 21(8):607–617.

ALLHAT Group (2002). Major Outcomes in High-Risk Hypertensive Patients Randomized to Angiotensin-Converting Enzyme Inhibitor or Calcium Channel Blocker vs Diuretic: The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA: The Journal of the American Medical Association*, 288(23):2981–2997.

Arias, E. (2014). United States Life Tables, 2010. *National Vital Statistics Reports*, 63(7).

Aron-Dine, A., Einav, L., Finkelstein, A., and Cullen, M. (2015). Moral Hazard in Health Insurance: Do Dynamic Incentives Matter? *The Review of Economics and Statistics*, 97(4):725–741.

Augenblick, N. and Nicholson, S. (2016). Ballot Position, Choice Fatigue, and Voter Behaviour. *The Review of Economic Studies*, 83(2):460–480.

Baicker, K., Mullainathan, S., and Schwartzstein, J. (2015). Behavioral Hazard in Health Insurance. *The Quarterly Journal of Economics*, 130(4):1623–1667.

Bibbins-Domingo, K., Grossman, D. C., Curry, S. J., Davidson, K. W., Epling, J. W., García, F. A. R., Gillman, M. W., Kemper, A. R., Krist, A. H., Kurth, A. E., Landefeld, C. S., LeFevre, M. L., Mangione, C. M., Phillips, W. R., Owens, D. K., Phipps, M. G., and Pignone, M. P. (2016). Statin Use for the Primary Prevention of Cardiovascular Disease in Adults: US Preventive Services Task Force Recommendation Statement. *JAMA*, 316(19):1997–2007.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2013). Salience and Consumer Choice. *Journal of Political Economy*, 121(5):803–843.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2020). Memory, Attention, and Choice. *The Quarterly Journal of Economics*, 135(3):1399–1442.

Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., and Kolstad, J. T. (2017). What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics. *The Quarterly Journal of Economics*, 132(3):1261–1318.

Buckles, K. S. and Hungerman, D. M. (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *The review of economics and statistics*, 95(3):711–724.

Calverley, P. M., Anderson, J. A., Celli, B., Ferguson, G. T., Jenkins, C., Jones, P. W., Yates, J. C., and Vestbo, J. (2007). Salmeterol and Fluticasone Propionate and Survival in Chronic Obstructive Pulmonary Disease. *New England Journal of Medicine*, 356(8):775–789.

Chandra, A., Gruber, J., and McKnight, R. (2010). Patient Cost-Sharing and Hospitalization Offsets in the Elderly. *American Economic Review*, 100(1):193–213.

Chandra, A., Handel, B., and Schwartzstein, J. (2019). Behavioral economics and health-care markets. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:459–502.

Chandra, A. and Skinner, J. (2012). Technology Growth and Expenditure Growth in Health Care. *Journal of Economic Literature*, 50(3):645–680.

Chandra, A. and Staiger, D. O. (2020). Identifying Sources of Inefficiency in Healthcare. *The Quarterly Journal of Economics*, 135(2):785–843.

Chernew, M. E., Rosen, A. B., and Fendrick, A. M. (2007). Value-based insurance design. *Health Affairs (Project Hope)*, 26(2):w195–203.

Chernew, M. E., Shah, M. R., Wegh, A., Rosenberg, S. N., Juster, I. A., Rosen, A. B., Sokol, M. C., Yu-Isenberg, K., and Fendrick, A. M. (2008). Impact Of Decreasing Copayments On Medication Adherence Within A Disease Management Environment. *Health Affairs*, 27(1):103–112.

Choudhry, N. K., Avorn, J., Glynn, R. J., Antman, E. M., Schneeweiss, S., Toscano, M., Reisman, L., Fernandes, J., Spettell, C., Lee, J. L., Levin, R., Brennan, T., and Shrank, W. H. (2011). Full Coverage for Preventive Medications after Myocardial Infarction. *New England Journal of Medicine*, 365(22):2088–2097.

Choudhry, N. K., Bykov, K., Shrank, W. H., Toscano, M., Rawlins, W. S., Reisman, L., Brennan, T. A., and Franklin, J. M. (2014). Eliminating Medication Copayments Reduces Disparities In Cardiovascular Care. *Health Affairs*, 33(5):863–870.

Currie, J. and Schwandt, H. (2013). Within-mother analysis of seasonal patterns in health at birth. *Proceedings of the National Academy of Sciences*, 110(30):12265–12270.

Cutler, D., Skinner, J. S., Stern, A. D., and Wennberg, D. (2019). Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending. *American Economic Journal: Economic Policy*, 11(1):192–221.

Cutler, D. and Zeckhauser, R. (2000). *Handbook of Health Economics, Volume IA*. Elsevier, Amsterdam.

Dahlöf, B., Devereux, R. B., Kjeldsen, S. E., Julius, S., Beevers, G., de Faire, U., Fyhrquist, F., Ibsen, H., Kristiansson, K., Lederballe-Pedersen, O., Lindholm, L. H., Nieminen, M. S., Omvik, P., Oparil, S., Wedel, H., and LIFE Study Group (2002). Cardiovascular morbidity and mortality in the Losartan Intervention For Endpoint reduction in hypertension study (LIFE): a randomised trial against atenolol. *Lancet (London, England)*, 359(9311):995–1003.

Einav, L. and Finkelstein, A. (2018). Moral Hazard in Health Insurance: What We Know and How We Know It. *Journal of the European Economic Association*, 16(4):957–982.

Einav, L., Finkelstein, A., and Polyakova, M. (2018). Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D. *American Economic Journal: Economic Policy*, 10(3):122–153.

Einav, L., Finkelstein, A., and Schrimpf, P. (2015). The Response of Drug Expenditure to Nonlinear Contract Design: Evidence from Medicare Part D. *The Quarterly Journal of Economics*, 130(2):841–899.

Evans, C. and Crawford, B. (1999). Patient Self-Reports in Pharmacoeconomic Studies. *PharmacoEconomics*, 15(3):241–256.

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., and Group, O. H. S. (2012). The Oregon Health Insurance Experiment: Evidence from the First Year*. *The Quarterly Journal of Economics*, 127(3):1057–1106.

Gabaix, X. (2019). Behavioral inattention. *Handbook of Behavioral Economics: Applications and Foundations 1*, 2:261–343.

Gaffney, A., Bor, D. H., Himmelstein, D. U., Woolhandler, S., and McCormick, D. (2020). The Effect Of Veterans Health Administration Coverage On Cost-Related Medication Nonadherence: The effect of the Veterans Health Administration's low-cost drug coverage on cost-related medication nonadherence among people participating in the system. *Health Affairs*, 39(1):33–40.

Geruso, M., Layton, T. J., and Wallace, J. (2020). Are All Managed Care Plans Created Equal? Evidence from Random Plan Assignment in Medicaid. Working Paper 27762, National Bureau of Economic Research.

Goldin, J., Lurie, I. Z., and McCubbin, J. (2020). Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach. *The Quarterly Journal of Economics*, 136(1):1–49.

Gross, T., Layton, T., and Prinz, D. (2020). The Liquidity Sensitivity of Healthcare Consumption: Evidence from Social Security Payments. Technical Report w27977, National Bureau of Economic Research.

Group, S. (1991). Effect of Enalapril on Survival in Patients with Reduced Left Ventricular Ejection Fractions and Congestive Heart Failure. *New England Journal of Medicine*, 325(5):293–302.

Hajjar, E. R., Cafiero, A. C., and Hanlon, J. T. (2007). Polypharmacy in elderly patients. *The American Journal of Geriatric Pharmacotherapy*, 5(4):345–351.

Handel, B. and Schwartzstein, J. (2018). Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care? *Journal of Economic Perspectives*, 32(1):155–178.

Huh, J. and Reif, J. (2017). Did Medicare Part D reduce mortality? *Journal of Health Economics*, 53:17–37.

Iyengar, S. S. and Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, 94(7):530–539.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291.

Kaplan, C. M. and Zhang, Y. (2017). Anticipatory Behavior in Response to Medicare Part D's Coverage Gap. *Health Economics*, 26(3):338–351.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction Policy Problems. *American Economic Review*, 105(5):491–495.

Laibson, D. (1997). Golden Eggs and Hyperbolic Discounting. *The Quarterly Journal of Economics*, 112(2):443–478.

Lewey, J., Shrank, W. H., Bowry, A. D. K., Kilabuk, E., Brennan, T. A., and Choudhry, N. K. (2013). Gender and racial disparities in adherence to statin therapy: A meta-analysis. *American Heart Journal*, 165(5):665–678.e1.

Miller, S., Johnson, N., and Wherry, L. R. (2019). Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data. Technical Report w26081, National Bureau of Economic Research.

Mullainathan, S. and Obermeyer, Z. (2019). A Machine Learning Approach to Low-Value Health Care: Wasted Tests, Missed Heart Attacks and Mis-Predictions. Technical Report w26168, National Bureau of Economic Research.

Mullainathan, S. and Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2):87–106.

Natarajan, P., Young, R., Stitziel, N. O., Padmanabhan, S., Baber, U., Mehran, R., Sartori, S., Fuster, V., Reilly, D. F., Butterworth, A., Rader, D. J., Ford, I., Sattar, N., and Kathiresan, S. (2017). Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation*, 135(22):2091–2101.

Newhouse, J. P. and Group, I. E. (1993). Free for All?: Lessons from the RAND Health Insurance Experiment.

O'Connell, J., Burke, , Mulryan, N., O'Dwyer, C., Donegan, C., McCallion, P., McCarron, M., Henman, M. C., and O'Dwyer, M. (2018). Drug burden index to define the burden of medicines in older adults with intellectual disabilities: An observational cross-sectional study. *British Journal of Clinical Pharmacology*, 84(3):553–567.

O'Donoghue, T. and Rabin, M. (1999). Doing It Now or Later. *American Economic Review*, 89(1):103–124.

Oliver, T. R., Lee, P. R., and Lipton, H. L. (2004). A Political History of Medicare and Prescription Drug Coverage. *The Milbank Quarterly*, 82(2):283–354.

Packer, M., Bristow, M. R., Cohn, J. N., Colucci, W. S., Fowler, M. B., Gilbert, E. M., and Shusterman, N. H. (1996). The Effect of Carvedilol on Morbidity and Mortality in Patients with Chronic Heart Failure. *New England Journal of Medicine*, 334(21):1349–1355.

Peeters, A., Mamun, A. A., Willekens, F., and Bonneux, L. (2002). A cardiovascular life history. A life course analysis of the original Framingham Heart Study cohort. *European Heart Journal*, 23(6):458–466.

Raffoul, M., Moore, M., Kamerow, D., and Bazemore, A. (2016). A Primary Care Panel Size of 2500 Is neither Accurate nor Reasonable. *The Journal of the American Board of Family Medicine*, 29(4):496–499.

Ridker, P. M., Danielson, E., Fonseca, F. A., Genest, J., Gotto, A. M., Kastelein, J. J., Koenig, W., Libby, P., Lorenzatti, A. J., MacFadyen, J. G., Nordestgaard, B. G., Shepherd, J., Willerson, J. T., and Glynn, R. J. (2008). Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated C-Reactive Protein. *New England Journal of Medicine*, 359(21):2195–2207.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131.

Young, B. A., Lin, E., Von Korff, M., Simon, G., Ciechanowski, P., Ludman, E. J., Everson-Stewart, S., Kinder, L., Oliver, M., Boyko, E. J., and Katon, W. J. (2008). Diabetes Complications Severity Index and Risk of Mortality, Hospitalization, and Healthcare Utilization. *The American journal of managed care*, 14(1):15–23.

Zhang, Y., Devore, E. E., Strohmaier, S., Grodstein, F., and Schernhammer, E. S. (2019). Birth month, birth season, and overall and cardiovascular disease mortality in US women: prospective cohort study. *BMJ*, 367:l6058.

Zinman, B., Wanner, C., Lachin, J. M., Fitchett, D., Bluhmki, E., Hantel, S., Mattheus, M., Devins, T., Johansen, O. E., Woerle, H. J., Broedl, U. C., Inzucchi, S. E., and EMPA-REG OUTCOME Investigators (2015). Empagliflozin, Cardiovascular Outcomes, and Mortality in Type 2 Diabetes. *The New England Journal of Medicine*, 373(22):2117–2128.

Tables

Table 1: Sample Summary Statistics

	Main Sample	Dual/LIS Sample
Panel A: Observations	358,706	127,089
Panel B: Demographics		
% Female	0.592	0.614
% White	0.896	0.605
% Black	0.048	0.191
% Hispanic	0.006	0.081
Zip code median income	59,274	47,655
Panel B: Part D Utilization and Health		
% in stand alone PDP	0.517	0.827
One-year total spending	1,478	2,438
One-year mortality	0.009	0.026
Panel C: Top Diagnoses†		
Disorders of lipid metabolism	0.366	0.288
Hypertension	0.357	0.411
Heart disease	0.177	0.211
Eye disorders	0.166	0.142
Non-traumatic joint disorders	0.16	0.187
Diabetes (w/o complication)	0.159	0.221

† CCS level 2 from Part B claims, for those in standalone PDP (non-MA) plans.

Table 2: Enrollment Month Effect on Coinsurance and Mortality: Reduced Form and 2SLS

	(1)	(2)	(3)
	Base Rate	*Enrollment Month Effect*	
	Mortality (Monthly, p.p.)	Coinsurance (p.p.)	Mortality (Monthly, p.p.)
Panel A: Reduced Form			
Predicted to be in initial coverage	0.047	-0.0622*** (0.00393)	-0.000292 (0.00191)
Predicted to be in coverage gap	0.132	-2.27*** (0.025)	-0.0113** (0.00521)
Predicted to be in catastrophic	0.279	1.52*** (0.131)	0.0465** (0.0218)
Panel B: Two-Stage Least Squares		*Coinsurance Effect (Instrumented)* Mortality (Monthly, p.p.)	
Entire sample, first stage by pctile bins of predicted spending		0.00435** (0.00209)	
Panel C: Interpretation	Mortality (Monthly, p.p.)	First Stage (Median, p.p)	Median First Stage × 2SLS Est.
Predicted compliers (Coverage gap and catastrophic)	0.147	11.0	0.048
		Median First Stage × 2SLS ÷ Monthly Mortality (×100)	
Percent change in monthly mortality		32.7%	

$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$

Notes: Panel A: reduced form results of the effect of enrollment month within three bins of predicted spending, corresponding to initial coverage (1-70 percentiles), coverage gap (71-97), and catastrophic (98-100). Column (1) reports mean monthly mortality in December by bin (p.p.); column (2) reports first stage estimates of enrollment month on December coinsurance rate (p.p.); column (3) reports the reduced form estimates of enrollment month on December mortality (p.p.). Panel B: 2SLS estimate of mortality change (in p.p.) per p.p. coinsurance change. Here, the first stage incorporates fine percentile bins of predicted spending (Figure 3) interacted with enrollment month. First stage F-statistic is 210.7. Panel C: Scales the 2SLS estimate (Panel B) by observed sample means of 'compliers': those predicted to spend enough to approach either the coverage gap or catastrophic coverage (top 30% of predicted spending). Column (1) reports the mean monthly mortality in December (p.p.); column (2) reports the median absolute difference in coinsurance rate (p.p.) between February and September enrollees (using percentile bins of predicted spending); column (3) multiplies the 2SLS estimate from Panel B by the median first stage. The final row divides the estimate in column (3) by the monthly mortality in column (1) to give the percent change in mortality between February and September. Standard errors are heteroskedasticity-robust.

Table 3: Enrollment Month Effects on Mortality: Falsification Tests

	(1)	(2)
	Mortality (Monthly, p.p.)	
	Mean	*Enrollment Month Effect (Reduced Form)*
Panel A: Dual-Eligibles, Age 65, December		
Predicted to be in initial coverage	0.169	0.00666
		(0.00647)
Predicted to be in coverage gap	0.399	-0.0144
		(0.015)
Predicted to be in catastrophic	0.786	-0.0188
		(0.0614)
Panel B: Main Sample, Age 65, Third Month of Enrollment		
Predicted to be in initial coverage	0.045	0.00152
		(0.00189)
Predicted to be in coverage gap	0.113	0.00642
		(0.00495)
Predicted to be in catastrophic	0.349	-0.0164
		(0.0249)

$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$

Notes: Falsification checks to determine whether enrollment month appears to be correlated with mortality via channels other than the price changes captured in our main analysis. We test this in two settings. Panel A: 65-year-old dual-eligibles in December of their first year of enrollment, who face uniform cost-sharing throughout the year (no coverage gap). Panel B: 65-year-olds in our main sample, before they begin to approach spending thresholds at which cost-sharing begins. In Panel B, we predict spending using the first two months of enrollment, rather than the first three in our main analysis, and test for mortality effects in the third month. Column (1) reports the mean mortality rate (p.p.) in the month/sample studied, and column (2) reports the reduced form estimate of monthly mortality on enrollment month. We do not estimate a 2SLS model because there is no first stage (i.e., no price variation due to enrollment month).

Table 4: Enrollment Month Effects on Mortality: Robustness to Design Choices in Main Analysis

	(1)	(2)	(3)
	Enrollment Month Effect (Reduced Form, p.p.)		*Coinsurance Effect (2SLS, p.p.)*
	Coverage Gap	Catastrophic	
Main Specification	-0.0113**	0.0465**	0.00435**
	(0.00521)	(0.0218)	(0.00209)
No Plan Fixed Effects	-0.0119**	0.0486**	0.00428**
	(0.00513)	(0.0209)	(0.00203)
Data Used to Predict Spending			
30 days initial claims	-0.00986**	0.031	0.00574**
	(0.00481)	(0.0218)	(0.0026)
60 days initial claims	-0.0127**	0.0449**	0.00535**
	(0.00512)	(0.0212)	(0.00224)
Instrument Specification			
Use birth month as instrument	-0.00999*	0.0539***	0.00443**
	(0.00519)	(0.0193)	(0.00218)
Include only 'on time' enrollees	-0.0147**	0.0564**	0.0054**
	(0.00628)	(0.0223)	(0.00246)
Outcome Period			
Use Dec-Jan as outcome period	-0.0162**	0.0648**	0.00506*
	(0.00722)	(0.0303)	(0.00269)
Use Dec-Feb as outcome period	-0.0166**	0.0727**	0.00561*
	(0.00841)	(0.0346)	(0.00313)

$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$

Notes: Each row presents estimates from a separate specification aimed to test the robustness of our main estimates to several design choices: with and without Part D plan fixed effects; with varying number of days used to predict spending; different ways to specify the instrument with respect to birth or enrollment month ('on time' refers to patients that enroll in the month that they turn 65); and longer time periods over which mortality effects are estimated. Columns (1) and (2) report reduced form estimates of December mortality on enrollment month for those predicted to be in the coverage gap and catastrophic coverage, respectively. Column (3) reports 2SLS estimates from the first stage specification using fine-bins of predicted spending.

Table 5: Estimates of Drug Utilization on Mortality: OLS vs. 2SLS

	(1)	(2)	(3)
	Independent Variable (OLS)/ Endogenous Variable (2SLS)		
	Total Spending	Number of Fills	≥ 5 Fills
OLS, with increasing controls			
None	0.000148***	0.0262***	0.166***
	(0.0000409)	(0.00354)	(0.0223)
Demographic	0.000148***	0.0265***	0.167***
	(0.0000411)	(0.00355)	(0.0223)
Predicted Drugs	0.000109**	0.0188***	0.115***
	(0.0000427)	(0.00441)	(0.0239)
Demographic + Predicted Drugs	0.000107**	0.0185***	0.113***
	(0.0000428)	(0.0044)	(0.0238)
Two-Stage Least Squares			
	-0.000788**	-0.0993*	-0.422
	(0.000328)	(0.0543)	(0.365)

$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$

Notes: Comparison of OLS vs. 2SLS estimates of utilization on mortality. OLS regressions consider mortality in January (first month of year 2) as a function of December (last month of year 1) utilization measures, because those who die in December will have mechanically fewer fills (because they are alive for less days in the month). Two-stage estimates consider mortality in December as a function of December utilization measures (instrumented by enrollment month). Controls included in OLS regressions include basic demographic variables (race and sex), and number of drugs a beneficiary is predicted to be on in December (using a similar prediction strategy to the one outlined in the main text, but predicting number of drugs rather than spending; the intuition is similar to controlling for a propensity score). Column (1) uses total December spending on the right hand side for OLS, and as the endogenous variable in 2SLS; column (2) uses number of December drug fills; column (3) uses an indicator for whether a patient made 5 or more fills in December.

Table 6: Demand Response to Cost-Sharing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		2SLS Estimates			*Mortality Contribution*		
	Dec. Mean	Coins. Effect	2SLS× Med. 1S	% Fills Decr.	RCT Effect	% on Class	Total Contrib.
Panel A: Overall							
Any Fill (p.p.)	84.1	-0.2*** (0.0152)	-2.21	-	-	-	-
Number of Fills	3.29	-0.0308*** (0.00132)	-0.34	-	-	-	-
Total Spending ($)	270.8	-5.54*** (0.269)	-61.2	-	-	-	-
Panel B: By Class							
Statins	0.378	-0.004*** (0.00026)	-0.044	-11.7%	-26.7%	0.584	1.82%
β-blockers	0.198	-0.0011*** (0.00018)	-0.013	-6.6%	-82%	0.319	1.72%
ACE Inhibitors	0.178	-0.00093*** (0.00017)	-0.01	-5.6%	-37.3%	0.298	0.63%
Thiazides	0.155	-0.00099*** (0.00016)	-0.011	-7.1%	-37.3%	0.264	0.7%
Diabetes (oral)	0.199	-0.0023*** (0.00026)	-0.025	-12.5%	-42.7%	0.23	1.23%
ARBs	0.12	-0.0013*** (0.00014)	-0.015	-12.5%	-37.3%	0.202	0.94%
Inhalants	0.085	-0.0015*** (0.00014)	-0.016	-18.9%	-26.5%	0.175	0.88%
CCBs	0.099	-0.00081*** (0.00013)	-0.009	-9.1%	-37.3%	0.158	0.53%
Total			**-0.143**				**8.44%**

$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$

Notes: Panel A: Column (1) presents the December mean of each variable (for 'compliers': top 30% of sample in terms of predicted spending); column (2) presents 2SLS estimates of December utilization on the coinsurance rate; column (3) multiplies 2SLS estimates by median absolute difference in coinsurance between February and September enrollees. Panel B: Column (4) calculates percentage reduction in fills for each class (= (4) ÷ (1)); column (5) is the percentage mortality reduction from clinical trials for each class (treatment-on-treated; see Appendix Table D.1 for calculation and sources); column (6) is our estimate of the total fraction of our sample on a given drug class (formed using the first 90 days of data, and not simply the December mean which is net of enrollment month effects); column (7) is the percentage mortality increase attributable to each class in our sample (= (4)×(5)×(6)). The final row of columns (4) and (7) sums over each row in Panel B.

Table 7: Demand Response for Specific Drug Classes, by Risk of Underlying Illness and Income

	(1)	(2)	(3)
	2SLS: Number of Fills Estimate, by Group		
	Overall	By Risk	By Zip5 Income
Cardiovascular by risk/income:			
Coinsurance rate	-0.00921***	-0.00323***	-0.00925***
	(0.000588)	(0.000714)	(0.000747)
Coinsurance rate*Top 1/3	-	-0.00907***	0.000298
		(0.00105)	(0.0012)
Diabetes by risk/income:			
Coinsurance rate	-0.00288***	-0.00236***	-0.00328***
	(0.000306)	(0.0005)	(0.000402)
Coinsurance rate*Top 1/3	-	-0.00068	0.00106*
		(0.00063)	(0.000613)
Respiratory by risk/income:			
Coinsurance rate	-0.00226***	-0.00207***	-0.00239***
	(0.000209)	(0.000285)	(0.000263)
Coinsurance rate*Top 1/3	-	-0.000337	0.000343
		(0.000418)	(0.00043)

$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$

Notes: Each column presents 2SLS estimates where the dependent variable is the number December fills in a given category and the endogenous variable is the December coinsurance rate. Column (1) presents the overall demand response for the category, while columns (2) and (3) present heterogeneity in this effect by risk and zip code income, respectively. Cardiovascular drugs include: statins, β-blockers, angiotensin-converting enzyme (ACE) inhibitors, calcium channel blockers (CCBs), angiotensin ii receptor blockers (ARBs), and thiazide diuretics. Diabetes drugs include insulin and other blood glucose lowering drugs. Respiratory drugs include drugs for obstructive airway diseases.

Table 8: Demand Response by Characteristics of Medications

	(1)	(2)	(3)
	2SLS Estimate		*Observations*
Panel A: Heuristics			
	Least Expensive	Most Expensive	
By total cost	-0.00479***	-0.00699***	N = 236,014
	(0.000191)	(2e-04)	
	First Filled	Last Filled	
By fill order	-0.00594***	-0.00433***	N = 203,243
	(0.000214)	(0.000191)	
Panel B: Sophistication			
	Generic	Branded	
By branded status	-0.0107***	-0.0193***	N = 358,706
	(0.000968)	(0.000637)	
	No Side Effects	Side Effects	
By side effects	-0.0104***	-0.0198***	N = 358,706
	(0.000559)	(0.00103)	
	January Fills	Days to First Fill	
Year2	0.0144***	-0.123***	N = 358,706 (292,365)
	(0.0013)	(0.00643)	

$^{*}p < .1, ^{**}p < .05, ^{***}p < .01$

Notes: Panel A: The "by total cost" and "by fill order" rows have outcome variables that are indicators for whether a specific drug that was filled in the first 90 days of enrollment was also filled in December (measured at the ndc9 level). To be included in the by cost regression, a patient must have made at least two fills of different costs in the first 90 days; to be included in the by fill order regression, a patient must have made at least two fills on different days in the first 90 days. Panel B: Generic, branded, no side effects, and side effects are the number of fills (in each category) made in December of year 1. January fills is the number of fills made in January of year 2. Days to first fill is measured in January through March of year 2; patients that make no fills in this period are excluded from the regression.

Figures

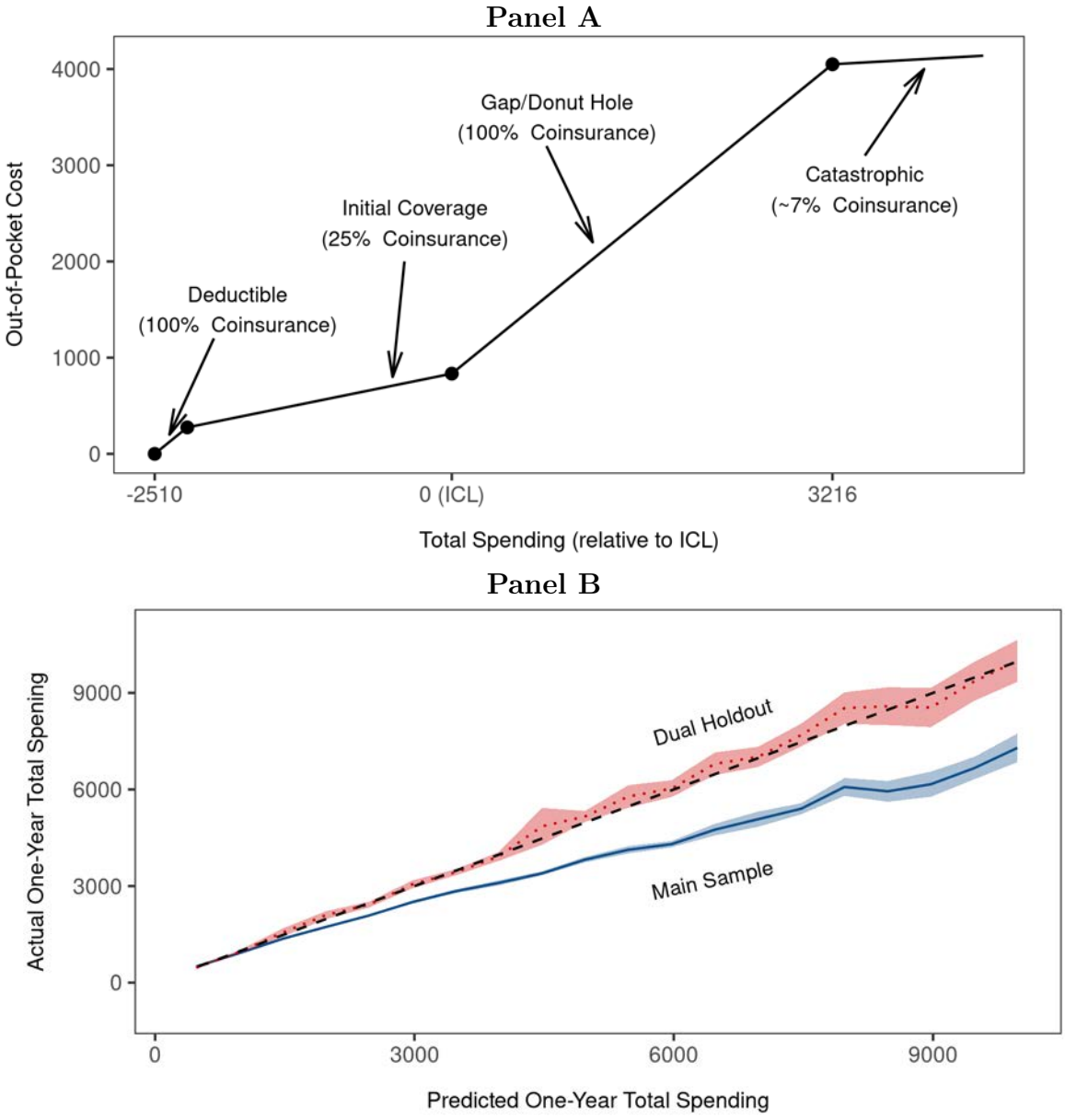


Figure 1: Part D Design and Spending Prediction

Notes: Panel A is adapted from Einav et al. (2015), and illustrates the Part D standard benefit design using 2008 program details. Panel B plots average realized 12-month spending, by $500 bins of predicted spending (from a machine learning model, fit in a separate sample of 65-year-old dual-eligible enrollees with no cost-sharing). For the dual holdout sample (dotted line), 12-month spending is measured from enrollment while for the main sample (solid line), it is measured in calendar year 2 of enrollment (January-December). The dashed line is on the 45 degree angle (actual = predicted). We right-censor the x-axis (at > $10,025, omitting 0.8% of the main sample, 3.0% of the dual sample), for readability.

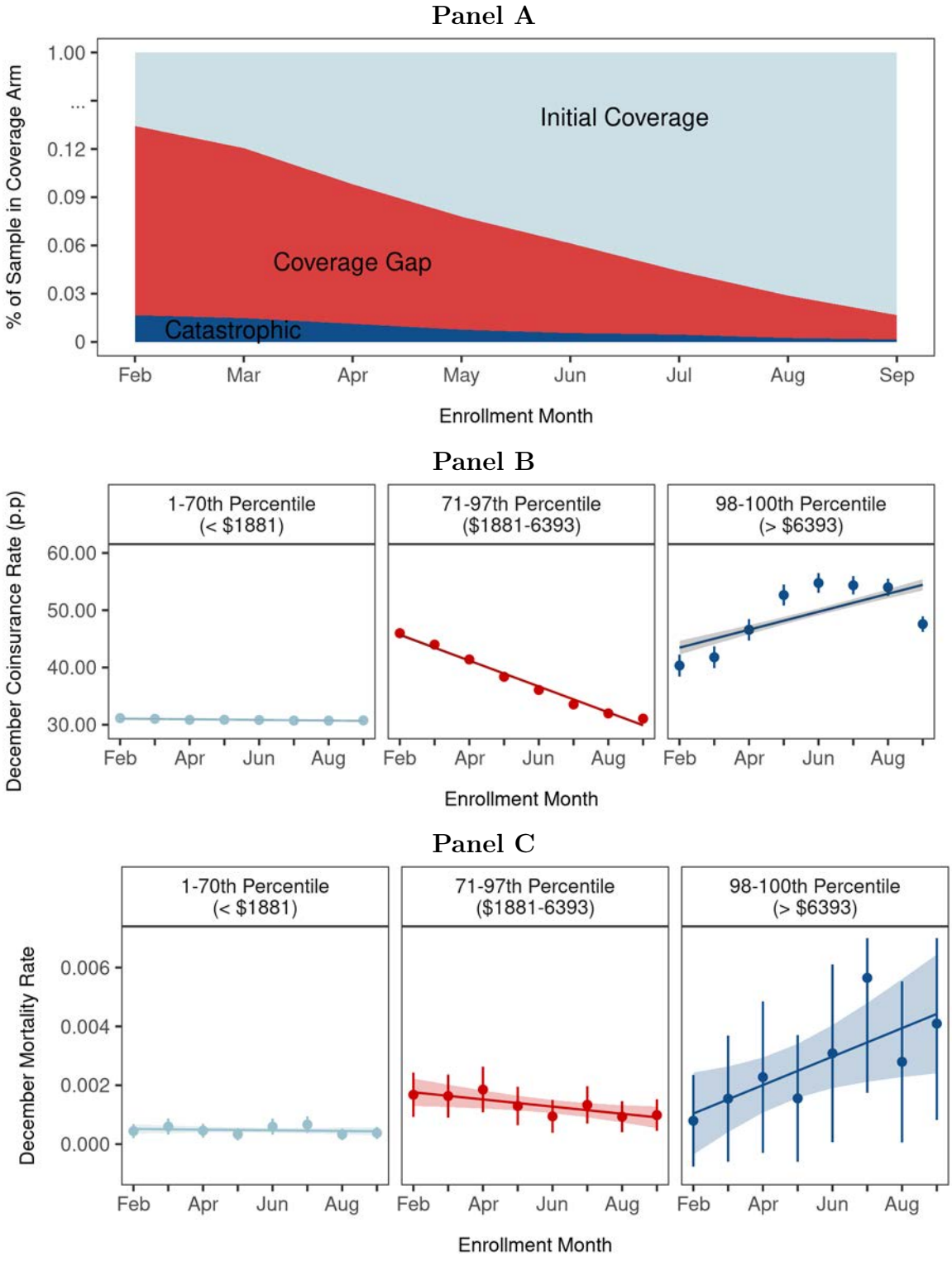


Figure 2: Reduced Form: Effect of Enrollment Month on Coverage Arm, Coinsurance, and Mortality

Notes: Panel A plots the December coverage arm by enrollment month. Panel B plots the December coinsurance rate (in p.p.) by enrollment month and bin of predicted spending. Panel C plots the December mortality rate by enrollment month and bin of predicted spending. Predicted spending bins are based on Figure 1 Panel B. Some confidence intervals (CIs) in the right-most panel of Panel C are truncated for better visibility (all CIs are symmetric).

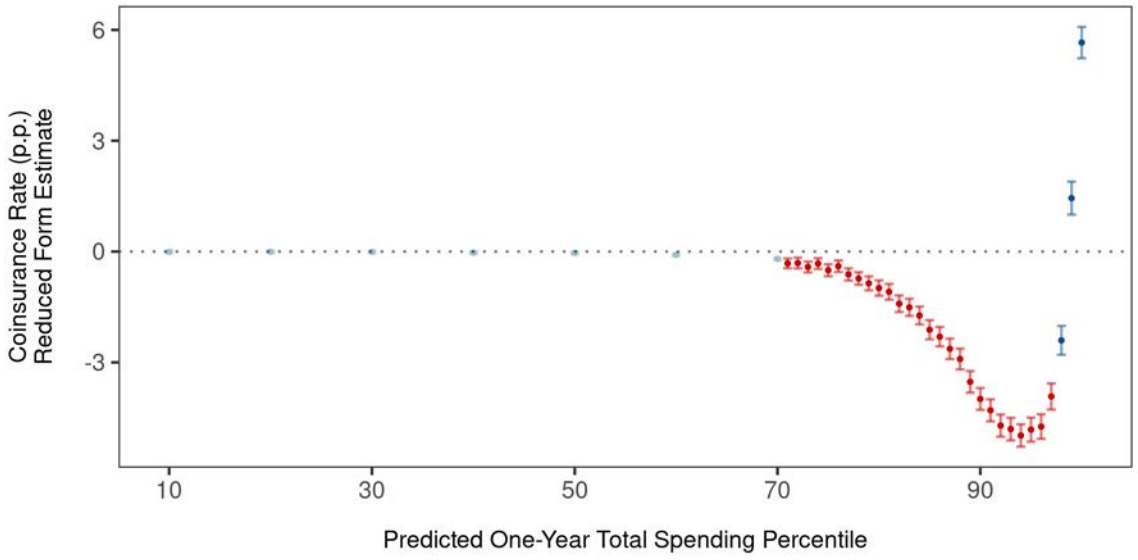


Figure 3: Enrollment Month Effect on Coinsurance Rate, by Percentiles of Predicted Spending

Notes: Each point represents a linear coefficient from a regression of December coinsurance rate (in p.p.) on enrollment month, within a percentile of predicted spending (i.e., our first stage equation (4)). Predicted spending levels below the 71st percentile are grouped into deciles because there is not enough variation at these levels of predicted spending to define unique percentiles. Coefficients on the horizontal line (at zero) indicate no effect of enrollment month on coinsurance rate for those with predicted one-year spending well below the first spending threshold (the transition from initial coverage to the coverage gap). Colors match the three bins of predicted spending in Figure 2.

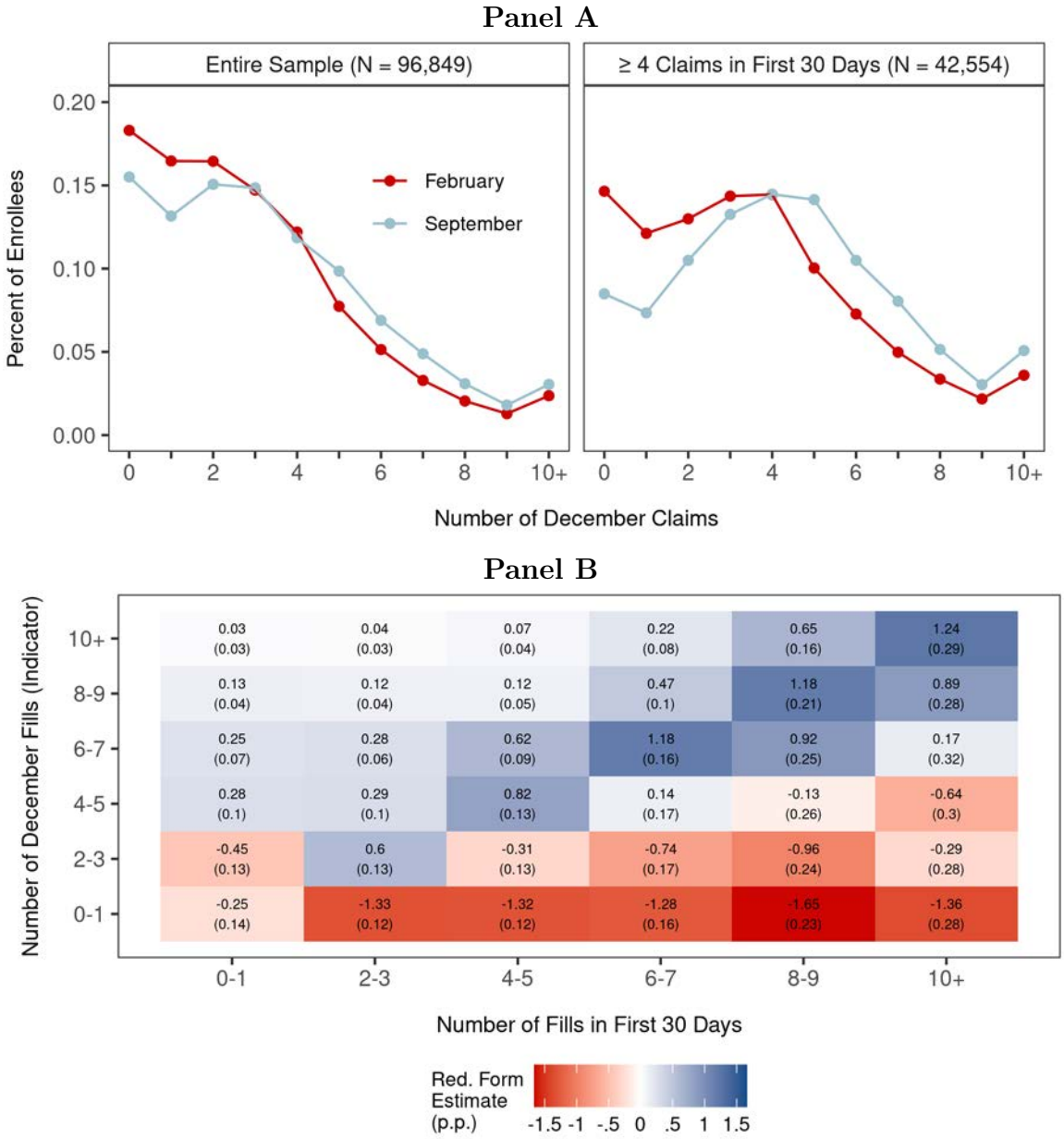


Figure 4: Number of December Drug Fills, by Enrollment Month and Number of Baseline Fills

Notes: This figure restricts to those whose spending is predicted to be in the coverage gap (71-97 percentile of predicted spending). Panel A plots the distribution of December claims for February and September enrollees, both overall (left) and for those with four or more claims at baseline (first 30 days of enrollment; right). Panel B presents estimates (in p.p.) of multiple regressions where indicators for filling a certain number of December claims (left hand side, shown on the y-axis: 0-1, 2-3, ... $\geq$10) are regressed on enrollment month interacted with indicators for the number of claims in the first 30 days (right hand side, shown on the x-axis: : 0-1, 2-3, ... , $\geq$10). Each row thus presents estimates from a separate regression. Negative estimates, shown in red and shaded by the magnitude, mean that those with higher prices (earlier enrollment months) are more likely to fill y claims in December. Positive estimates are shown in blue and mean that those with lower prices (later enrollment months) are more likely to fill y claims in December.