Initial Learning and Eventual Substitution: A Behavior Study in Medicare Part D*

Tina Marsh Dalton† Yi Zhong‡

May 22, 2018

Abstract

Large out-of-pocket cost regions have been gaining traction in insurance plans as a way to influence enrollee behavior and control medical costs. In this paper, we analyze great dimensional richness in learning and behavioral change for Medicare Part D enrollees responding to nonlinear changes in their out-of-pocket costs. In particular, we identify a previously unaddressed learning region : initial purchases. Enrollees fill more often during initial purchases and respond more sharply to a large price shock than they do later in the pricing schedule. We identify enrollee reactions to the price schedule using a discontinuity approach that simultaneously controls for underlying disease mix and enrollee expected end-of-year price. Besides initial purchases, we deconstruct the large spending drop at the "doughnut hole" threshold, and show the decrease is largely driven by dropping branded drugs. Enrollees reduce out-of-pocket costs by simply filling the same mix of drugs less frequently instead of dropping fills in certain disease categories. There is some substitution to generic options, but we show that linking these patterns into disease-level substitution ability reveals substantial heterogeneity.

JEL Codes: I13, D83

Keywords: learning; nonlinear; discontinuity; generic substitution; pharmaceuticals

*The authors thank for helpful comments: Colleen Carey, Gautam Gowrisankaran, Maryam Saeedi, Meghan Skira, Ali Shourideh, Robert Town and participants at the Midwest Health Economics Conference. Funding for this project was provided by the Wake Forest University/ Blue Cross Blue Shield Pilot Grant. Excellent research assistance was provided by Ryan Angi.

†204B Kirby Hall, Department of Economics, Wake Forest University. tina.marsh.dalton@gmail.com

‡Gardner Hall, Department of Economics, University of North Carolina-Chapel Hill

1 Introduction

The Medicare Part D drug benefit has grown tremendously, with participation rising to over 70 percent of eligible enrollees in just 10 years after its introduction in 2006 (Hoadley et al., 2015). The most significant feature of the program is its use of a large cost-sharing region for enrollees, or the "doughnut hole". The standard Part D structure has three distinct coverage regions. The enrollees enjoy a coverage region initially, and, after passing a threshold ($2510 in drug purchases), they enter the infamous "doughnut hole" where all coverage of drug purchases end. The enrollees must continue paying for all purchases completely out-of-pocket until they accumulate $4,050 in total out-of-pocket expenses. After that, generous coverage resumes. Because price changes as the enrollee moves in and out of each region, the pricing schedule for the enrollee is nonlinear with distinct thresholds where out-of-pocket incidence changes. This large cost-sharing feature is in line with other recent approaches in health insurance using increased out-of-pocket payments as a way of encouraging beneficiaries to adjust their spending behaviors (see Brot-Goldberg et al. (2017) for a high-deductible health insurance plan). What these policies can achieve depends on how consumers adjust their drug purchasing behaviors. One of the main goals of a large cost-sharing region is to induce beneficiaries to make substitutions within their drug portfolio into cheaper alternatives to address the issue of moral hazard. If beneficiaries do not make these types of substitutions, then Part D may instead be functioning merely as a short-term subsidy to enrollees with high drug spending, who merely reduce or drop fillings after cost-sharing increases. If substitutions are not made, which prescriptions are dropped? Are beneficiaries stopping medications for particular disease categories or just lowering drug adherence rates across all prescriptions? These details have implications for what policies to balance cost and consumer welfare. In this analysis, we examine the behaviors in detail which would result in the observed spending drop and also document a previously-unexplored region of enrollee response in Part D discussions– enrollees' initial purchases.

This work makes two contributions to the discussion around the responses to the Medicare Part D nonlinear pricing. There is much interest in how enrollees adjust their purchases to the nonlinear structure of the pricing schedule. Previous work focuses on the changes that occur upon entry into the doughnut hole, or examines how well enrollee purchases seem to incorporate the dynamic nature of the nonlinear pricing (Einav et al., 2015; Dalton et al., 2015; Abaluck et al., 2015; Joyce et al., 2013; Jung et al., 2013; Zhang et al., 2009b; and Zhang et al., 2009a). However, we present evidence that learning about the price schedule may also

occur during the enrollee's initial encounters with purchases in the pricing schedule. Initial purchases are another channel for enrollees to learn about the price schedule and incorporate this information into their spending behaviors. In both the initial purchases and the doughnut hole, our second contribution is identification of enrollee reactions to the price schedule using a discontinuity approach that controls for underlying disease mix and expected end-of-year price.

Identifying behavioral responses in spending is difficult when facing a nonlinear price schedule. The first difficulty is that enrollees with different end-of-year spending levels may differ in their decision-making through their beliefs early-in-the-year about the set of prices they will face later. In this work, we isolate a group of enrollees who all face full out-of-pocket prices at the end of the year, and thus should have had reasonably similar expectations about their end-of-year prices. We create an estimation sample of individuals who pass through the doughnut hole between March and August of the coverage year (Dalton et al., 2015 and Abaluck et al., 2015). Thus, any pre-doughnut hole behavior comparisons are over individuals who uniformly ended up facing entry into the doughnut hole region.

The second difficulty presented by nonlinear pricing schedules is that, because price and aggregated spending are linked, enrollees further along the pricing schedule likely have higher unobserved levels of illness. Identifying any behavior changes due to price must also be able to effectively control for underlying illness. Within the sample of enrollees with a reasonable expectation of crossing into the doughnut hole, we use individual fixed effects to analyze within-enrollee behavior as the enrollee responds to price changes along her aggregate spending. That is, we identify responses by comparing enrollee behavior before a price change to her own behavior after the price change, using individual fixed effects along the enrollee's aggregate spending distribution.

An additional advantage of our empirical setting is that the pricing schedules are particularly straightforward in regards to tradeoffs between branded and generic substitution. We examine Part D plans in particular that have nonlinear pricing for branded purchases, but full coverage for generic drugs in the doughnut hole. This means that the choice of generic had the same underlying structure throughout the plan year, with only changes in branded purchases' relative costs.

In accordance with previous work, we document substantial drops in spending at the doughnut hole (Einav et al., 2015; Dalton et al., 2015; Abaluck et al., 2015; Joyce et al., 2013; Jung et al., 2013; Zhang et al., 2009b; and Zhang et al., 2009a). However, we also document large drops in spending within the first few purchases or the first several hundred dollars, even

though prices don't change during these periods. We find evidence of learning, as enrollee fillings visits are most frequent during this period. When enrollees face a large spending shock (without an accompanying price change), we observe that substitution into generic options is more pronounced during early purchases compared to later. At the doughnut hole, we find that an average enrollee drops nearly a tenth of branded filling, but only substitutes about 90 percent of this filling with generic. We do not find a statistically significant change in the number of disease types filled in visits before and after the price change, which suggests that fillings are merely delayed without changes in the overall portfolio of fills. Within specific disease categories, high-blood pressure, heart disease, and diabetes show the greatest responses in substitution into generic fillings. Substitution is most likely in middle-tier expense branded drugs (between 25 and 100 dollars). We find that generic availability contributes to heterogeneity in substitution. Substitution behaviors are low in disease categories at the extremes of either high or low availability of generics. Only categories of 'medium' generic availability show a meaningful behavioral response to increased costs in the doughnut hole.

Our work contributes to the literature on understanding beneficiary behavior in Medicare Part D. Previous work has focused on enrollee's choice of plans (Heiss et al., 2010; Abaluck and Gruber, 2011; Kling et al., 2012; Ketcham et al., 2012; Abaluck and Gruber, 2016; Ho et al., 2017; Polyakova, 2016; and Ketcham et al., 2015), and on the dynamic optimization of enrollees (Einav et al., 2015; Dalton et al., 2015; Abaluck et al., 2015; and Schroeder et al., 2014). The approach here is similar to Dalton et al. (2015), who use a discontinuity-based analysis in a more confined fashion to establish patterns around the doughnut hole in order to inform a dynamic model of patient spending patterns. Dalton et al. (2015) find strong evidence of myopia around the doughnut hole, undergirding the focus in this work on behaviors surrounding the doughnut hole. This paper also contributes to discussions around price salience (see Chetty et al. (2009) and Bordalo et al. (2012) for more theoretical analysis).

The rest of the paper proceeds as follows. Section 2 describes the Part D setting and related literature, Section 3 lays out a simple model of enrollee behavior which motivates the choice of the initial period and doughnut hole region and organizes predictions about enrollee behavior. Section 4 describes our empirical approach. Section 5 describes our data and Section 6 describes the results and implications. Section 7 concludes.

2 Background and Part D Setting

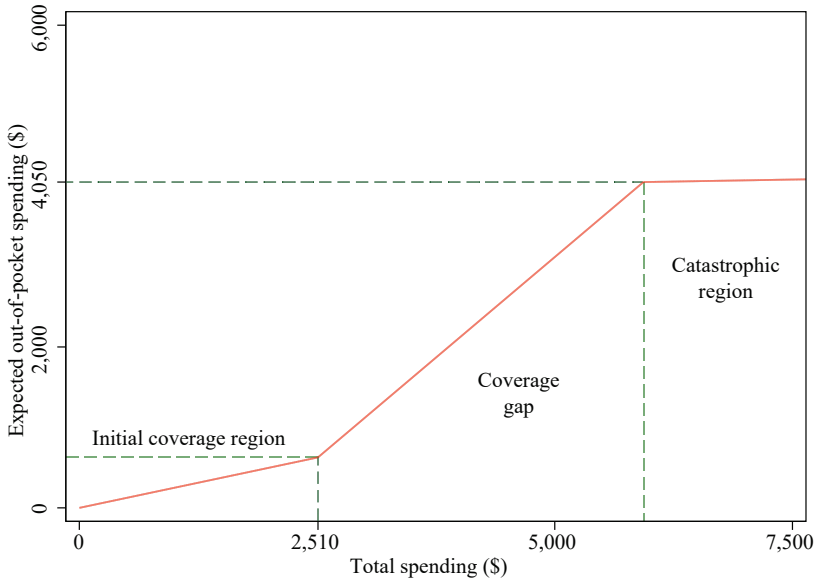
When the Medicare Part D drug benefit was introduced in 2006, it was the largest extension of benefits since Medicare's inception in 1965. The program has grown immensely in popularity since its introduction. The design of Part D plans incorporates cost-sharing which changes as the enrollee increases aggregate drug spending and a large region of full out-of-pocket spending, infamously known as the "doughnut hole." Figure 1 illustrates the nonlinear price structure of Part D in 2008, the year of our data. The standard Part D structure has three distinct coverage regions. The initial spending is subject to a tiered copay system, usually with lower copays for generics or preferred brands and higher copays for non-preferred brands and specialty drugs. This "coverage area," as we will refer to it, continues until the enrollee accumulates $2,510 in drug purchases.[1] At this point, the enrollee moves into the infamous "doughnut hole," where all coverage of drug purchases ends. In the doughnut hole region, the enrollee must pay for all purchases 100 percent out-of-pocket. This is a large price increase, given the copays for generic and most branded drugs in the coverage region usually fall below $40. The enrollee then must continue paying for all purchases completely out-of-pocket until the enrollee has accumulated $4,050 in total out-of-pocket expenses (In our data, this translated on average to the enrollee purchasing at least $1,000 of drug fillings at full out-of-pocket prices). Finally, after over $4,050 in accumulated out-of-pocket expenditure, the enrollee enters the "catastrophic coverage" region, with copays of $2.25 per generic and $5.60 per branded purchase.[2] When enrollees face regions of large out-of-pocket payments, this may encourage them to re-evaluate spending or choose cheaper alternative therapies. Examples of this approach include high-deductible plans, out-of-pocket maximums, and gaps in coverage, where prices change for beneficiaries in a nonlinear manner and often by a large amount (Brot-Goldberg et al., 2017; Dafny et al., 2017; Aron-Dine et al., 2015).

Given the large change to the enrollee's budget and the nonlinear nature of the standard plan, some private providers offer plans that deviate from the standard Medicare plan. Changes to the standard structure are permitted as long as the plan is actuarially equivalent to the standard plan for representative enrollees. In this analysis, we focus on two plans which deviate from the standard plan by removing the doughnut hole feature selectively

[1] Many Part D plans also feature a small deductible before initial coverage begins. In the CMS baseline example plan the deductible is $275. In the plans used in this analysis, one does not have any deductible and the other has a small deductible of $100.

[2] Entering the catastrophic coverage region is rarer under the Part D program, and we focus here only on the doughnut hole threshold.

Figure 1: Relationship between Total Expenditure and Expected Out-of-Pocket Spending in Medicare Part D



within the plan pricing structure. In particular, these plans continue to cover generic drug fills throughout the doughnut hole, essentially changing the price schedule for generic fills into a straightforward linear structure, with constant copays, irregardless of the total spending of the enrollee. The analytical advantage of this plan feature is that the nonlinear pricing structure only changes for branded drugs, whereas generic alternative cost remains constant throughout the nonlinear pricing plan.[3]

Previous work has shown that the introduction of Medicare Part D plan generates significant increases in drug utilization and decreases in patient out-of-pocket spending (Yin et al., 2008; Lichtenberg and Sun, 2007; Ketcham and Simon, 2008; Duggan and Morton, 2010; and Liu et al., 2011). Additionally, Part D plans also improve social welfare as Zhang et al. (2009a) find that the increased drug spending is accompanied with decreases in other medical spending; and Duggan and Morton (2010) find Part D plan leads to a substantial relative price decreases in average branded pharmaceutical prices.

However, accompanying these nonlinear price schedules are concerns about their effectiveness to optimally change beneficiary behavior. In general medical deductibles, Brot-Goldberg

[3]The juxtaposition of the linear and nonlinear schedules is also interesting more generally, since the nonlinear structure is applied selectively in other insurance settings. For example, recent high deductible plans usually include linear pricing on preventative care, with acute care applied to the deductible.

et al. (2017) find that enrollees under-consume before the deductible, even with near-certainty of entering full coverage. There is concern that beneficiaries may reduce expenditures in a way that worsens health outcomes. Joyce et al. (2013) find the coverage gap disrupts drug utilization among seniors with diabetes, even if the effect is modest and concentrated among higher cost and brand-name medications in short-run, the change of drug prescription behavior could harm care plans. Zhang et al. (2009b) find beneficiaries with chronic conditions are more likely to reach the doughnut hole, and those who do reduce adherence to drug prescriptions. Chandra et al. (2010) suggest a potential externality issue with increased prescription cost-sharing, finding increased hospital utilization in response to higher copayments for physicians and prescription drugs, especially among the most ill population.

Most recently, a literature focuses on how well beneficiaries plan their purchases dynamically in these settings (Einav et al., 2015; Dalton et al., 2015; Abaluck et al., 2015; and Schroeder et al., 2014). If the enrollee has the ability to understand the entirety of the price schedule, she may adjust spending early on to avoid the future cost-sharing region. This literature does not find substantial evidence for forward-looking behavior in drug purchases from Medicare enrollees. Dalton et al. (2015) find nearly complete myopia, Abaluck et al. (2015) find a high degree of myopia, while Einav et al. (2015) find evidence showing that enrollees plan their consumption dynamically across the year.

Given the evidence for a large proportion of enrollee response occurring at the doughnut hole region, we need a better understanding of what particular changes contribute to an overall decrease in spending. There are a few studies which break down drug adherence and substitution behaviors for a narrow drug class (Jung et al., 2013 and Fung et al., 2010). In this work we take a different approach, which is to parse out general trends across all drug purchases in an individual portfolio. In this way, we make general comparisons among drug categories of which characteristics lead to changes in purchase behaviors. We compare an individual's own behavior as she moves along the nonlinear pricing schedule.

3 Theoretical Setting and Behavioral Predictions

Changes in spending are a broad measure of consumer behavior, but policy may be more effective if it reflects the deeper dimensionality of drug purchases. Enrollees can react to price changes through quantity adjustments, by dropping prescription fills or filling less often– reacting on the extensive margin. They may also react on the intensive margin by substituting their prescriptions into less expensive drug choices. In this work, we want to decompose the

multidimensional behavior of spending decreases to learn more about how enrollees react to the complex nonlinear price schedule. Besides response at the doughnut hole, we also introduce another important region, which has been largely ignored, where enrollees may react to insurance pricing: initial purchases. In this section, we lay out a baseline framework for examining behavior in these two regions: initial purchases and the region around doughnut hole entry.

Pharmaceutical consumption is an especially valuable area to study price response more generally. First, many drug purchases happen frequently. Over 80 percent of the fillings in our datasets are for drugs to treat chronic conditions. Drugs for chronic conditions need to be filled with regularity and generally on a set schedule. This means that individuals receive frequent price signals. Frequent purchases may help beneficiaries learn about the pricing schedule. Secondly, individuals may be able to predict their short-term consumption when pharmaceutical treatments are prescribed over an extended period or for chronic disease. This contrasts with other types of health consumption, such as the diversity of inpatient and outpatient services in medical insurance plans (e.g. Brot-Goldberg et al. (2017)). Learning the price schedule may generate a higher payoff to the enrollees since current purchases have a greater certainty of being repeated as compared to treatments for a broken leg, for example. On the other hand, repeated purchases without accompanying price changes may also lull the enrollees into reduced price monitoring.

To provide a baseline framework for an individual's behavior and highlight the importance of the two regions of interest, consider the following stylized model for one purchase day. Each individual has a set of health conditions, $h \epsilon H_i$, which each have a utility cost of c_h. The individual can choose a drug to mitigate this utility cost from the set of drugs available to treat the condition, $j \epsilon J_h$, which includes the outside option of no purchase. Choosing drug j will counteract the individual's health condition utility cost by ϕ_{hij}, h, but will also have a monetary cost oop_{jn_i} to the individual. The monetary cost to the individual will vary as the individual moves along her purchase days, since the nonlinear price schedule changes as the individual's aggregated spending increases. Thus the out-of-pocket price varies both by the type of drug j and by the individual's count of aggregated purchases, n_i. Let n_i indicate individual i's aggregated number of purchase days, with the first purchase day under the plan's price schedule equal to 1.[4] The sum of drug purchases over all drugs j purchased on

[4]We could also treat the individual as making decisions based on the previous purchase day's information, or $n_i - 1$. Here, the effect of n_i is simply to place the individual within their own series of choices, so this interpretation would just be a normalization of the current structure.

day n_i must satisfy the budget constraint for income allocated to that day, I_{n_i}.[5] Out-of-pocket price oop_{jn_i} changes utility though a price salience / learning term σ_{n_i}.

The enrollee's problem on a filling day n_i is:

$$\sum_{h \epsilon H_i} \max_{j \epsilon J_h} (\phi_{hij} - c_h - \sigma_{n_i} \, oop_{jn_i}) \tag{1}$$

where the set of optimal choice js are subject to

$$\textstyle\sum_{h \epsilon H_i} oop_{jn_i} \leq I_{n_i}$$

The model highlights three behavioral responses. The first is the role of price salience, or learning about the price schedule, the second is the choice not to fill, and the third is the ability to substitute between drug choices.

The price salience term σ_{n_i} incorporates a role for an enrollee to learn about her price schedule from her purchases. The individual may initially mis-perceive her out-of-pocket price, so there is wedge between her true payment and what she believes her payment will be, $\sigma_{n_i} < 1$.[6] As the individual learns, perception approaches reality and σ_{n_i} increases, with $\sigma_{n_i} = 1$ being a perfectly correct understanding of the out-of-pocket incidence of her drug choice. Learning results from two components, the number of purchase days, n_i, and the difference in out-of-pocket price since the last visit, $\sum_{H_i} oop_{jn_i} - \sum_{H_i} oop_{jn_i - 1}$.

Define $\sigma_{n_i} \epsilon \, [0, 1]$ as

$$\sigma_{n_i} = f\big(\, n_i, \textstyle\sum_{H_i} oop_{jn_i} - \sum_{H_i} oop_{jn_i - 1} \big) \tag{2}$$

An increasing number of visits provides more price signals and more learning about the price schedule, so price salience increases with the number of purchase days. That is, $\partial \sigma_{n_i} / \partial n_i > 0$. Learning is likely highest for the individual's initial purchases; if the effect drops off, then $\partial^2 \sigma_{n_i} / \partial {n_i}^2 < 0$.

Frequent visits with the same out-of-pocket prices may lull the enrollee into not monitoring her place in the price schedule. For chronic conditions, individuals purchase drugs

[5] The role of the binding budget constraint requirement here is to maintain the existence of tradeoffs between spending dimensions for the individual within a purchase day. Income for the day can be interpreted broadly as any general limit to spending that would maintain these tradeoffs.

[6] In the doughnut hole threshold setting, out-of-pocket incidence only has potential to increase, so we do not model the opposite effect, where the individual could perceive her out-of-pocket price to be more expensive than it is in reality.

frequently, and the initial pricing period maintains consistent prices over large swaths of initial spending. However, a change in out-of-pocket price may pull the individual out of this complacency and motivate her to invest effort into learning about the price schedule more than on other purchase days. Thus, salience increases when there is a price increase, $\partial\sigma_n/\partial(\sum_{H_i} oop_{jn_i} - \sum_{H_i} oop_{jn_i-1}) > 0$. Larger changes should induce larger learning responses, $\partial^2\sigma_n/\partial(\sum_{H_i} oop_{jn_i} - \sum_{H_i} oop_{jn_i-1})^2 > 0$.

These two components of price salience suggest two distinct regions of behavioral response. In initial purchases, the effect of the number of visits should dominate, since out-of-pocket prices are constant in the price schedule during initial coverage. However, as the enrollee moves along the spending distribution and an additional visit has diminishing marginal impact, the largest shock to price salience will be the transition into the doughnut hole with the accompanying out-of-pocket price changes.

The second behavioral response above is whether the enrollee stops filling her prescriptions, choosing the outside option for j_h. This could be by filling less often, that is, stretching the time between purchase days, or by choosing to stop filling for certain diseases, h, within her set of diseases H_i. If the out-of-pocket cost of a drug increases enough, the negative effect on enrollee utility may not outweigh the benefit ϕ_{hij}. Here the total drug portfolio has great influence, where an enrollee with a heavily-branded portfolio will face a larger out-of-pocket price change at the doughnut hole compared to one with a generic portfolios, without a price change. The immediate perceived benefit of a drug, or the difference between curative benefit ϕ_{hij} and health cost c_h, may signal which disease categories are dropped from an enrollee's portfolio. Delaying or stopping fillings has health implications for drug adherence, but also has policy implications if it shows individuals stop filling versus make substitutions. A straightforward response of delaying fills implies that the doughnut hole is simply acting as a short-term subsidy rather than a useful price signal to encourage substitution in these disease categories.

Finally, we study drug substitution behavior. The model highlights two channels that could facilitate more substitution within drug choices. First, larger changes in the out-of-pocket price should encourage more searching for drug substitutes, i.e. expensive branded fills will have a larger percentage change in out-of-pocket price upon entering the donut hole compared to cheaper branded fills. Most importantly, the choice set of J_h for a given disease h expands or restricts individual substitution behavior. Switching may be easier if there are more generic substitutes available. Finally, each enrollee has her own ϕ_{hij} so that drug benefit can vary among the choices in J_h. The ability to switch to a cheaper drug may depend

on disease-level characteristics, and those enrollees who do not make the initial switch to a generic may have an atypical ϕ_{hij} that makes substitution more difficult. Ostensibly, the goal of these nonlinear price plans is to encourage enrollees to substitute to cheaper alternatives. Understanding how much substitution behavior appears possible is important to evaluating these policy goals.

4 Empirical Approach

We have identified several dimensions of behavioral response underlying a decrease in spending and two regions in particular where we should expect these responses to be most important. Empirically, however, two major difficulties arise when analyzing drug purchasing behavior within a nonlinear price schedule. The first problem is, when individuals differ by their end-of-year spending, they may expect different end-of-year prices to which they are responding in their current purchase. The second problem is that price and aggregated spending are linked through the price schedule, so there is selection by illness severity as the individual moves into a new pricing region. Teasing out the effect of price must also control for differences in underlying illness between individuals above and below the doughnut hole threshold.

To address the problem of different expected end-of-year prices, we create an estimation sample composed of individuals who pass into the doughnut hole region between March and August of the coverage year. Arriving at $2,510 by March means that the individual should have had a reasonable idea of passing into the doughnut region, and that she will remain in the doughnut hole region paying full price for the remainder of the year.[7] This also ensures that comparisons of any pre-doughnut hole behavior are based over a group of enrollees with uniform relative certainty of their end of year prices; all pre-doughnut hole behavior is that of individuals who ended up in the full price region solidly in the middle of their spending year.

We employ a discontinuity-based approach to controlling for underlying illness severity, as in Dalton et al. (2015). To identify the effect of learning or price changes, we track an individual's behaviors as she moves along her own spending distribution, and compare changes in her own behaviors before and after the discontinuous change in price at the doughnut hole. In this way, we can use individual fixed effects to control for underlying

[7]We exclude a small number of enrollees who ended the year in the catastrophic coverage region, who would have faced more generous end-of-year prices.

disease mix while identifying the magnitude of behavioral response specifically due to changes in plan structure along the spending distribution.

Our outcomes of interest are the dimensions of spending adjustment that match learning, dropping fillings, and substitution behaviors from Section 3. This includes outcomes such as responses to previous spending, number of fillings, time between purchase dates, and the mix of generic and branded purchases. Unless otherwise noted, an observation is a purchase day of an enrollee from the estimation sample, where a day is indexed by t and an individual is indexed by i. The estimation specification is:

$$
\begin{aligned}
Y_{it} = FE_i \quad & + \sum_{1}^{K} \lambda_k \text{ Coverage Region}_{kt} + \epsilon_{it} \qquad (3) \\
Y_{it} = FE_i \quad & + \lambda_1 \, \mathbb{1}\{-1,000 \leq m_{it} < -500\} \\
& + \lambda_2 \, \mathbb{1}\{-500 \leq m_{it} < 0\} \\
& + \lambda_3 \, \mathbb{1}\{0 \leq m_{it} < 500\} \\
& + \lambda_4 \, \mathbb{1}\{-500 \leq m_{it} < 1,000\} + \epsilon_{it}
\end{aligned}
$$

where Y_{it} is a dimension of interest in pharmaceutical spending observed at the enrollee purchase day level, including number of fillings, branded versus generic fillings, length of time between fillings, the number of disease categories included in the purchase day, and other dimensions of composition of enrollee's purchases. The λ coefficients are over a series of indicator variables for the coverage region in which the purchase day falls. Specifically, we divide coverage regions by how far away the individual's accumulated spending is from the plan's doughnut hole threshold, denoted as m_{it}. Accumulated spending before the doughnut hole has negative values, and m_{it} becomes positive after aggregate spending pushes above the threshold. For estimation around the doughnut hole region, we most commonly denote "Early Coverage" as between \$1,000 to \$500 away from the doughnut hole, "Late Coverage" is \$500 before and "Early Doughnut" is \$500 after, and "Late Doughnut" is \$500 beyond the threshold up to \$1,000 above. To examine initial purchases we will also include a "Beginning" period including early purchases up to the "Early Coverage" region.

To further examine enrollee behavior in the initial purchasing region, we also examine enrollees' price salience through shocks to spending. For this analysis, we interact the coverage region of the purchase day with a lagged indicator of high spending on the previous purchase day. This specification is:

$$Y_{it} = FE_i \quad + \sum_{1}^{K} \gamma_{k1} \{\text{Coverage Region}_{kt}\} \times 0\{\text{High Spending}_{t-1}\} \quad (4)$$
$$+ \sum_{1}^{K} \gamma_{k2} \{\text{Coverage Region}_{kt}\} \times \mathbb{1}\{\text{High Spending}_{t-1}\} + \epsilon_{it}$$

where the coefficients of interest are the set of γ_{k2}s which indicate the individual's response in today's purchases to unusually high spending in the previous purchase day. If there is learning over initial purchases, we should expect the γ_{k2} for the early coverage regions to be larger than the γ_{k2} in later coverage regions, as individuals show larger responses and take advantage of possible drug purchase adjustments early on.

5 Data

The plans in this analysis are employer-sponsored Medicare Part D plans provided to retirees and eligible employees as a part of the employer's benefit package. Nearly 7 million enrollees in Part D participate through employer-sponsored plans, or 15 percent of enrollment in 2008 (MedPac, 2009). The data here is obtained from Express Scripts, currently the largest pharmacy benefit manager (PBM) in the U.S. and includes all information about a drug filling, including the components of price, out-of-pocket incidence, the date of the filling, the method of filling, and the aggregate spending amounts as of that fill date. It also includes some basic demographic information about the patient including age and sex. These plans here are from two different employers and are of particular interest because they featured no change to the pricing of generic drugs before and after the threshold.

Table 1 lists the pricing schedules for the plans in this analysis, along with the standard structure set up as a baseline by the Centers for Medicare and Medicaid Services for Part D plans. A standard plan in 2008 featured a deductible, followed by a relatively generous coverage region up to a doughnut hole threshold of $2,510, after which all drugs, both generic and branded, were subject to full out-of-pocket pricing. The two plans here had relatively generous coverage up to doughnut hole thresholds of $2,510 and $2,600 for CL1B and CL3D, respectively. Both plans had more generous pricing for generics, a $10 copay for CL3D and full coverage for CL1B, which remained constant across all spending levels. Branded drugs, however, moved to full out-of-pocket costs after crossing the doughnut hole in both plans.

Table 1: Price Schedules for Medicare Part D Plans

Drug Type	Cost Structure					
	Deductible	Coverage Region OOP	Doughnut Hole Start	Doughnut Hole OOP	Catastrophic Coverage Start	Catastrophic Coverage OOP
Standard Plan						
Branded	\$275	\$30-\$50	\$2,510	100%	\$4,050	\$5.60
Generic	\$275	\$5-\$10	\$2,510	100%	\$4,050	\$2.25
CL1B Plan						
Branded	\$100	\$25-25%	\$2,510	100%	\$4,050	\$5.60
Generic	\$0	\$0	NA	\$0	\$4,050	\$2.25
CL3D Plan						
Branded	\$0	\$30-\$50	\$2,600	100%	\$4,050	\$5.60
Generic	\$10	\$10	NA	\$10	\$4,050	\$2.25

Table 2 lists summary statistics for our two plans. The top panel includes statistics for the full dataset, while the lower panel includes only those in the estimation sample. In total, we observe 18,961 and 6,889 individuals for CL3D and CL1B, respectively. Both the plans are about 60 percent female. An average enrollee in CL1B is about 70 years old with just over 9 different diseases types, and has average total spending of \$3,692, which means about 60 percent of these enrollees end up hitting the doughnut hole. CL3D enrollees are slightly older, and the average enrollee has 5.3 disease categories for which she fills over 1.6 scripts for each visit. CL3D enrollees are less likely to have spending bringing them into the doughnut hole, with only 20 percent of enrollees falling into a range of \$500 before the doughnut hole. Together, two plans contain over 25,800 enrollees.

The estimation sample is defined as those enrollees who hit the doughnut hole threshold between the beginning of March and the end of August. As such, this sample includes a subset of each plan which is slightly sicker. For example, the average ages are slightly higher than the plan average, and the average number of prescriptions filled per visit increase. There are a handful of individuals who in fact eventually leave the coverage gap and enter the catastrophic coverage region, which we remove from the sample.[8] We also remove any individuals who have a single filling day with spending over \$2,000. These individuals may have limited control over behaviors around the doughnut hole threshold since one filling moved the individual nearly half way through our total estimation window. This is only 31

[8]We remove the individuals who arrive at the catastrophic region because they may be considering different end-of-year prices in their doughnut hole region decisions.

individuals out of 3,506.

Table 3 displays the makeup of drug fills in terms of generic, branded, and maintenance characteristics. In total, we observe 260,045 fills in the estimation sample and around 80 percent of the fills are from CL1B enrollees. Generic fills make up 69 to 77 percent of these fills. Within branded fills, we split the fills into tiers based on their cost. A branded drug is categorized as "expensive" if the drug cost more than \$150 for a single filling. "Medium" brand drugs are those less than or equal to \$150 but greater than \$50 per filling. Finally, a drug is characterized as "cheap" if a single filling costs less than \$50. Both plans have the highest share in the medium branded fills, 59 and 49 percent for CL1B and CL3D, respectively. CL3D has a slightly higher share in expensive branded fill than CL1B. Both plans have around 80 percent of fills for maintenance drugs, drugs to treat chronic conditions, as opposed to drugs to treat acute conditions.

Table 4 lists the diseases most commonly indicated for the most-filled drugs in the two plans. The largest category, with over 39,000 fillings from the two plans, is drugs for Heart Disease/ High Blood Pressure. The most commonly filled drugs are generally for chronic diseases, such as high cholesterol and diabetes. In the top five disease types, only Pain is a category that may include also some acute uses. Other common disease categories are ulcer disease, depression, thyroid disorders, and infections.

6 Results

In this section, we first document how enrollees' spending changes along the aggregate spending distribution. Consistent with our theoretical model, the initial purchases and doughnut hole threshold show the greatest changes in spending behaviors. We then look into each of these regions to decompose these spending changes into components of enrollee's behavior: learning, quantity changes, and substitution behaviors.

6.1 Aggregate Spending Changes

We first establish how enrollees respond to the changes in price and incentives to switch to generic drugs based on the nonlinear price schedule. Figure 2 shows average weekly spending in the plans from initial purchases up to \$5,000 in aggregate drug spending for our estimation sample (i.e., all enrollees who hit the doughnut hole between March and August of the year and do not enter catastrophic coverage area). Each graph shows average weekly spending

Table 2: Summary Statistics by Plan and Overall

	CL1B	CL3D	Overall
Full Data			
Pct. Female	0.6	0.6	0.6
Avg. Age	70.36	76.22	74.66
Avg. Number Fillings per visit	2.11	1.63	1.75
Avg. Number of diseases types	9.19	5.33	6.36
% hitting early coverage ($500-1000 from doughnut hole)	0.82	0.28	0.42
% hitting late coverage ($0-500 from doughnut hole)	0.74	0.20	0.34
% hitting early doughnut ($0-500 passing doughnut hole)	0.64	0.12	0.26
% hitting late doughnut ($500-10000 passing doughnut hole)	0.52	0.07	0.18
% in Doughnut Hole	0.63	0.12	0.25
Avg. Total Spending	3,692.18	1,209.23	1,870.94
Avg. Total Out-Of-Pocket	966.92	486.08	614.22
Total Spending			
25th pctile	1,936.73	263.83	388.26
50 pctile	3,115.82	768.87	1,206.76
75th pctile	4,592.72	1,769.47	2,635.66
Out-of-Pocket Spending			
25th pctile	224.82	134.53	150.00
50 pctile	525.00	330.20	370.00
75th pctile	1,327.71	628.14	733.62
N	6889	18961	25850
Estimation Sample Only			
Pct. Female	0.61	0.62	0.61
Avg. Age	70.56	77.06	72.25
Avg. Number Fillings per visit	2.13	1.83	2.05
Avg. Number of diseases types	10.79	9.11	10.35
% hitting early coverage ($500-1000 from doughnut hole)	1	1	1
% hitting late coverage ($0-500 from doughnut hole)	1	1	1
% hitting early doughnut ($0-500 passing doughnut hole)	1	1	1
% hitting late doughnut ($500-10000 passing doughnut hole)	0.97	0.85	0.94
% in Doughnut Hole	1	1	1
Avg. Total Spending	5,091.82	4,124.78	4,840.81
Avg. Total Out-Of-Pocket	1506.06	1837.41	1592.07
Total Spending			
25th pctile	3,945.61	3331.24	3753.83
50 pctile	4,712.39	3897.99	4,495.21
75th pctile	5,859.33	4,650.88	5,554.29
Out-of-Pocket Spending			
25th pctile	743.69	1,139.53	869.53
50 pctile	1351.38	1,683.78	1,435.43
75th pctile	2116.08	2,400.50	2,195.49
N	2573	902	3475

Table 3: Structure of Fills by Plan and Overall (Estimation sample)

	CL1B	CL3D	Overall
Total fills	210,065	49,980	260,045
Share of generic fills	0.77	0.69	0.76
Share of branded fills	0.23	0.31	0.24
Share of expensive branded fills	0.29	0.42	0.32
Share of medium branded fills	0.59	0.49	0.57
Share of cheap branded fills	0.12	0.09	0.11
Share of fills for maintenance drugs	0.80	0.83	0.81

Table 4: Most Common Diseases Indicated for Drug Fillings (Estimation Sample)

Disease	Number of Fillings	Percent of Total Fills	Cumulative Percent
HIGH BLOOD PRESS/HEART DISEASE	39,101	15.05	15.05
PAIN	23,862	9.18	24.23
DIABETES	19,999	7.7	31.93
HIGH BLOOD CHOLESTEROL	17,770	6.84	38.77
DEPRESSION	15,114	5.82	44.58
ULCER DISEASE	11,760	4.53	49.11
HIGH BLOOD PRESSURE	10,748	4.14	53.25
INFECTIONS	8,746	3.37	56.61
THYROID DISORDERS	8,350	3.21	59.82
FLUID RETENTION	8,193	3.15	62.98
SEIZURES	7,583	2.92	65.9
VITAMINS AND MINERALS	6,487	2.5	68.39
HEART DISEASE	5,993	2.31	70.7
BLOOD MODIFYING	5,821	2.24	72.94
URINARY DISORDERS	5,519	2.12	75.06
ASTHMA	5,443	2.09	77.16
MENTAL/NEURO DISORDERS	5,372	2.07	79.23
ANTICOAGULANT	4,569	1.76	80.98
PAIN AND INFLAMMATION	4,446	1.71	82.69
ALLERGIES	4,243	1.63	84.33

with a smoothed trendline for one of the plans.[9] The top two panels show overall spending, with CL1B on the left and CL3D on the right. The first trend to note is that the overall spending drops significantly in the initial purchase stage (i.e., aggregate spending lower than $800) for both plans.[10] Following the striking initial drop, both plans show relatively flat spending levels between $1,000 until at least $1,500. Spending levels then drop significantly again at the doughnut hole, where CL1B shows a slightly more gradual adjustment toward the threshold than CL3D.

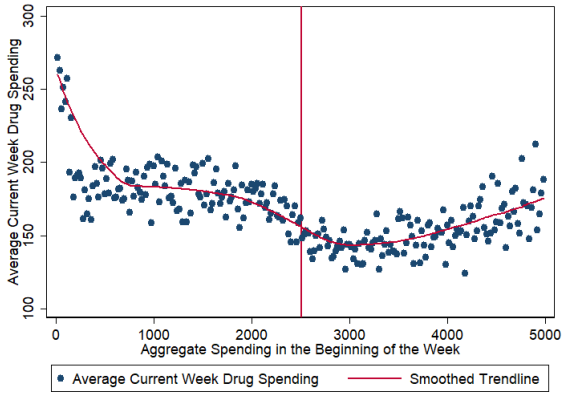
The middle panels focus only on branded spending, showing the average weekly branded spending as aggregate spending grows. Here, the most striking feature is that the branded spending drops closely mirror the overall drops in spending, indicating that nearly all the spending adjustments in the initial stage and at the threshold are driven by drops in branded spending. Generic behavior is displayed in the final two figures at the bottom of Figure 2. Generic spending behaviors have some differences between the two plans. CL3D shows no break in the trendline at either the initial stage or the threshold of doughnut hole, with a slightly upward trend by the end of the spending course. CL1B has a significant drop in generic spending in the initial stage, no change at the doughnut hole, and an upward trend after the threshold.

The spending trends suggest the most important regions for behavior changes are in the initial stage and at the doughnut hole, which is consistent with the implications of our theoretical model. Enrollees tend to respond to the nonlinear price schedule of branded drugs by reducing branded spending in the beginning while learning more information about price. At the doughnut hole, they reduce branded spending as a result of the out-of-pocket price changes for branded drugs. The change in generic spending looks to be modest overall, with some upward trends further into the doughnut hole. The reduction of generic spending in CL1B plan in the initial purchases may be due to learning about the total portfolio costs in addition to branded substitution learning.

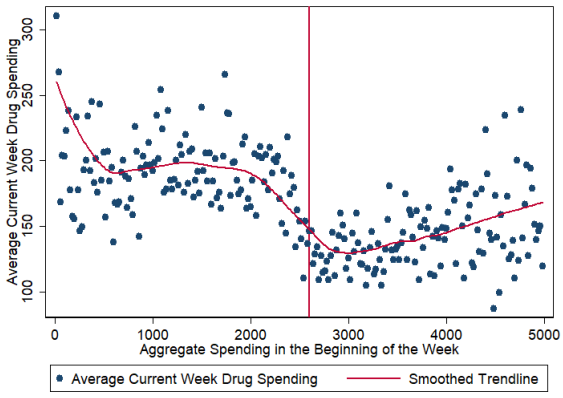
Table 5 displays these trends using the estimation method described in Section 4, with individual fixed effects and indicators for each price region. The table shows how individual behavior changes across seven coverage segments, which are each defined in relation to the plans' doughnut hole thresholds. "Beginning" is the initial $710 (for CL1B, and $800 for

[9] Trendline is from a Lowess regression of weekly spending on lagged aggregate spending, with a bandwidth of 0.03.

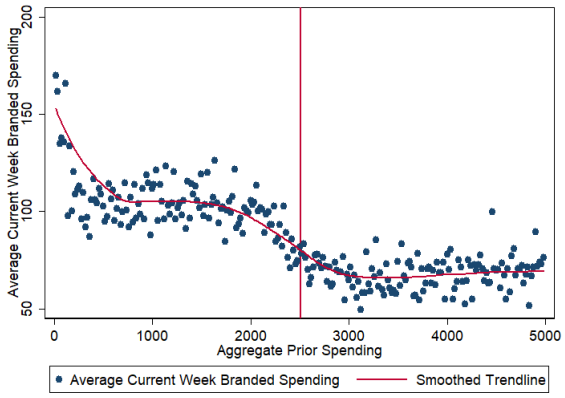
[10] We performed robustness checks by (1) excluding the first 3 purchases for every individual and (2) using other plans lacking generic coverage in the doughnut hole and find the same spending drop pattern in the initial purchases.



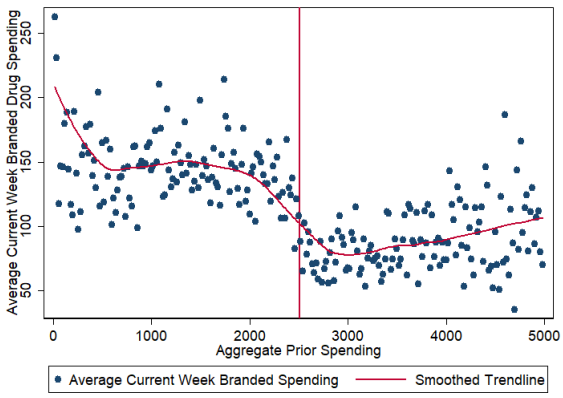
(a) Overall spending, CL1B



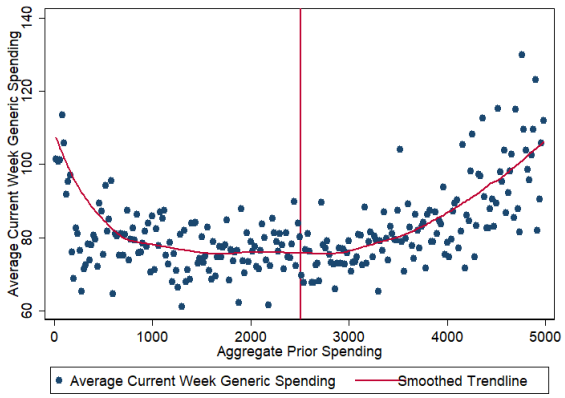
(b) Overall spending, CL3D



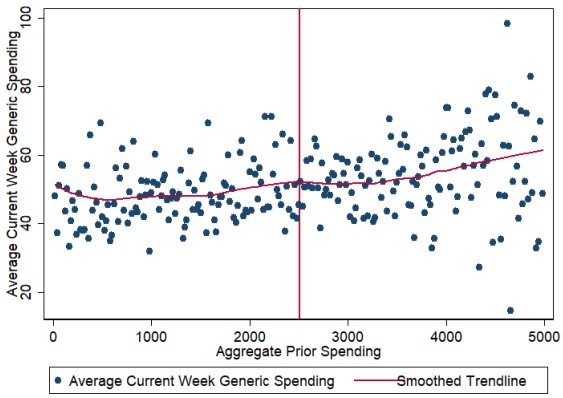
(c) Branded spending, CL1B



(d) Branded spending, CL3D



(e) Generic spending, CL1B



(f) Generic spending, CL3D

Figure 2: Different Types of Weekly Purchase by Aggregate Prior Spending

CL3D) spending, or when aggregate spending starts at $0 moving up to $1,800 below the doughnut hole threshold. "Coverage 1" is $1,800 to $1,200 below the plan's threshold, thus ranges from $710 to $1,310 in prior aggregate spending for CL1B and ranges from $800 to $1,400 in prior aggregate spending for CL3D. "Coverage 2" is $1,200 to $600 below the plan threshold and "Coverage 3" is $600 below up to the threshold level. Correspondingly, "Doughnut 1," "Doughnut 2," and "Doughnut 3" are defined as prior aggregate spending at the threshold up to $600 above, $600 to $1,200 above, and then $1,200 to $1,800 above, respectively. We make Coverage 2 a reference category. Recall from Figure 2, this region shows a flat trend in spending, so serves as a nice platform of comparison for changes.[11]

The first column of Table 5 shows three distinct regions of overall spending behavior. The coefficients report the change in spending in that region compared to the Coverage 2 region. In both the Beginning and Coverage 1 region, enrollees spend more in total spending during each purchase visit compared to the Coverage 2 region, but the magnitude is decreasing as they spend $26 more in the Beginning region while only $5 more in the Coverage 1 region, showing the drop in spending in the initial period to arrive at Coverage 2. Additionally, Beginning is the only region where enrollees spend more on branded drugs in each visit compared to Coverage 2 region. Beginning at $1,200 below the doughnut hole, there is very little adjustment in overall spending, including the Coverage 3 region, which shows only a $5 (or 4 percent) drop in spending per visit compared to the Coverage 2 region. However, there is a huge drop of $27 (or 20 percent) in overall spending after crossing into Doughnut 1 region, which continues into Doughnut 2 and 3 as well. This drop is mainly driven by a decrease of branded spending of $30. Generic spending increases slightly, by $3, making up the difference. This substitution behavior seems to persist further into the doughnut hole, into $1,200 to $1,800 above. Branded spending reacts by more than $33 compared to Coverage 2's levels when reaching $600 and $1,800 into the doughnut hole.[12]

[11] We have also performed the analysis by measuring days, instead of dollars, away from the doughnut hole threshold. The drop in spending crossing the doughnut hole is similarly present. However, defining day-based regions much beyond the threshold is problematic because of heterogeneity in the total number of days an enrollee is observed filling scripts. We choose instead to focus on the spending regions which correspond to price.

[12] The same regressions broken down by plan are reported in Appendix Table 18. The patterns are similar, with larger magnitudes of adjustment in the beginning for CL1B and at doughnut hole threshold for CL3D.

Table 5: Spending Changes Along Coverage Regions

	Total Spending	Branded Spending	Generic Spending
Beginning	26.08**	15.55**	10.53**
	(1.658)	(1.406)	(0.856)
Coverage 1	4.882**	2.045	2.837**
	(1.752)	(1.486)	(0.904)
Coverage 2	0	0	0
	(.)	(.)	(.)
Coverage 3	-5.231**	-6.879**	1.647
	(1.738)	(1.474)	(0.897)
Doughnut 1	-26.54**	-29.73**	3.191**
	(1.672)	(1.418)	(0.863)
Doughnut 2	-29.70**	-33.31**	3.604**
	(1.736)	(1.472)	(0.896)
Doughnut 3	-30.98**	-33.91**	2.935**
	(1.911)	(1.621)	(0.986)
Constant	131.9**	82.73**	49.20**
	(1.229)	(1.042)	(0.634)
N	113,117	113,117	113,117

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Beginning" is defined as prior aggregate spending from \$0 up to \$1,800 below the doughnut hole threshold; "Coverage 1" is prior aggregate spending \$1,800 - \$1,200 below the threshold, "Coverage 2" is \$1,200 - \$600 below, and "Coverage 3" is \$600 below up to the threshold; "Doughnut 1" is prior aggregate spending from the threshold up to \$600 above, "Doughnut 2" is \$600 to \$1,200 above, and "Doughnut 3" is \$1,200 to \$1,800 above.

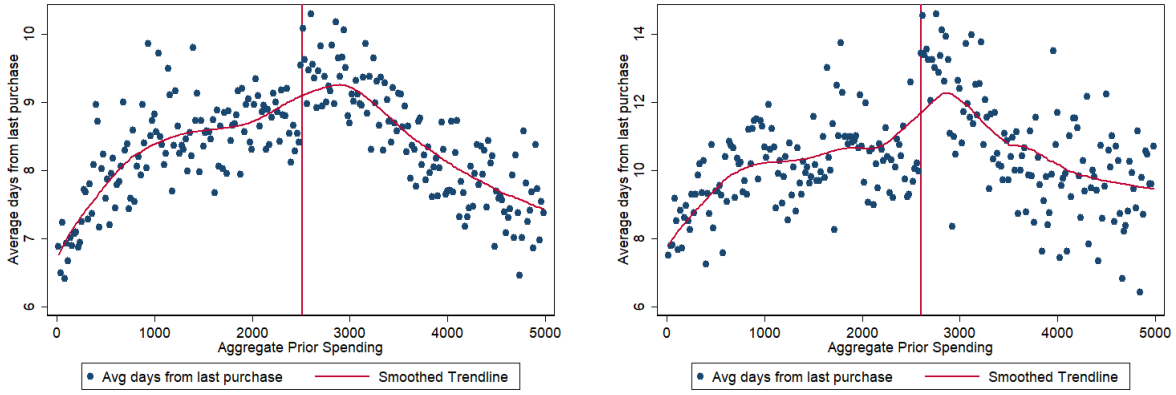
6.2 Evidence for Learning in Initial Purchases

Given the largest changes in spending are in the initial period and doughnut hole, we now turn to examine initial purchases. In particular, initial purchases are an important region in which to examine learning. Here, any changes in purchasing are not confounded by accompanying changes in current prices, as we might see in the doughnut hole. Relative tradeoffs between branded and generic purchases do not change during the initial period, yet we see spending responses as enrollees make more purchases in the plan, as suggested by the model.[13] Figure 2 shows that enrollees in both plans reduce their spending in their initial purchase stage (i.e., the initial $1,000 spending). In this section, we present evidence that this spending reduction results from learning via two channels: (1) enrollees learn through more frequent visits about the price and coverage schedule and (2) enrollees become more price salient after a large spending shock in the initial purchases.

Figure 3 shows the average days between visits matched along the aggregate spending distribution. The general shape is an inverted "V", meaning that the most frequent visits occur in the early spending periods, at about every 7-8 days. Enrollees from both plans visit more frequently in initial purchases, then slowly increase the number of days between fillings. The longest spacing between visits occurs just after crossing into the doughnut hole. The most rapid increase in time between visits occurs during the first $1,000 in spending. The next largest increase in delay occurs in the days around the doughnut hole, and the distance between visits gradually starts to decrease when further from the doughnut hole.

The change of visit pacing is greatest in the early period. Table 6 reports the fixed-effect estimation results for days between visits across coverage regions. The base category starts at $0 in prior aggregate spending and the final category ends at $1800 above the doughnut hole threshold. The coefficients report the change in days between visits in that region compared to the Beginning region. All coefficients show that the length of time between visits increases as aggregate spending increases. Along the spending distribution, overall or by plan, enrollees on average visit most frequently in the "Beginning" stage (at about every 8 days). The largest drop in visits is in between the initial purchases and the first coverage region. Enrollees tend to wait 0.7 more days in Coverage 1 compared to initial purchases. They delay a third of a day moving from Coverage 1 to Coverage 2, and another tenth of a day in the last coverage region. In initial entry into the doughnut hole, enrollees are waiting nearly two days longer to fill a prescription than they were in the initial purchases.

[13] Although the CL1B plan did have a deductible on branded drugs for the initial $100 period, we will take this into account below and show that these results are robust to spending after the initial $100.



(a) Average Days between Visits, CL1B

(b) Average Days between Visits, CL3D

Figure 3: Average Days Between Visits by Aggregate Prior Spending

Accompanied by the findings that spending drops precipitously during the initial period, this suggests that there is another important region for learning in pharmaceutical pricing. When enrollees visit frequently, they likely encounter more opportunities to learn about the pricing schedule of their plan and how their own portfolio of drug purchases fits into this pricing schedule. As enrollees match their spending to the pricing schedule, we see spending adjustments being made, and then visits dropping apace.

To further investigate how enrollees learn about their pricing and drug portfolio costs, we turn to examine the effect of a spending shock on enrollee purchase behavior. Specifically, we look at enrollee behavior after a daily purchase that falls into the upper 25th percentile of spending during this period. If the initial period's spending is being influenced by learning, we should expect the spending reactions to shocks to be more important or larger during the initial purchases.

We examine the effect of a shock on enrollees' choice of generics, both in shares of total spending and shares of total number of fills per visit. An enrollee is recorded as receiving a spending shock if her previous purchase day fell in the upper 25th percentile of average spending in the initial \$1,500 spending range for that plan. For CL1B, that means greater than \$160.43 on a fill day, or greater than \$169.15 on a fill day for CL3D. Because CL3D had a small deductible of \$100, we exclude the first \$100 of purchases in these regressions in order to avoid any possible effects on behavior from the deductible pricing. Table 7 reports the individual fixed effect regression results to explain the share of spending which is generic, and the share of total scripts filled which are generic. We use interaction terms between the

Table 6: Days between Visits Along Coverage Regions

	Total Spending	Branded Spending	Generic Spending
Beginning	0	0	0
	(.)	(.)	(.)
Coverage 1	0.737**	0.919**	0.649**
	(0.096)	(0.226)	(0.103)
Coverage 2	1.004**	1.436**	0.833**
	(0.095)	(0.228)	(0.102)
Coverage 3	1.106**	1.511**	0.940**
	(0.095)	(0.228)	(0.102)
Doughnut 1	1.834**	2.997**	1.425**
	(0.091)	(0.216)	(0.099)
Doughnut 2	1.389**	2.224**	1.115**
	(0.095)	(0.242)	(0.101)
Doughnut 3	0.940**	1.657**	0.705**
	(0.105)	(0.292)	(0.109)
Constant	8.003**	9.074**	7.717**
	(0.066)	(0.150)	(0.072)
N	111,086	25,974	85,112

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Beginning" is defined as prior aggregate spending from \$0 up to \$1,800 below the doughnut hole threshold; "Coverage 1" is prior aggregate spending \$1,800 - \$1,200 below the threshold, "Coverage 2" is \$1,200 - \$600 below, and "Coverage 3" is \$600 below up to the threshold; "Doughnut 1" is prior aggregate spending from the threshold up to \$600 above, "Doughnut 2" is \$600 to \$1,200 above, and "Doughnut 3" is \$1,200 to \$1,800 above.

lagged spending shock in the last purchase and each individual spending region to capture the heterogeneous responses in each region.

The estimation results support our hypothesis about the dissipation of learning effects in response to a spending shock along the spending distribution. The omitted category is enrollees who do not experience a spending shock in the initial $100-500 range. In this early range, those enrollees who experienced a spending shock at their previous filling end up with a 9 percentage point higher share of generics in their fills at their next visit, as compared with enrollees who did not experience such a shock. This corresponds with approximately 13 percentage points more generic in the spending share compared with those who did not have a shock. However, the effect of a shock on increasing generic spending decreases further into the aggregate spending distribution. The interaction between a shock in the previous period and spending in the $500-1,000 range only increases the generic fill share by 4.5 percentage points, or 6 percentage points of generic spending share. This reduced magnitude is similar for a shock in the $1,000-$1,500 region, at about 4 percentage points increase in generic use measured either in fillings or in spending. The effect of a shock in the initial period is double that of later periods. The relative pricing of branded versus generic is the same for each of these regions, however, the response to a shock is not. This suggests that initial purchases have more influence in changing behaviors, which is evidence for learning.[14]

6.3 Behavior Changes at the Doughnut Hole

In this section, we turn to the other two behavioral predictions from Section 3: the decision not to fill and substitution. For these behaviors, we examine the regions around the doughnut hole threshold, where prices change discontinuously. This means that each branded drug now has a larger impact on an enrollee's budget and that the relative prices of branded drugs have increased disproportionately compared to generic drugs. We define four coverage segments around the price change threshold, and examine the within-enrollee behavior as the enrollee moves between regions. Recall that "Early coverage" is $1,000 below the plan's threshold up to $500 below, "Late coverage" is $500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to $500 above, and then $500 above to $1000 above the threshold, respectively.

[14]We also find similar results with more detailed coverage segments and after including the initial $100 spending. We also perform a similar analysis of spending shocks only in terms of out-of-pocket spending, to consider only what the enrollee would see leave her wallet, versus the full information that she might glean from a pharmacy visit. We find similar behavior patterns with out-of-pocket spending, where the initial period's interacted coefficients are largest.

Table 7: Behavior After a Spending Shock: Beginning

	Generic filling share	Generic spending share
\$100-500 × High Spending$_{t-1}$=0	0	0
	(.)	(.)
\$100-500 × High Spending$_{t-1}$=1	0.0895**	0.127**
	(0.011)	(0.012)
\$500-1000 × High Spending$_{t-1}$=0	-0.0208**	-0.0326**
	(0.007)	(0.007)
\$500-1000 × High Spending$_{t-1}$=1	0.0451**	0.0632**
	(0.009)	(0.010)
\$1000-1500 × High Spending$_{t-1}$=0	-0.0278**	-0.0593**
	(0.006)	(0.007)
\$1000-1500 × High Spending$_{t-1}$=1	0.0434**	0.0424**
	(0.008)	(0.009)
Constant	0.703**	0.648**
	(0.005)	(0.006)
N	34,448	34,447

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: High Spending$_{t-1}$=1 if last fill day spending higher than \$160.43 for CL1B and higher than \$169.15 for CL3D.

Note: The initial \$100 spending is excluded from the analysis.

Table 8 reports the results of individual fixed-effect regressions for spending and filling behaviors before versus after the doughnut hole threshold. The omitted category of comparison is Early Coverage, or $1,000 to $500 below the threshold. The upper half of Table 8 indicates little to no spending adjustment happens in the coverage region immediately before the threshold, with coefficients close to zero and not significant across total spending, and within branded and generic spending. The biggest drop in spending occurs in the immediate region after the threshold, most of which is driven by a decrease in branded spending of nearly $19. However, the reduction in branded spending may emerge from either of two types of behavior: enrollees may buy fewer fills or instead may price shop and substitute to cheaper branded or generic drugs. Which mechanism dominates has different policy implications, where shopping for cheaper substitutes was part of the original motivation for a large out-of-pocket region, but simply stopping fillings may have adverse health outcomes. Our next set of analyses break this dollar decrease into behavioral components of the number of fillings and the type of fillings that result.

6.3.1 Number of Fillings

The lower half of Table 8 reports changes at the doughnut hole in number of fillings made per purchase day. Compared to the Early coverage region, an average enrollee drops nearly one-tenth of a branded filling after crossing the doughnut hole, and substitutes 0.08 of this with a generic. This balances out to a small, not statistically significant change in overall number of fillings per purchase day after the doughnut hole. Moving further into the doughnut hole, an average enrollee buys approximately 0.2 more generic fillings, but continues to drop almost a tenth of a branded filling. This means that total fillings in the late doughnut hole increase approximately one-tenth of a fill.[15]

The results from Table 8 are at the purchase-visit level. However, another dimension to reduce fillings for an enrollee would be to fill less frequently. Previous Table 6 discussed that this is also taking place, where the average enrollee waits more than half a day longer after reaching the doughnut hole to fill a prescription. Together, these results imply a significant reaction to the price changes at the doughnut hole, where enrollees fill both less frequently, and slightly fewer scripts on each delayed visit.

When fills decrease following the doughnut hole price changes, this could be because of dropping specific drugs or disease categories or because the enrollee simply fills the same drugs less often. Policy implications differ for each behavior. Delaying all fillings may have

[15] We also perform this analysis for the plans separately and the results are reported in Table 19 in Appendix.

Table 8: Spending and Filling Changes Around The Doughnut Hole

	Total Spending	Branded Spending	Generic Spending
Early coverage	0	0	0
	(.)	(.)	(.)
Late coverage	-0.511	-0.0904	-0.420
	(2.014)	(1.720)	(0.995)
Early doughnut	-15.86**	-18.68**	2.822**
	(1.926)	(1.645)	(0.952)
Late doughnut	-8.989**	-19.39**	10.40**
	(1.926)	(1.645)	(0.952)
Constant	133.4**	83.65**	49.73**
	(1.411)	(1.205)	(0.698)
	Total Fills	Branded Fills	Generic Fills
Early coverage	0	0	0
	(.)	(.)	(.)
Late coverage	0.0140	-0.00350	0.0175
	(0.018)	(0.008)	(0.016)
Early doughnut	-0.0135	-0.0926**	0.0791**
	(0.018)	(0.008)	(0.016)
Late doughnut	0.108**	-0.0928**	0.201**
	(0.018)	(0.008)	(0.016)
Constant	1.916**	0.534**	1.381**
	(0.013)	(0.006)	(0.011)
N	58,578	58,578	58,578

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.

small effects spread across disease categories, but dropping certain categories of disease can have heterogeneous health effects depending on the importance of care continuity in the prescribed treatment.

To examine if certain disease categories are dropped, we first compare the breakdown of fills for acute versus chronic conditions. More acute conditions may induce a differential price response based on a more immediate effect of the disease on the enrollee's perception of her health. We use an indicator variable of whether the drug is generally used to treat an acute diagnosis - to arrest a disease in progress, or whether the drug is generally prescribed for chronic diseases – a maintenance drug used to manage a constant condition. The welfare effects of changing fillings in the acute or maintenance category are different. We might expect fillings for chronic conditions to have less of an immediate impact on the enrollee's health perception, but to have a larger long-term health impact by affecting long-term management of the condition. Poor long-term management can lead to comorbidities of other related conditions.

Table 9 reports filling behavior for acute and maintenance drugs. The table lists changes in within-enrollee behavior both for fillings of acute and maintenance drugs as well as the share acute drugs contribute to the overall purchase day spending and filling portfolio. Neither category shows significant changes crossing the doughnut hole. This means that the makeup of fillings between acute and maintenance drugs does not seem to be an important dimension of enrollee behavior. There is a slight increase in maintenance fillings in the late doughnut region, which might reflect within-enrollee changes in disease accumulation, but the change in the accompanying share of maintenance drugs in the portfolio is very close to zero. One explanation for why we may not see behavioral changes on this dimension is that a vast majority of fillings in this population are for chronic disease, at over 80 percent of the fillings. A median enrollee purchase day does not even have an acute filling. Even with price changes, the acute versus maintenance balance does not have much impact.

We next discuss how the mix of diseases is affected by the drop in spending. We assess a count of the number of unique disease types attributable to an enrollee's set of drugs in each purchase date. Table 10 reports changes in the number of diseases for which an enrollee fills as she passes through the different coverage regions. We report both total overall numbers as well as the total within a subset of the top 20 diseases in the dataset, as previously reported in Table 4. Generally, entry into the doughnut hole is not accompanied with filling for fewer numbers of disease categories, either overall or when restricted to the top 20 diseases. If an individual finds themselves late in the doughnut hole, there is a slight positive coefficient,

Table 9: Acute versus Maintenance Drugs Across Coverage Regions

	Acute Fills	Maintenance Fills
Early coverage	0	0
	(.)	(.)
Late coverage	0.00191	0.0121
	(0.007)	(0.019)
Early doughnut	-0.00195	-0.0116
	(0.007)	(0.018)
Late doughnut	0.00199	0.106**
	(0.007)	(0.018)
Constant	0.354**	1.562**
	(0.005)	(0.013)
N	96,332	96,332

	Acute spending share	Acute fills share
Early coverage	0	0
	(.)	(.)
Late coverage	-0.00107	-0.00148
	(0.005)	(0.004)
Early doughnut	0.000743	-0.00109
	(0.004)	(0.004)
Late doughnut	-0.00680	-0.00958*
	(0.004)	(0.004)
Constant	0.225**	0.229**
	(0.003)	(0.003)
N	58,576	58,578

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.

indicating that an enrollee may accumulate more diseases over the course of the year.

Table 10: Number of Disease Types Filled

	All diseases			Top 20 diseases		
	Overall	CL3D	CL1B	Overall	CL3D	CL1B
Early coverage	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
Late coverage	0.00796	-0.0314	0.0200	0.00427	-0.0274	0.0139
	(0.015)	(0.027)	(0.018)	(0.014)	(0.027)	(0.017)
Early doughnut	-0.00378	-0.0395	0.00746	0.00651	-0.0187	0.0144
	(0.015)	(0.025)	(0.017)	(0.014)	(0.025)	(0.016)
Late doughnut	0.0954**	0.0401	0.111**	0.0957**	0.0628*	0.105**
	(0.015)	(0.027)	(0.017)	(0.014)	(0.027)	(0.016)
Constant	1.736**	1.611**	1.775**	1.446**	1.332**	1.481**
	(0.011)	(0.019)	(0.013)	(0.010)	(0.019)	(0.012)
N	58,578	13,847	44,731	58,578	13,847	44,731

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.

On a more detailed level, we examine specific disease categories to see if the overall lack of response in the count of diseases filled masks any underlying heterogeneity within individual diseases. Table 11 details filling behavior changes for the top five most common diseases, the top five chronic diseases, and the top 5 acute diseases. Diabetes and High Blood Pressure / Heart Disease (HBP/HD) fall into both the most common and most common chronic disease lists. These two categories are the only ones which show any significant changes in number of fills within an enrollee over the doughnut hole period. Diabetes fills drop by 0.03 in the early doughnut, and later by 0.05 in the late doughnut. The response of HBP/HD is even smaller, at only 0.02 decrease in fillings after crossing the doughnut hole threshold. For HBP/Heart Disease, we might expect less of a noticeable effect for the enrollee when filling compared to Depression or Pain, two of the other four top filling categories. Diabetes and HBP/HD may have more therapeutic alternatives, or Diabetes may have more peripheral purchases, such as testing accessories, which may be possible to use less intensively.

Table 11: Fillings On the Top Diseases

	Top 5 diseases				
	HBP/Heart Disease	Pain	Diabetes	High Choles.	Depression
Early coverage	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Late coverage	-0.0125	0.00538	0.00271	-0.00110	0.00434
	(0.010)	(0.012)	(0.016)	(0.009)	(0.010)
Early doughnut	-0.0232*	-0.00695	-0.0304*	-0.0120	-0.00287
	(0.010)	(0.012)	(0.015)	(0.008)	(0.010)
Late doughnut	-0.00652	0.00939	-0.0468**	-0.0113	0.0162
	(0.010)	(0.012)	(0.015)	(0.008)	(0.010)
Constant	1.243**	1.168**	1.237**	1.080**	1.085**
	(0.007)	(0.009)	(0.011)	(0.006)	(0.007)
N	13,892	8,046	7,083	7,188	5,629
	Top 5 maintenance diseases				
	HBP/Heart Disease	Diabetes	High Choles.	Depression	Ulcer Disease
Early coverage	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Late coverage	-0.0125	0.00271	-0.00110	0.00434	0.00546
	(0.010)	(0.016)	(0.009)	(0.010)	(0.007)
Early doughnut	-0.0232*	-0.0304*	-0.0120	-0.00287	-0.00878
	(0.010)	(0.015)	(0.008)	(0.010)	(0.007)
Late doughnut	-0.00652	-0.0468**	-0.0113	0.0162	0.00728
	(0.010)	(0.015)	(0.008)	(0.010)	(0.007)
Constant	1.243**	1.237**	1.080**	1.085**	1.031**
	(0.007)	(0.011)	(0.006)	(0.007)	(0.005)
N	13,892	7,083	7,188	5,629	4,649
	Top 5 acute disease				
	Pain	Infections	Inflam./Immune disorder	Pain/Cramps	Sleep Disorder
Early coverage	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Late coverage	0.00538	-0.0301*	0.000573	-0.00664	-0.0115
	(0.012)	(0.013)	(0.019)	(0.015)	(0.011)
Early doughnut	-0.00695	-0.0283*	0.0289	-0.00953	0.00771
	(0.012)	(0.012)	(0.018)	(0.014)	(0.010)
Late doughnut	0.00939	-0.0103	0.00538	0.00159	0.00226
	(0.012)	(0.013)	(0.017)	(0.014)	(0.011)
Constant	1.168**	1.062**	1.023**	1.036**	1.011**
	(0.009)	(0.009)	(0.012)	(0.010)	(0.007)
N	8,046	3,459	1,164	1,407	925

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.

6.3.2 Substitution Patterns

We have established that spending decreases substantially, and that this decrease does not seem to be principally driven by dropping particular disease categories. If an enrollee does not drop a disease category, spending may decrease instead through substitution of the filling into a cheaper drug choice. In this section, we investigate if enrollees are more likely to substitute to generic or cheap branded drugs when price changes at the doughnut hold threshold. First, we analyze potential price shopping behavior. That is, whether enrollees are more likely to substitute an expensive branded drug to a cheaper branded option or to a generic drug. Second, we examine heterogeneous substitution behaviors by disease type. Finally, we stratify substitution behavior based on the availability of generic drugs within a disease category.

Table 12 reports the within-enrollee filling behavior changes across coverage regions. The left hand panel shows the changes in fillings broken down by overall, branded and generic, while the right hand panel breaks down the subset of branded drugs only. Branded drugs are split into three price categories. Recall that a branded drug is categorized as "expensive" if the drug cost more than \$150 for a single filling, "medium" if it cost less than or equal to \$150 but greater than \$50 per filling, and "cheap" if it cost less than \$50 for a single filling.

As shown above, the average enrollee drops approximately 0.09 of a branded filling, and only replaces about 0.08 of this drop with generics. The right side panel reveals more about which branded drugs see the most change. Enrollees drop both expensive branded and medium branded drugs when going through the doughnut hole, with a slightly larger drop occurring in the medium branded category of 0.05. There is essentially no change in the cheap branded category. However, the average price of cheap branded drug is only \$29, which is almost the same as a generic drug. To put these cheaper drugs in the context of the plan's tiered copays, 40% of the cheap drugs in CL1B have an associated copay of \$25 pre-threshold, which is the copay associated with the "Preferred Brand" tier. Up to 93% of the cheap drugs in CL1B pre-threshold have a copay of \$25 or less. If a drug's cost is less than the copay, the usual payment for the enrollee is just the cost of the drug, not the full amount of the copay. A similar story holds for CL3D, with over 95% of the pre-threshold copays for cheap drugs being at or below the Preferred Brand copay. Therefore, the cheap drug category likely saw very little change in realized prices even after the nonlinear threshold came into place. The observed changes to the enrollee in the cheap brand likely appear similar to those in the generic-only fillings, so more realistically appeared as a linear price schedule both before and after the threshold. However, the fact that the cheap drug spending did not increase after

the threshold is indication that very little within-brand substitution occurred, at least from expensive or medium to cheap branded fills.

Table 12: Filling by Branded Drug Types Across Coverage Regions

	Total	All Generic	All Brand	All Brand		
				Expensive Brand	Medium Brand	Cheap Brand
Early coverage	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
Late coverage	0.0140	0.0175	-0.00350	0.00524	-0.0116	0.00282
	(0.018)	(0.016)	(0.008)	(0.005)	(0.006)	(0.003)
Early doughnut	-0.0135	0.0791**	-0.0926**	-0.0421**	-0.0533**	0.00271
	(0.018)	(0.016)	(0.008)	(0.005)	(0.006)	(0.003)
Late doughnut	0.108**	0.201**	-0.0928**	-0.0404**	-0.0591**	0.00672*
	(0.018)	(0.016)	(0.008)	(0.005)	(0.006)	(0.003)
Constant	1.916**	1.381**	0.534**	0.189**	0.296**	0.0496**
	(0.013)	(0.011)	(0.006)	(0.004)	(0.004)	(0.002)
N	58,578	58,578	58,578	58,578	58,578	58,578

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: A branded drug is categorized as "expensive" if the drug cost more than $150 for a single filling, "Medium" if cost less than or equal to $150 but greater than $50 per filling, and "cheap" if cost less than $50 for a single filling.

Note: "Early coverage" is $1,000 below the plans threshold up to $500 below, "Late coverage" is $500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to $500 above, and then $500 above to $1000 above the threshold, respectively.

Table 13 reports changes in branded/generic filling behavior for the top 5 most common diseases reported as branded drug fills before the threshold. The top half of the table speaks to changes in branded filling, while the bottom half shows if these branded drops ended up as generic substitutions. Branded fills drop in all top five branded drug disease, with four of the categories showing drops throughout the doughnut hole, and Blood Modifiers showing drops eventually, in the late doughnut region. The largest drops occur in Diabetes, at 0.06 less of a branded fill. Both High Cholesterol and Asthma drugs see over 0.05 a branded fill drop, and HBP/HD drops by 0.03 of a filling.

Correspondingly, the changes in generic fillings for those diseases are reported in the lower half of Table 13. Enrollees do not substitute to generic fillings immediately for Diabetes and

HBP/HD after the price change at the threshold, but they do fill more generics as they move into the late doughnut region. However, the increased amount of generic fills does not compensate for the total decrease in branded fillings. The substitution to generic fillings for High Cholesterol happens immediately in the early doughnut, and nearly replaces the full amount of branded filling losses. For Asthma, there is not much change in generic fillings, though there is a decrease in branded filling after entering into the doughnut hole region.

Lastly, we examine heterogeneous substitution behavior by a measure of generic substitution availability of the 23 most common disease categories filled in the data. Using pre-threshold behavior, the 23 disease categories are divided into four categories based on the empirically observed ability to substitute to a generic option within that disease category. These 23 disease categories represent 88.7 percent of total fillings in the data. In order to construct categories of generic substitution ability, we observe the rates of generic consumption versus branded consumption in the region before the doughnut hole. This gives an empirical measure of the rate of tradeoff between generic and branded before the dramatic change in the relative prices between branded and generic. Table 14 shows the 23 disease categories matched with the pre-threshold rate of generic filling. The final column lists a label for the four categories of generic substitution ability we use in our empirical investigation.

There is substantial heterogeneity in observed generic substitution in the 23 disease categories. Some disease categories have very high availability of generics, with nearly all pre-threshold fillings of that category being a generic fill. These include Pain, High Blood Pressure, and Vitamins and Minerals. Any disease category with over 97 percent generic fillings pre-threshold we denote as Complete percent generic, with little generic substitution possibility since nearly all fillings start with a generic choice. The second category is High generic prevalence, where generic fillings made up 90 to 92 percent of total fillings in pre-threshold fillings. This High generic category includes Heart Disease, Ulcer Disease, and Thyroid disorders. The next category has a greater potential for generic substitution, with only between 65 and 88 percent of scripts being filled with a generic before the threshold. This Medium category includes Diabetes, High Blood Pressure/Heart Disease, High Cholesterol, and Depression. The final category of diseases has a very low rate of generic use before the threshold, which likely implies that there are not very many generic substitutes available or that switching is difficult for an average enrollee. This Low generic category ranges from 15 percent to 33 percent generic in the fills pre-threshold. The largest generic shares are in Bone Conditions, Glaucoma, and Asthma.

Because the Complete generic category has little to no generic substitution possibility, we

Table 13: Fillings for the Top 5 Diseases in Branded Drug

	Number of Branded Fills				
	Diabetes	HBP/Heart Disease	High Choles.	Blood Mod.	Asthma
Early coverage	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Late coverage	-0.00166	-0.0107	-0.00583	0.0143	-0.0169
	(0.018)	(0.008)	(0.010)	(0.021)	(0.028)
Early doughnut	-0.0590**	-0.0336**	-0.0486**	-0.0354	-0.0534*
	(0.018)	(0.007)	(0.009)	(0.021)	(0.027)
Late doughnut	-0.0778**	-0.0431**	-0.0760**	-0.0418*	-0.0836**
	(0.018)	(0.007)	(0.010)	(0.021)	(0.028)
Constant	0.659**	0.215**	0.382**	0.987**	0.918**
	(0.013)	(0.005)	(0.007)	(0.015)	(0.019)
N	7,083	13,892	7,188	2,217	2,095
	Number of Generic Fills				
	Diabetes	HBP/Heart Disease	High Choles.	Blood Mod.	Asthma
Early coverage	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
Late coverage	0.00437	-0.00176	0.00473	-0.00992	0.0126
	(0.016)	(0.012)	(0.010)	(0.010)	(0.021)
Early doughnut	0.0286	0.0104	0.0367**	-0.000117	0.0251
	(0.015)	(0.011)	(0.009)	(0.009)	(0.021)
Late doughnut	0.0310*	0.0366**	0.0648**	0.0201*	0.00323
	(0.015)	(0.011)	(0.010)	(0.010)	(0.022)
Constant	0.578**	1.028**	0.698**	0.139**	0.221**
	(0.011)	(0.008)	(0.007)	(0.007)	(0.014)
N	7,083	13,892	7,188	2,217	2,095

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.

Table 14: Disease Categories and Empirically Observed Generic Consumption Pre-Threshold

Disease category	Maintenance vs. Acute	Percent of Generic Fills	Generic Prevalence
PAIN	A	98	complete
HIGH BLOOD PRESSURE	M	99	complete
FLUID RETENTION	M	100	complete
VITAMINS AND MINERALS	M	100	complete
ANTICOAGULANT	M	97	complete
INFLAMMATION/IMMUNE DISORDERS	A	100	complete
INFECTIONS	A	90	high
ULCER DISEASE	M	91	high
THYROID DISORDERS	M	92	high
HEART DISEASE	M	90	high
PAIN AND INFLAMMATION	M	90	high
HIGH BLOOD PRESS/HEART DISEASE	M	88	medium
HIGH BLOOD CHOLESTEROL	M	78	medium
DIABETES	M	65	medium
DEPRESSION	M	80	medium
SEIZURES	M	82	medium
ALLERGIES	M	78	medium
BLOOD MODIFYING	M	15	low
BONE CONDITIONS	M	53	low
GLAUCOMA	M	33	low
MENTAL/NEURO DISORDERS	M	16	low
URINARY DISORDERS	M	40	low
ASTHMA	M	24	low

focus on the High, Medium, and Low generic categories. To study generic substitution, we use the same previous individual-level fixed effect approach, but instead study the outcome of the choice to fill with a generic drug versus a branded option. We run fixed effect logits for choice of generic along the four regions before and after the doughnut hole using the earlier divisions of Early Coverage, Late Coverage, Early Doughnut, and Late Doughnut encompassing a total window of $2,000 of spending. Table 15 shows the logit coefficients indicating the change in probability of filling a generic drug among the four coverage regions.

The upper third of Table 15 shows the results for generic choice for disease categories with high rates of generic fills pre-threshold. Heart disease shows the most notable change here, with negative coefficients along the coverage regions, indicating substitution away from generics as an individual passes further into the doughnut hole region. This might be because generics have a high availability within this category, so it is more likely that patients are initially prescribed generic drugs. Patients only switch away to branded in the presence of problems with their generic prescription. The other categories show very little response either before or immediately after the threshold. There is some movement to generic drugs in the Late doughnut category, with positive coefficients in for Ulcer Disease, Pain and Inflammation, and Thyroid Disease. This may occur because budgets become tighter, which we will explore in the next section.

The Medium generic level category shows much more switching activity after the threshold price change, as shown in the middle third of Table 15. All diseases in this category actually show increased probability of filling a generic upon entering the late doughnut hole region. Four of the six, High Cholesterol, Diabetes, Depression, and Seizures, also show large early positive coefficients, indicating generic substitution immediately following the threshold price change.

The disease categories in Low generic availability show a much more mixed response in generic filling compared with the disease categories with a mid-range availability of generics. Because generics are seldom used in these disease categories pre-threshold, the ability to switch to generics following the threshold is likely low. This is borne out in the regression results in the lower third of Table 15. Only half of the six categories, Bone Condition drugs, Glaucoma, and Mental/Neuro disorders show significant increase in the probability of filling a generic after the threshold price change. Bone Conditions and Mental/Neuro disorders both show continued or growing generic substitution later into the doughnut hole, but it should also be noted that these categories are less populated than the diseases in the "Medium" availability category.

Table 15: Substitution Patterns by Generic Availability Disease Category

	"High" Availability					
	Infections	Ulcer Disease	Thyroid Disease	Heart Disease	Pain and Inflam.	
Choice of generic						
Early coverage	0	0	0	0	0	
	(.)	(.)	(.)	(.)	(.)	
Late coverage	-0.225	0.159	1.287	-0.942**	-0.434	
	(0.240)	(0.347)	(0.700)	(0.311)	(0.653)	
Early doughnut	0.328	0.659	1.340	-1.014**	0.267	
	(0.240)	(0.339)	(0.706)	(0.305)	(0.650)	
Late doughnut	0.366	1.613**	1.556*	-0.866**	2.180**	
	(0.254)	(0.350)	(0.707)	(0.297)	(0.770)	
N	657	333	92	458	107	
	"Medium" Availability					
	HBP/HD	High Choles.	Diabetes	Depression	Seizure	Allergies
Choice of generic						
Early coverage	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
Late coverage	0.0472	0.157	0.0150	0.00749	0.247	0.888
	(0.105)	(0.188)	(0.103)	(0.223)	(0.310)	(0.666)
Early doughnut	0.109	0.901**	0.223*	0.765**	1.166**	0.165
	(0.102)	(0.184)	(0.101)	(0.226)	(0.307)	(0.608)
Late doughnut	0.273**	1.404**	0.204*	1.465**	1.733**	2.144**
	(0.103)	(0.193)	(0.101)	(0.242)	(0.323)	(0.760)
N	3,623	1,145	3570	780	430	102
	"Low" Availability					
	Blood Mod.	Bone Cond.	Glaucoma	Mental Neuro	Urinary	Asthma
Choice of generic						
Early coverage	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
Late coverage	-0.724	0.633	0.423	-0.165	0.309	0.136
	(0.568)	(0.675)	(0.332)	(0.408)	(0.336)	(0.265)
Early doughnut	-0.00772	2.039**	0.699*	1.220**	0.585	0.229
	(0.515)	(0.706)	(0.330)	(0.404)	(0.330)	(0.257)
Late doughnut	1.018	2.935**	0.490	1.490**	0.987**	0.0858
	(0.540)	(0.705)	(0.346)	(0.410)	(0.334)	(0.270)
N	133	164	332	287	327	575

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Fixed effect logit. Dependent variable equal to 1 if generic is chosen, 0 otherwise.

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.

6.4 Discussion and Alternative Explanations

Simply observing a change in spending following the doughnut hole may not be enough detail to determine the policy effectiveness of a nonlinear pricing scheme. In this work, we have decomposed a drop in aggregate spending into filling frequency, types of drugs filled, and substitution behaviors in order to learn more about the types of behavior induced by the nonlinear pricing of Medicare Part D. The first dimension of spending adjustment is the relatively-ignored region of initial purchases and its contribution towards learning about the plan's pricing. We show that one of the largest drops in spending occurs over the first few purchases by enrollees, below $500. This corresponds with an increasing dropoff in the frequency with which enrollees fill their prescriptions. We present this as evidence of the ability for the enrollee to learn about costs in their drug portfolio in a new coverage year and the nonlinear pricing. More frequent visits facilitate this learning early on. This means an important part of generic substitution and learning can be found in initial purchases, not only leading up to or at the doughnut hole.

For purchases at the doughnut hole, enrollees seem to make fewer fills, but mainly because of an increase in delay between fillings. We did not find evidence of large differences within categories of acute versus chronic disease or evidence for dropping entire disease categories in fills at the individual level. Within the branded fillings, both very expensive and middle range drugs saw significant changes, but the switch into generics was not complete. We see that just making generic drug coverage more generous has heterogenous effects. The availability of generic substitutes or the ease of substitution varies among disease categories and heavily influences any generic changes that happen as prices increase at the doughnut hole. In particular, the most consistent generic switching occurred in the "Medium" generic availability division. This might be because categories with nearly all generic fillings do not have capacity to switch, while categories with high availability might have early switching to generic, so any patient still using branded at the doughnut hole is not a good candidate for the generic option.

One consideration merits review regarding the initial purchase period. An alternative explanation for a rapid drop in spending at the beginning of the year is that the patients could be stockpiling in their early purchases. Perhaps enrollees had waited for the new plan to start, and their first purchases are items they wanted to stockpile for use over the year. We rule out this stockpiling story from three aspects. First, Lowess graphs of the weekly spending distribution which exclude the first three purchases still show a continuing trend of decreasing weekly purchases, so this stockpiling story would have to have a decreasing

component as well, which is generally not in line with an idea of preplanning large initial purchases. Second, we examine the types of diseases indicated for the early fillings. In Table 16, the top half shows that the number of scripts filled per visit does not change significantly from the initial purchases before $500 to the period up to $1,000. The lower half of the table shows that the number of diseases found in a purchase day also does not change when exiting initial spending. This means that the reductions in initial spending are not quantity driven by adding extra fills into a purchase day or by filling for an uncommonly-filled disease category. The spending decreases must instead then occur either through less frequent visits or through substitution to cheaper drugs. Finally, we might think that stockpiling behavior could occur for a patient's less-frequently filled diseases, which are only filled when the pricing schedule resets. To examine this, we estimate the individual fixed effect model for the number of fillings among diseases ranked by commonality in the data. Table 17 reports filling changes over an initial $1,500 spending. We find no reduction in the number fillings for diseases in the least common categories of diseases between the initial period and later. The types of diseases which are filled in the initial period look similar to those filled later in the purchase cycle.

7 Conclusion

Large out-of-pocket cost areas have been gaining traction in insurance plans such as Medicare Part D as an attempt to influence enrollee behavior and control health care costs. Here, we examine a setting with great richness in both the dimensions of enrollee behavior but also a high potential for enrollees to react to price changes because of the frequent and relatively predictable nature of pharmaceutical spending in chronic disease. We present here a general framework that could be adapted for other settings of nonlinear pricing, where we control for individual-level disease mix through fixed effect regression, and examine the effects of price changes through the within-individual reactions moving through different pricing regions. We control for expectations about end-of-year price by collecting a sample of individuals who all pass through the doughnut hole between March and August, then use this fixed-effect approach to break down the dimensions of behavior which lead to our observed decreases in spending.

In particular, we highlight a new region where enrollees appear to be learning about prices and respond with spending decreases: the period of initial purchases. Enrollees visit more often, and react more sharply to large price shocks in this region than later. At the

Table 16: Initial Period Behaviors

Fillings per purchase day			
	Overall	CL3D	CL1B
\$100-500	0	0	0
	(.)	(.)	(.)
\$500-1000	0.00382	0.0640	-0.0154
	(0.022)	(0.039)	(0.026)
\$1000-1500	0.214**	0.214**	0.213**
	(0.021)	(0.038)	(0.025)
Constant	1.962**	1.754**	2.029**
	(0.016)	(0.029)	(0.020)
N	36,569	9,025	27,544

Disease categories per purchase day			
	Overall	CL3D	CL1B
\$100-500	0	0	0
	(.)	(.)	(.)
\$500-1000	-0.0195	0.0311	-0.0357
	(0.018)	(0.032)	(0.021)
\$1000-1500	0.152**	0.145**	0.154**
	(0.017)	(0.031)	(0.021)
Constant	1.786**	1.615**	1.841**
	(0.014)	(0.024)	(0.016)
N	36,569	9,025	27,544

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Table 17: Fills on the top 5, 6-10, 11-15, 16-20, 20+ diseases

	Top 5	Top 6-10	Top 11-15	Top 16-20	21+
$100-500	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)
$500-1000	0.0191	0.00673	0.0267	0.0190	-0.0158
	(0.016)	(0.013)	(0.014)	(0.015)	(0.013)
$1000-1500	0.0842**	0.0205	0.0617**	0.0410**	0.00632
	(0.016)	(0.013)	(0.013)	(0.015)	(0.013)
Constant	1.549**	1.231**	1.182**	1.159**	1.191**
	(0.012)	(0.010)	(0.010)	(0.011)	(0.009)
N	20,590	10,267	7,264	5,930	9,217

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

doughnut hole threshold, we document large decreases in spending, which is supported by previous work. Breaking down these spending decreases, enrollees seem to be reducing their costs by simply filling the same mix of drugs less frequently, as opposed to dropping fills for certain diseases. There is some substitution to generic drugs, but because individual disease categories are heterogenous in the availability of generics, just making generic coverage more generous is shown to have limited ability to influence switching behaviors.

We hope this detailed behavioral analysis will layer on top of dynamic optimization discussions to help fill out a picture of consumer pharmaceutical spending. In particular, understanding that fills occur with less frequency, but across all drug categories, combined with myopa findings, tells us that the doughnut hole region here may actually be functioning more as a fixed subsidy system for those with high branded spending rather than a within-year method of inducing generic substitution.

References

Abaluck, J. and Gruber, J. (2011). Heterogeneity in choice inconsistencies among the elderly: Evidence from prescription drug plan choice. *American Economic Review*, 101(3):377–381.

Abaluck, J. and Gruber, J. (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *Americal Economic Review*, 106(8):2145 – 2184.

Abaluck, J., Gruber, J., and Swanson, A. (2015). Prescription drug utilization under Medicare Part D: a linear model of nonlinear budget sets. NBER Working Working Paper 20976.

Aron-Dine, A., Einav, L., Finkelstein, A., and Cullen, M. (2015). Moral hazard in health insurance: Do dynamic incentives matter? *Review of Economics and Statistics*, 97(4):725–741.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2012). Salience theory of choice under risk. *Quarterly Journal of Economics*, 127(3):1243–1285.

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. *Quarterly Journal of Economics*, 132(3):12611318.

Chandra, A., Gruber, J., and McKnight, R. (2010). Patient cost-sharing and hospitalization offsets in the elderly. *American Economic Review*, 100:193–213.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: theory and evidence. *American Economic Review*, 99(4):1145–1177.

Dafny, L., Ody, C., and Schmitt, M. (2017). When discounts raise costs: the effect of copay coupons on generic utilization. *American Economic Journal: Economic Policy*, 9(2):91–123.

Dalton, C. M., Gowrisankaran, G., and Town, R. (2015). Myopia and complex dynamic incentives: Evidence from Medicare part D. NBER Working Working Paper 21104.

Duggan, M. G. and Morton, F. S. (2010). The effect of Medicare Part D on pharmaceutical prices and utilization. *American Economic Review*, 100(1):590–607.

Einav, L., Finkelstein, A., and Schrimpf, P. (2015). The response of drug expenditures to contract design in Medicare Part D. *Quarterly Journal of Economics*, 130(2):841–899.

Fung, V., Mangione, C. M., Huang, J., Turk, N., Quiter, E. S., Schmittdiel, J. A., and Hsu, J. (2010). Falling into the coverage gap: Part D drug costs and adherence for Medicare Advantage prescription drug plan beneficiaries with diabetes. *Health Service Research*, 45(2):355–375.

Heiss, F., McFadden, D., and Winter, J. (2010). Mind the gap! Consumer perceptions and choices in Medicare Part D prescription drug plans. In Wise, D., editor, *Research Findings in the Economics of Aging*. University of Chicago Press.

Ho, K., Hogan, J., and Morton, F. S. (2017). The impact of consumer inattention on insurer pricing in the Medicare Part D program. *RAND Journal of Economics*, 48(4):877–905.

Hoadley, J., Cubanski, J., and Neuman, T. (2015). Medicare Part D at 10 years. Kaiser family foundation issue brief, Kaiser Family Foundation.

Joyce, G. F., Zissimopoulos, J., and Goldman, D. P. (2013). Digesting the doughnut hole. *Journal of Health Economics*, 32(6).

Jung, K., Feldman, R., and McBean, A. M. (2013). Demand for prescription drugs under non-linear pricing in Medicare Part D. *International Jounral of Health Care Finance and Economics*, 14(1):19–40.

Ketcham, J. and Simon, K. (2008). Medicare Part D's effects on elderly patients' drug costs and utilization. *American Journal of Managed Care*, 14(11 Suppl):SP14–22.

Ketcham, J. D., Lucarelli, C., Miravete, E. J., and Roebuck, M. C. (2012). Sinking, swimming, or learning to swim in Medicare Part D. *American Economic Review*, 102(6):2639–2673.

Ketcham, J. D., Lucarelli, C., and Powers, C. A. (2015). Paying attention or paying too much in Medicare Part D. *American Economic Review*, 105(1):204–233.

Kling, J., Mullainathan, S., Shafir, E., Vermeulen, L., and Wrobel, M. (2012). Comparison friction: Experimental evidence from Medicare drug plans. *Quarterly Journal of Economics*, 127(1):1180–1210.

Lichtenberg, F. and Sun, S. (2007). The impact of Medicare Part D on prescription drug use by the elderly. *Health Affairs*, 16:1735–1744.

Liu, F., Alexander, C., Crawford, S., Hedeker, D., Pickard, S., and Walton, S. (2011). The impact of Medicare Part D on out-of-pocket costs for prescription drugs, medication utilization, health resource utilization, and preference-based health utility. *Health Services Research*, 46(4):1104–1123.

MedPac (2009). Report to the Congress: Medicare payment policy. A status report on Part D for 2009.

Polyakova, M. (2016). Regulation of insurance with adverse selection and switching costs: evidence from Medicare Part D. *American Economic Journal: Applied Economics*, 8(3):165–195.

Schroeder, M. C., Brooks, J. M., and Cook, E. A. (2014). Heart attacks and subsequent medication choices: the effect of Medicare Part D benefit phase and time of year. Mimeo, University of Iowa.

Yin, W., Basu, A., Zhang, J. X., Rabbani, A., Meltzer, D. O., and Alexander, G. C. (2008). The effect of the Medicare Part D prescription benefit on drug utilization and expenditures. *Annals of Internal Medicine*, 148(3):169–177.

Zhang, Y., Donohue, J. M., Lave, J. R., O'Donnell, G., and Newhouse, J. P. (2009a). The effect of Medicare Part D on drug and medical spending. *New England Journal of Medicine*, 361:52–61.

Zhang, Y., Donohue, J. M., Newhouse, J. P., and Lave, J. R. (2009b). The effects of the coverage gap on drug spending: A closer look at Medicare Part D. *Health Affairs*, 28(2):317–325.

8 Appendix results

Table 18: Spending Changes Along Coverage Regions By Plan

	Total Spending			Branded Spending			Generic Spending		
	Overall	CL3D	CL1B	Overall	CL3D	CL1B	Overall	CL3D	CL1B
Beginning	26.08**	15.33**	29.58**	15.55**	14.56**	15.51**	10.53**	0.762	14.06**
	(1.658)	(3.700)	(1.843)	(1.406)	(3.443)	(1.496)	(0.856)	(1.405)	(1.036)
Coverage 1	4.882**	2.156	5.623**	2.045	2.162	1.893	2.837**	-0.00623	3.729**
	(1.752)	(3.991)	(1.933)	(1.486)	(3.713)	(1.568)	(0.904)	(1.515)	(1.087)
Coverage 2	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Coverage 3	-5.231**	-6.021	-4.924**	-6.879**	-9.248*	-6.096**	1.647	3.227*	1.172
	(1.738)	(4.026)	(1.907)	(1.474)	(3.746)	(1.547)	(0.897)	(1.529)	(1.072)
Doughnut 1	-26.54**	-41.63**	-21.74**	-29.73**	-48.10**	-23.83**	3.191**	6.466**	2.087*
	(1.672)	(3.811)	(1.844)	(1.418)	(3.546)	(1.496)	(0.863)	(1.447)	(1.037)
Doughnut 2	-29.70**	-46.30**	-25.14**	-33.31**	-48.35**	-29.24**	3.604**	2.049	4.105**
	(1.736)	(4.208)	(1.882)	(1.472)	(3.915)	(1.527)	(0.896)	(1.598)	(1.058)
Doughnut 3	-30.98**	-46.70**	-27.03**	-33.91**	-48.82**	-30.33**	2.935**	2.120	3.304**
	(1.911)	(5.011)	(2.040)	(1.621)	(4.663)	(1.656)	(0.986)	(1.903)	(1.147)
Constant	131.9**	143.9**	128.3**	82.73**	108.3**	74.90**	49.20**	35.54**	53.37**
	(1.229)	(2.830)	(1.351)	(1.042)	(2.633)	(1.096)	(0.634)	(1.074)	(0.760)
N	113,117	26,431	86,686	113,117	26,431	86,686	113,117	26,431	86,686

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Beginning" is defined as prior aggregate spending from \$0 up to \$1,800 below the doughnut hole threshold; "Coverage 1" is prior aggregate spending \$1,800 - \$1,200 below the threshold, "Coverage 2" is \$1,200 - \$600 below, and "Coverage 3" is \$600 below up to the threshold; "Doughnut 1" is prior aggregate spending from the threshold up to \$600 above, "Doughnut 2" is \$600 to \$1,200 above, and "Doughnut 3" is \$1,200 to \$1,800 above.

Table 19: Spending and Filling Changes Around Doughnut Hole (by plan)

	Total Spending			Branded Spending			Generic Spending		
	Overall	CL3D	CL1B	Overall	CL3D	CL1B	Overall	CL3D	CL1B
Early coverage	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Late coverage	-0.511	-1.324	-0.175	-0.0904	-2.072	0.595	-0.420	0.747	-0.771
	(2.014)	(4.716)	(2.200)	(1.720)	(4.396)	(1.801)	(0.995)	(1.746)	(1.184)
Early doughnut	-15.86**	-24.81**	-12.98**	-18.68**	-29.93**	-15.06**	2.822**	5.121**	2.080
	(1.926)	(4.398)	(2.121)	(1.645)	(4.099)	(1.736)	(0.952)	(1.628)	(1.142)
Late doughnut	-8.989**	-19.31**	-6.038**	-19.39**	-28.93**	-16.59**	10.40**	9.619**	10.55**
	(1.926)	(4.609)	(2.092)	(1.645)	(4.297)	(1.713)	(0.952)	(1.707)	(1.126)
Constant	133.4**	145.5**	129.6**	83.65**	110.0**	75.51**	49.73**	35.55**	54.12**
	(1.411)	(3.262)	(1.548)	(1.205)	(3.040)	(1.267)	(0.698)	(1.208)	(0.833)
	Total Fills			Branded Fills			Generic Fills		
	Overall	CL3D	CL1B	Overall	CL3D	CL1B	Overall	CL3D	CL1B
Early coverage	0	0	0	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Late coverage	0.0140	-0.0180	0.0239	-0.00350	-0.00687	-0.00220	0.0175	-0.0111	0.0261
	(0.018)	(0.033)	(0.022)	(0.008)	(0.018)	(0.010)	(0.016)	(0.029)	(0.020)
Early doughnut	-0.0135	-0.0483	-0.00256	-0.0926**	-0.128**	-0.0814**	0.0791**	0.0792**	0.0788**
	(0.018)	(0.031)	(0.021)	(0.008)	(0.017)	(0.009)	(0.016)	(0.027)	(0.019)
Late doughnut	0.108**	0.0551	0.123**	-0.0928**	-0.122**	-0.0841**	0.201**	0.177**	0.207**
	(0.018)	(0.032)	(0.021)	(0.008)	(0.017)	(0.009)	(0.016)	(0.028)	(0.019)
Constant	1.916**	1.760**	1.964**	0.534**	0.589**	0.517**	1.381**	1.171**	1.447**
	(0.013)	(0.023)	(0.015)	(0.006)	(0.012)	(0.007)	(0.011)	(0.020)	(0.014)
N	58,578	13,847	44,731	58,578	13,847	44,731	58,578	13,847	44,731

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.

Table 20: Weekly Spending and Filling Changes Around Doughnut Hole

	Total Spending	Branded Spending	Generic Spending
Early coverage	0	0	0
	(.)	(.)	(.)
Late coverage	-4.121	-2.478	-1.643
	(2.896)	(2.425)	(1.402)
Early doughnut	-25.92**	-27.68**	1.763
	(2.750)	(2.303)	(1.331)
Late doughnut	-12.37**	-26.98**	14.61**
	(2.716)	(2.275)	(1.315)
Constant	187.1**	117.5**	69.65**
	(2.025)	(1.696)	(0.980)
	Total Fills	Branded Fills	Generic Fills
Early coverage	0	0	0
	(.)	(.)	(.)
Late coverage	-0.00630	-0.00910	0.00279
	(0.028)	(0.012)	(0.024)
Early doughnut	-0.0752**	-0.134**	0.0585*
	(0.026)	(0.011)	(0.023)
Late doughnut	0.139**	-0.126**	0.265**
	(0.026)	(0.011)	(0.023)
Constant	2.668**	0.743**	1.926**
	(0.019)	(0.008)	(0.017)
N	43,446	43,446	43,446

Standard errors are in parentheses

* $p < 0.05$, ** $p < 0.01$

Note: "Early coverage" is \$1,000 below the plans threshold up to \$500 below, "Late coverage" is \$500 below up to the threshold level; "Early doughnut" and "Late doughnut" are at the threshold up to \$500 above, and then \$500 above to \$1000 above the threshold, respectively.