Take-Up, Drop-Out, and Spending in ACA Marketplaces

Rebecca Diamond[1] *, Michael J. Dickstein[2] *, Timothy McQuade[1] and Petra Persson[3] *

[1]Stanford Graduate School of Business
[2]Stern School of Business, New York University
[3]Department of Economics, Stanford University
*NBER

First draft: September 5, 2017
Last compiled: November 9, 2018

Abstract

The Affordable Care Act (ACA) established health insurance marketplaces where consumers can buy individual coverage. Leveraging novel credit card and bank account micro-data, we identify new enrollees in the California marketplace and measure their health spending and premium payments. Following enrollment, we observe dramatic spikes in individuals' health care consumption. We also document widespread attrition, with roughly half of all new enrollees exiting coverage before the end of the plan year. Some enrollees who drop out re-time discretionary health spending to the months of insurance coverage. This drop-out behavior can generate a new type of adverse selection: insurers face high costs relative to the premiums collected when they enroll strategic consumers. We develop a model to illustrate how this pattern of attrition can undermine market stability and lead to substantial price increases, even absent differences in enrollees' underlying health risks. Further, using data on plan price increases, we show that insurers largely shift the costs of attrition to non-drop-out enrollees, whose inertia generates low price sensitivity. Our results suggest that campaigns to improve use of social insurance may be more effective when they jointly target take-up and attrition.

We would like to thank Anais Galdin, Ziao Ju, Jiayu Lou, Lucienne Oyer, and Ryan Shyu for outstanding research assistance. Diamond appreciates the support of the Laura and John Arnold Foundation. We would also like to thank seminar participants at the 2017 AEA Annual Meeting, NYU, Stanford University, RAND, UNC-Chapel Hill, Duke University, University of Michigan, and Texas A&M University for helpful comments and suggestions.

1 Introduction

Social insurance programs are large, government-provided or government-subsidized insurance schemes that protect constituents against adverse risks to their income or health. One important component of social insurance design is to ensure that insurance de facto "reaches" the intended beneficiaries. Empirical evidence suggests this targeting is imperfect: take-up of social insurance is often low among intended beneficiaries, due to factors including stigma, a lack of information, and transaction costs (Currie, 2006).[1] The academic interest in take-up is also mirrored in real-world policy; many changes to the administration of social insurance programs are motivated in part by a desire to increase participation among intended beneficiaries.

Focusing only on take-up, however, misses a crucial ingredient to the success of social insurance: intended beneficiaries who take up insurance must *stay* enrolled for the program to be effective. The importance of studying drop-out in addition to take-up seems most relevant in settings in which enrollees must pay a premium for coverage, such as in SCHIP, Medicare Part D, and ACA health insurance plans. In settings where attrition is large, campaigns to boost take-up may have little effect on ultimate coverage. In this paper, we take a two-sided approach. We measure how individuals both take up and drop out of insurance coverage, and examine the consequences for the value of insurance.

We study attrition in the context of government-subsidized individual health insurance. The Patient Protection and Affordable Care Act (ACA), passed in March 2010, aimed to expand health insurance coverage in the United States. In particular, to increase take-up of insurance by individuals, the ACA established federal and state health insurance marketplaces where individual consumers could shop for health coverage. Among other elements, the law also regulated the types of plans available for purchase and the ability of insurers to reject applicants.[2] As a result of these reforms—and a large promotional campaign to encourage take-up—millions of Americans enrolled in marketplace health plans.[3] In 2014 and 2015, the first years of the marketplaces, the share of US residents covered by individual market insurance rose by 50% and 75%, respectively, relative to 2013.

This paper provides a nuanced analysis of what happened thereafter. While we document sharp increases in new enrollees' consumption of health care, we also observe widespread attrition, with more than half of all new enrollees dropping out before the end of the plan year. Consumers may leave coverage when they no longer view the insurance as valuable or obtain more comprehensive or cheaper forms of insurance, including employer-sponsored plans or public insurance through state Medicaid programs. They may also act strategically, consuming extra health care or re-timing health expenditures to the initial months of

[1] Currie (2006) reviews take-up of various programs, which is approximately 75% for Earned Income Tax Credit (EITC) and as low as 6-14% for the State Children's Health Insurance Program (SCHIP). Also see Kleven and Kopczuk (2011) and Bhargava and Manoli (2015).

[2] The ACA regulated the market for individual plans on both the supply and demand sides. On the supply side, insurers are now required to issue plans to all consumers who apply during an annual open enrollment period, without regard to a consumer's preexisting conditions. On the demand side, most US citizens and legal residents who do not receive insurance through an employer or government program must purchase plans in the individual insurance market or face a tax penalty. The tax penalty is set to fall to $0 in 2019.

[3] The promotional campaign included television ads, radio ads, Internet campaigns, and was spearheaded by President Barack Obama himself, who championed insurance on numerous television programs.

coverage and discontinuing coverage in the remaining months of the year. Using a theoretical framework, we show that drop-out behavior that involves re-timing can undermine the stability of the market for individual insurance and drive up premiums. Using data on insurance premiums by plan, we establish empirically that *non*-drop-out enrollees bear the brunt of the costs associated with drop-out behavior, through increased insurance premiums.

Conducting an empirical analysis of enrollee behavior in the individual insurance market is challenging because researchers typically face data constraints. The marketplaces feature a range of different insurance plans, administered by various insurers. In California, for example, ten different insurance companies offered plans on the exchange in 2014, the first year of open enrollment. Thus, beneficiaries' claims data are disaggregated across insurers. Further, claims data never contain information on individuals before or after enrollment in health insurance, making it challenging to study the causes and consequences of enrollment and drop-out.

We overcome this challenge by exploiting novel credit card and bank account micro-data. These data come from a firm that provides financial software to banks. Banks that adopt this software then allow our data provider to observe the credits and debits on users' bank and credit card accounts. Thus, selection into our dataset depends on the identities of banks who partner with our financial software company, and not on household decisions. Many of the largest US banks partner with our financial software company.[4]

Within these data, we focus on 850,000 account holders in California, the state with the largest individual insurance market. We identify new enrollees in the California individual insurance market by capturing premium payments made to the insurers participating on the Covered California marketplace.[5] By directly observing premium payments, we circumvent the challenge that information about who enrolls is disaggregated across insurers. Further, we exploit the text description of each financial transaction to capture health and drug spending. More precisely, starting from the universe of each account holder's credit card and bank transactions, we apply a machine learning algorithm to identify out-of-pocket health care spending and out-of-pocket drug spending. Our estimates of annual out-of-pocket health care spending obtained using this method closely match the Medical Expenditure Panel Survey, validating our data and methods. For a subset of our analysis that focuses on households with children, we use a similar approach to distinguish households that have (young) children.

Leveraging this unique data, we observe a sharp increase in new enrollees (new premium payments) at the end of 2013, corresponding to the start of the ACA's open enrollment period. The new enrollees are generally poorer households that receive government premium subsidies, suggesting that the marketplaces *de facto* served some of the populations that, before the ACA, were the most likely to be uninsured.[6]

[4]Unlike credit card and bank account data that come from platforms requiring active opt-in, our data suffers from no selection on usage of personal finance software or on other measures of financial sophistication. This feature of the data is important, as the low- to moderate-income households who stand to benefit most from the ACA may be the least likely to use personal finance software and, consequently, to share data with on-line platforms. As we discuss in more detail in Section 3, our sample is a random sample of active account holders, with the sample size corresponding to 7 percent of the California population.

[5]We describe the process through which we identify enrollees in detail in Section 3. The resulting market shares by each insurer in our data closely match those reported by Covered California.

[6]According to the National Health Interview Survey published by the Centers for Disease Control, 20.4% of U.S. households

In the first part of our analysis, we study the behavior of new ACA enrollees in terms of their health care spending and attrition rates. Once new enrollees gain insurance coverage, we observe sharp spikes in their consumption of health care: the number of health transactions increases 24% and the dollar value of out-of-pocket health care and drug spending increases. We can use the estimated increase in the quantity of health care transactions as a measure of the moral hazard effect of ACA insurance.[7] Using a difference-in-difference design that compares the change in health care consumption of new enrollees to the change in a "reference group" of households that never held individual insurance, we find ACA coverage increases health care consumption by over 50% among households with less than $20k of income, while the effect declines to 27% for households with $40k-$60k of income and declines further, to zero, for households with $100k-$200k of income.[8]

While new enrollees generally increase their health care spending once covered, they rarely pay consistent monthly premiums. Instead, attrition is widespread: only about half of all new enrollees in the 2014 and 2015 open enrollment periods pay a full year of premiums, with the sharpest drop-out observed after only *one month* of payment. Attrition is large even among the poorest newly enrolling households who receive government premium subsidies. The rate of drop-out also appears to be growing in each subsequent enrollment period among new enrollees. We re-estimate the health care consumption effects of ACA coverage focusing on those who drop out and find large moral hazard effects. Not only does health care consumption increase after enrollment, but this consumption is concentrated in the months of coverage, and then reverts back to the pre-enrollment level after dropping out. Indeed, short-term enrollees consume *more* health care per month during their period of coverage than those who enroll for more than a year.

Finally, we compare this behavior to the observed health care consumption around sign-up and drop-out in the period prior to the opening of ACA insurance exchanges. Households could purchase individual insurance prior to the ACA, but insurance companies could also set premiums based on an enrollees' prior health. In the period before the ACA, we observe fewer dropouts. Among those who do drop out, our moral hazard estimates are much smaller. The distinct pattern suggests ACA reforms made strategic sign-up and drop-out easier.

Consumers may exit marketplace plans for other forms of insurance, including state Medicaid programs. To gauge the drivers of the drop-out behavior, we first assess whether affordability or changes in household income explain the pattern we observe. Specifically, we analyze whether households experience decreases in income around the time they drop coverage. Interestingly, while we find drop-out to be strongly correlated with income losses in the *pre*-ACA period, we find a substantially weaker correlation post-ACA.

---

were uninsured in 2013.

[7] We validate this method using data from the Medical Expenditure Panel Survey (MEPS) where all health care charges are observed, not just out-out-pocket charges. While the MEPS sample is too small to study the ACA insurance in detail, it shows that increases in the count of health care transactions tracks closely the increase in total health charges for those who sign up for ACA coverage and for the overall MEPS sample.

[8] As we discuss in detail in Section 4, we use the term "reference group" as opposed to "treatment group" to underscore that we are not assuming quasi-random assignment of households into the group of ACA enrollees. Indeed, a key aspect of our study is to examine the reasons why certain households sign up for, and potentially leave, ACA-provided insurance. Utilizing this reference group ensures that the changes to ACA enrollees' health care consumption do not simply reflect time trends in the overall population.

This change suggests the ACA lessens the need for households that suffer temporary economic hardship to drop individual market coverage. Second, we explicitly study how household-level predicted eligibility for California's Medicaid's program, Medi-Cal, evolves around the time households drop coverage. We find modest increases in Medi-Cal eligibility around drop-out. These changes in Medi-Cal eligibility are small relative to the share of households that drop out, suggesting that churn into Medi-Cal does not account for an economically important share of the attrition from ACA plans. Finally, to argue that at least some of the drop-out we observe arises when consumers re-time expenditures, we hire Mechanical Turk workers to score the words we observe in the transaction descriptions as more or less essential and/or urgent. We then compare the spending of households that drop coverage and households that remain insured for the entire plan year, before and after enrolling in a marketplace plan. We find the average household that ultimately drops coverage appears to re-time less essential and less urgent health care spending to the period in which they maintain coverage.

Having documented the scope of attrition from the health insurance plans purchased in the individual market, we next present a theoretical framework to analyze how the presence of drop-out enrollees affects market stability. The model illustrates that the presence of strategic drop-out can destabilize the market and yield substantial plan price increases. This pattern arises even in a health insurance market in which all consumers are *ex ante* identical in their expected annual health care expenditures; that is, even when we eliminate the traditional adverse selection channel.

Our description of consumers as multi-dimensional – where households willing to drop out also have strong moral hazard effects – is related to past empirical work testing for sources of informational advantages in health markets. Finkelstein and McGarry (2006) demonstrate advantageous selection in the market for long-term care insurance, driven by the fact that consumers of low risk also have risk preferences that generate a taste for insurance. Fang, Keane, and Silverman (2008) find that consumers who purchase Medigap plans are also advantageously selected, in part because the decision to purchase depends on cognitive ability; in their setting, more cognitively able enrollees are healthier. Shepard (2016) finds that plans with generous hospital networks are adversely selected, as consumers with preferences for "star" hospitals incur higher medical expenses when sick. In our setting, the unknown proportion of the population that will re-time its health care expenditures and subsequently drop coverage generates adverse selection.

In the final part of the paper, we examine the pass-through of the costs associated with the presence of drop-out types. We show that *non*-drop-out enrollees bear the brunt of the costs associated with drop-out behavior through increased insurance premiums. Non-drop-out enrollees become inertial in their health plan choices in future years, allowing insurers to charge high mark-ups. These dynamics suggest that the cost of providing ACA health insurance could be reduced by adopting alternative penalty designs. One example is setting different penalty rates for non-insurance coverage depending on whether a household never signs up for ACA health insurance or whether it signs up and subsequently drops coverage mid-year. Charging a higher penalty for lack of insurance coverage after dropping out of ACA coverage could help alleviate the upward pressure on costs driven by the presence of enrollees who strategically drop out.

Our paper contributes to an emerging literature on the consequences of the ACA.[9] Antwi, Moriya, and

[9] Outside of the literature on the ACA but also related is work by Finkelstein, Hendren, and Shepard (2017), who analyze low-

Simon (2015) and Simon, Soni, and Cawley (2017) analyze changes in health behaviors exploiting variation from the ACA's Medicaid expansion. In Medicaid, the drop-out problem is less relevant, as enrollees are covered as long as they remain qualified. Dafny, Gruber, and Ody (2015), Dickstein et al. (2015), and Tebaldi (2017) provide early evidence on the relationship between entry and pricing, but characterize neither drop-out behavior nor examine the effect of dropouts on market stability.

More broadly, our analysis of drop-out behavior relates to the literature on selection on moral hazard in insurance markets. The novelty in our setting is that this adverse selection arises due to enrollees' ability to drop out of coverage, even in the absence of heterogeneity in risk. This finding relates most closely to work by Cabral (2017), who shows that re-timing of claims in the dental market can generate adverse selection, albeit through a different mechanism. In her setting, consumers strategically delay treatments to minimize out-of-pocket costs, but cannot drop out. In our setting, in contrast, households consume care upon sign-up, and subsequently cease to pay premiums.

The remainder of the paper proceeds as follows. In Section 2, we begin by describing how the ACA reformed the market for individual insurance in the United States. Section 3 describes our data and the machine-learning algorithm that we use to identify out-of-pocket health and drug spending. In Section 4, we study the spending behavior of new ACA enrollees and document widespread drop-out. In Section 5 we develop a framework to examine how the presence of dropouts affects the stability of the individual insurance market, and illustrate adverse selection and insurers' responses via premium setting. Section 6 examines the pass-through of the costs associated with the presence of drop-out types through premium increases, and Section 7 discusses penalty design in light of this threat to market stability. Section 8 concludes.

2 Institutional Details

We examine consumer behavior in the individual insurance market. In 2013, prior to the implementation of the Affordable Care Act (ACA), just 4% of the U.S. population purchased individual insurance coverage. Of the remainder, 50% purchased coverage through their employer, 33% received coverage through public insurance programs, including Medicaid, Medicare, and military and veterans health care, and 13% lacked insurance coverage (The Kaiser Family Foundation, 2016). After the ACA, in 2014 and 2015, the share covered by individual market insurance rose to 6 and 7%, respectively, an increase in percentage terms of 50 and 75% relative to 2013. We describe the key features of the ACA that led to these coverage changes in the United States. We then look more closely at regulations of the individual market in California, the setting for our empirical work.

2.1 The Affordable Care Act

The US Congress passed The Affordable Care Act in March 2010 with the goal of expanding health insurance coverage in the United States and reducing the number of uninsured. The law altered the insurance system

income households' willingness to pay for health insurance using administrative data from Massachusetts' subsidized insurance exchange. They do not document drop-out behavior, however.

in four key ways. First, it expanded Medicaid, the means-tested insurance program, to all Americans under 138% of the federal poverty guidelines beginning in 2014.[10] Second, the ACA mandated that most US citizens and legal residents obtain health insurance coverage. For those not covered by public insurance or insurance through their employer, this mandate required that individuals purchase plans in the individual insurance market or face a tax penalty.[11] Third, the act imposed requirements on private health plans. Non-group health insurance plans, for example, were required to issue plans to all applicants, regardless of past illnesses, cover ten categories of essential health benefits, and eliminate lifetime or annual limits on the dollar value of coverage.[12] Fourth, the ACA required employers with 50 or more full time employees to offer health coverage or face a tax penalty (The Kaiser Family Foundation, 2018).

We focus on changes to the market for individual insurance plans. Prior to the passage of the ACA, the requirements on insurers serving the individual market differed by state. In June 2012, for example, California, along with 31 other states, had no rate restrictions on premiums in the individual market. In 2012, only six states required insurers to "guarantee issue" individual insurance to any applicant (The Kaiser Family Foundation, 2012).

The ACA harmonized the rules across states, eliminating insurance underwriting in the individual market in all states. Insurers must now issue plans to all consumers who apply during an annual open enrollment period, without regard to a consumer's preexisting conditions.[13] The law also limits the ability of insurers to set prices freely. Premiums can vary only according to four factors: (1) age, with at most a 3:1 ratio of premiums for the oldest to youngest enrollees; (2) geographic rating area; (3) family composition; and (4) tobacco use, limited to a 1.5:1 ratio. In particular, not only are insurers required to provide coverage to individuals with pre-existing conditions, they are further not allowed to charge such individuals higher premiums.

Insurers wishing to offer plans in the individual market must declare their interest in entering a particular geographic market and detail specific plan options and monthly premiums before the plan year. Those premiums will then be fixed over the plan year.[14] The plan offerings themselves also must fit into standardized bins based on actuarial value (AV). Those bins include: Bronze (60% AV), Silver (70% AV), Gold (80% AV), and Platinum (90% AV).[15]

[10] A subsequent Supreme Court ruling in 2012, however, allowed states to opt out of the expansion; as of November 2017, 33 states including the District of Columbia had chosen to expand Medicaid.

[11] The tax bill signed into law in December 2017 eliminated the tax penalty underlying the individual mandate as of January 2019, but left other features of the law intact.

[12] In 2014, the out-of-pocket maximum could not exceed $6,350 per individual and $12,700 per family. The essential benefits must include: ambulatory patient services, emergency services, hospitalization, maternity and newborn care, mental health and substance use care, prescription drugs, rehabilitative and habilitative services and devices, laboratory services, preventive care and chronic disease management, and pediatric services, including dental and vision care (*Patient Protection and Affordable Care Act 1302* 2010).

[13] Insurers must also allow individuals 60-day special enrollment periods for potential enrollees with a qualifying life event. Such life events include losing health coverage, moving, getting married, having a baby, or adopting a child. In our empirical work, we focus on enrollment decisions in the open enrollment period separately from those occurring after qualifying life events.

[14] Dickstein et al. (2015) describe how individual states define rating areas within their borders.

[15] Catastrophic plans with lower actuarial value could be offered to young adults under 30 (*Patient Protection and Affordable Care Act 1302* 2010).

On the demand side, consumers have the option of purchasing insurance plans either directly from an insurer or from a marketplace. Consumers purchasing in the marketplaces may be eligible for premium subsidies and cost-sharing subsidies, depending on their income and the specific plan they choose.[16] Premium subsidies, which depend on both the individual's income and the premium of a benchmark plan in the consumer's geographic region, are "advanceable" —in most cases, the government pays the subsidy directly to the consumer's chosen insurer. Because we observe premiums paid by bank or credit accounts in our data, our measure of premiums is net of any subsidies the federal government pays the insurer in advance. Consumers of income between 100% to 250% of the poverty line are also eligible for cost-sharing subsidies when they purchase a silver tier plan. These subsidies come in the form of lower deductibles, co-payments, and coinsurance.

2.2 Covered California

In our empirical work, we focus on California's individual insurance market, including its state-based marketplace, Covered California (CC). California imposes several additional regulations on insurers beyond those required under the ACA. In particular, when insurers choose to serve a geographic market in California, they must offer at least one plan of each metal type (Tebaldi, 2017). Furthermore, and in contrast to other state marketplaces, the plan offerings are uniform within a metal tier. That is, California dictates the exact co-payments, deductibles, and out-of-pocket maximums for all plans in a given metal tier. Only the premiums and the network of physicians and hospitals included in the plan's network may differ by insurer. We show these plan design features in Appendix Tables A1 and A2 for the 2014 and 2015 enrollment years.

3 Data

Our data come from a financial services company that provides services to large banks, including five of the top ten U.S. banks. As part of providing these services, the company collects transaction-level data from the banks' users, including all individual-level transactions on their bank accounts and linked credit cards. Notably, unlike other platforms such as Mint.com, which requires active user opt-in, our third-party data provider gathers transaction information directly from the financial institutions. We are thus less concerned about the selection of consumers into our dataset, other than through their choice of financial institution.

More specifically, our data identify users with a unique user code, which links together all of the user's bank accounts and credit card accounts within a given financial institution. Within this financial institution, we

[16] In detail, the premium subsidy design sets a cap on how much of an individual's income must be spent to enroll in the benchmark plan, for individuals with income between 100 and 400% of the federal poverty guidelines. In 2017, for example, an individual with income between 100 and 133% of the poverty line would be required to pay no more than 2.04% of her income. If the cost of the benchmark plan, the second cheapest silver plan, exceeds an individual's premium cap, the federal government pays a subsidy equal to the amount beyond the cap. Thus, the subsidy equals the difference between the income cap and the benchmark premium. The individual need not buy the benchmark plan, however; she can use the subsidy toward any plan sold in the marketplace.

see every line-item transaction with the date, the dollar amount, and a text description of the transaction. For many of the transactions, we have a separate variable with the merchant's name, and for transactions that occur at a physical location, we have geographic information on the location, including city, state, and sometimes zip code. Table A3 contains an example of credit card transaction data and some relevant variables we observe.

Our sample comprises 8.5 million "active" user households in the United States.[17] In this paper, we focus on the 850 thousand households in California. Comparing our sample size count to the number of households in California, our data represent a 6.7% sample of the California population.[18] Figure 1 shows the geographic distribution of our users across the counties in California. The counties in the Bay Area, such as Alameda, San Francisco, and Santa Clara counties, are overly represented, where we have over a 10% sample of the population. In contrast, in many rural counties our sample represents less than 3% of the total population. This geographic variation in coverage intensity relates to the set of banks that have partnered with our data provider.

3.1 Measuring Household Income

To create measures of monthly and annual household income, we start with all deposit transactions to users' bank accounts and remove non income-related transactions such as transfers between accounts and loan disbursements.[19] Figure 2a compares our 2014 income distribution to the 2014 California household income distribution as reported by the American Community Survey (ACS). Our data has a distribution slightly skewed to the left, likely because the income transactions we observe in our data are after-tax. We also have a larger mass of very low income users relative to the share reported by the ACS, likely reflecting that those users' income and consumption occurs more often in cash; cash income and spending lie outside our banking data. Overall, our data appear to be quite representative of the household income distribution. Comparing the geographic variation in median incomes across counties between our data and the ACS, Figure 2b shows a correlation of 0.75, with many counties falling close to the 45-degree line.

3.2 Measuring Insurance Premiums

Next, we measure health insurance premium payments. We compiled a list of all health insurance providers participating in Covered California (CC) from 2014 to 2015.[20] For each insurer, we manually searched for premium payments from both bank and credit card accounts based on certain keywords or phrases appearing

[17] The financial services company provided us a random sample of "active" users, where "active" is defined as having frequent transactions during 2014.

[18] While a user in our data could represent either an individual or a household, the income data discussed below matches the household income distribution much better than the individual income distribution, suggesting that our bank users should be viewed as households.

[19] The Data Appendix provides a complete list of the types of payments excluded from income. We also exclude households with (after-tax) incomes over $200,000 as they are unlikely to be impacted by the ACA.

[20] The companies are Anthem Blue Cross of California (ABC), Blue Shield of California (BS), Kaiser Permanente (KP), Health Net (HN), L.A. Care Health Plan, Molina Healthcare, Chinese Community Health Plan, Contra Costa Health Plan, Sharp Health Plan, Valley Health Plan, and Western Health Advantage.

in the description of the transactions. We were able to identify all of the participating insurers in the data.[21] Based on the dates of the first insurance payment, we identify those consumers who select an insurance plan during open enrollment.

Figure 2 plots the 2014 open enrollment market shares by insurer as measured according to the premium payments in our data against the market shares as reported by Covered California. For all insurers other than Kaiser, our market shares are within 1-4 percentage points of those reported by Covered California. Our Kaiser market share is likely low due to the difficulty of separating insurance premium payments from payments for health care services, as Kaiser is a vertically integrated insurer and health care provider.[22] Overall, our data appear quite representative of Covered California enrollees.

3.3 Measuring Health and Drug Spending

Our final key variables are measures of individual health care and pharmacy spending. To create these measures, we must identify those transactions from individual users' bank accounts and credit cards that are health care or pharmacy/drug related. We use the Laplacian Corrected Naive Bayes machine learning algorithm to classify each transaction as either 'health,' 'drug,' or 'other'. This requires first creating a training dataset for each category, in which we manually sort a sample of transactions into these three categories based on their descriptions.[23] For each category, we then calculate the frequency of each word appearing in these transactions. For example, for each word x, we could measure the probability that this word occurred within the health training sample as: $\mathbb{P}(x|health)$. From the three training sets, we can also measure the overall share of transactions that are health, as represented by $\mathbb{P}(health)$, and the overall frequency of the word x occurring across all categories, $\mathbb{P}(x)$. We use these inputs in the classifier below.

The Naive Bayer Classifier takes an unclassified transaction t, parses the description string into a set of individual words, $[x_1, x_2, ..., x_{S_t}]$, and measures the probability this set of words came from the distribution of words within each of the three categories of transactions. The transaction is classified into the category that has the highest probability of observing this set of words in a transaction. A key simplifying assumption of the Naive Bayes classifier is that it assumes that the set of words appearing in a given transaction are independent of one another. Thus, the classifier does not take into account the joint probabilities of words appearing together in the same transactions, but simply uses the products of each word's marginal probability. While this process ignores some information, it dramatically simplifies the calculation.[24]

[21] In Section A.2 of the Data Appendix, we provide a complete list of identified words and phrases used in classifying premium payments.

[22] While Kaiser insurance payment transactions often contain different descriptions from health care payments, they could be bundled with health-care payments and, in those cases, cannot be identified. See Section A.2 of the Data Appendix for further details.

[23] Our data provider created its own classification of spending categories. We use this classification as a baseline and then construct our final training set by manually removing transaction which do not appear to be health care related. Section A.3 of the Data Appendix provides further details on this and other issues related to constructing our machine-learning based classifier.

[24] We use the Laplacian Corrected version of the Naive Bayes classifier. This method assumes that all possible words appear at least once within the word probability distributions within each transaction category. Without this correction, if there were a word that never appears in our word bank for a given category, Bayes' rule would tell us there is zero probability that the

For example, the probability that transaction t came from the health category is measured as:

$$\mathbb{P}(health|t) = \mathbb{P}(health|[x_1, x_2, ..., x_{S_t}]) = \frac{\prod_{i=1}^{S_t} \mathbb{P}(x_i|health)}{\prod_{i=1}^{S_t} \mathbb{P}(x_i)} \mathbb{P}(health),$$

where $\mathbb{P}(x_i|health)$, $\mathbb{P}(x_i)$, and $\mathbb{P}(health)$ are all measured in the training datasets discussed above. Figure A1 shows word clouds of the top 100 words in each category, where font size indicates frequency. As these word clouds indicate, our empirical procedure appears to be quite successful at identifying relevant health and drug-related transactions. See Appendix A.3 for further discussion on creating the training datasets and our machine learning classifier.

To assess the accuracy of our methods for measuring health care consumption, we benchmark our estimates against those reported in the Medical Expenditure Panel Survey (MEPS). We compare average annual out-of-pocket health care and drug spending in our sample to those reported by MEPS households who live in the Western Region of the US in Table 1. [25] Our measure of annual health care spending is within 10% of the MEPS reported amount for all years 2012 through 2015. This difference is not statistically significant. Our estimates of drug spending, however, are substantially higher than MEPS. This is expected since our measure of drug spending includes all consumption at drug stores, and not just purchases of prescription drugs. Overall, our measure of health care consumption appears representative.

3.4 Identifying Households with Children

Similar to how we have classified a transaction as health or drug, we also predict if a household has children or not using the same Naive Bayes text classification method. However, instead of classifying transactions, we have combined all the transactions made by a single household in 2013 into one long text string and computed a score of the likelihood that this household has children based on the string. To create the training sample, we manually picked out transactions from merchants that sell child-oriented products and services. We then formed our child training sample using the top 10% of households with transactions at these merchants.

Our method classifies about 13% of all the households in our data as having young children. This value is fewer than the number reported by the Census Bureau because our definition of children mainly is driven by purchases indicative of young children. As shown in Figure A2, the most popular merchant names such as Gymboree and Carter's are retailers of baby clothing and toddler essentials.

3.5 Determining Medicaid Eligibility

In our analysis of drop-out in Section 4.4, we explore the interaction between attrition and a household's eligibility for Medi-Cal, California's Medicaid program. For this analysis, we need a measure of a household's Medi-Cal eligibility by month. Because we do not directly observe this outcome, we collect additional data

transaction came from this category, no matter what other words also appear in this transaction. For all words, we add one to the count frequency of each word to create our word banks. This prevents these zero probability events from dominating the classifier.

[25] MEPS does not provide state level geographic identifiers.

that allows us to impute, for each household and month, the *probability* that it is eligible for Medi-Cal. Appendix A.4 describes in detail how we create this measure; here, we provide a brief summary.

First, because Medi-Cal eligibility depends on pre-tax monthly income, we add taxes to our observed post-tax measure of income. To do so, we use the TAXSIM software developed by the National Bureau of Economic Research to determine the average income tax rates for federal, state, and payroll taxes that our households would face given their observed income. Tax rates vary by family size; however, our bank data does not provide us a direct count of family members. To account for unobserved family size, we compute alternative tax rates–and alternative levels of pre-tax income–for different hypothetical household sizes.

With our pre-tax income measure determined for a variety of potential family sizes for each household in our data set, we collect the monthly Federal Poverty Line values relevant to Medi-Cal eligibility.[26] These values differ by year and by family size. To compute a single eligibility probability for each household, we need additional data to allow us to weight by the probability that an account in our data represents each of the possible family sizes. To form these weights, we use statistics computed from the California Health Interview Survey for years 2013 through 2016. The survey provides data on the joint distribution of family size and income, specifically for purchasers of individual insurance.[27] Using this empirical distribution of family size by income group, we form a weighted sum of the eligibility determinations for each household and month.

3.6 Measuring the types of health care consumed

In our analysis of how dropouts and long-term enrollees differ in their health care consumption in Section 4.5, we also want to differentiate whether dropouts consume more discretionary or less urgent health care. These types of care may be more easily re-timed across months within a year. For this analysis, we need to define which health care transactions are more discretionary or urgent. We rely on an outside data source to measure these characteristics, for each of 1,522 different root words appearing in our health transactions.

More specifically, to perform this classification task in an objective way, we chose not to classify the words ourselves, but relied on the service of Amazon Mechanical Turk (MTurk) workers. MTurk is a marketplace that allows access to an on-demand scalable workforce, and thus allows us to perform thousands of classification tasks.

We designed a classification survey so that each health-related word is classified by 10 different workers. Each word is ranked on two scales ranging from 1 to 10: one for *essentialness* of the health care service; one for its *urgency*. We set the lowest level of urgency/essentialness equal to 1, while 10 represents an

[26] We define eligibility using the threshold of 138% of the Federal Poverty Line, applicable to all adults ages 19-64 under the ACA in California beginning in 2014. We do not account for the higher eligibility thresholds for infants, pregnant women, and the working disabled program, which could affect some members of the households we observe in the bank data.

[27] Using data from the California Health Interview Survey, we compute the family size distribution by income bin, among those families who purchase individual insurance plans in California. The distribution varies importantly across income categories. For example, for households with income less than $20,000, 57% contain only single adults, 14% contain two adults, and 29% represent families with between one and three children. At incomes between $40,000 and $70,000, only 27% of households contain a single adult, 38% contain two adults, and 35% of households have between one and three children.

extremely urgent/essential health care need. We also asked workers to tag whether a word was irrelevant for health care. Some words in the transaction descriptions of our health care transactions are generic words that do not tell us about the treatment, such as "California" or "general". We omit these words from our analysis.

To help workers in this classification task, we provided them with examples of what we consider urgent and essential. We defined urgent health care spending as a service a patient cannot postpone. A patient with a heart attack, for example, must go to the emergency room immediately or risk death. On the other hand, patients can postpone some visits, like an annual check-up. Similarly, some visits to the doctor are essential: a patient with a chronic condition like diabetes needs to receive regular treatment. Other health visits are much less essential; for example, visits for acne treatment. Our two measures of urgency and essentialness will turn out to be correlated, yet distinct.

To ensure that we focus our analysis on those transactions most likely to involve health care, we omit any root word for which at least 2 of the 10 MTurk workers flagged the word as irrelevant. This rule leaves us with 134 words. While applying this rule causes a large drop in our sample of words, when we examine our initial set of root words in the descriptions of health care transactions, many are generic and uninformative. Table 2 lists the top and bottom five words in the categories "essential" and "urgent," respectively, in terms of mean rank assigned by MTurk workers. Cancer, tumor, and cardiology rank in the top five for both essential and urgent. Indeed, highly urgent health care tends to also be highly essential. However, differences emerge in the lowest ranked words. Clinics and supplements appear at the bottom of the urgent ranking, but not at the bottom of the essential one, as patients may need these types of care but can easily postpone them. Relatedly, acupuncture and chiropractic rank at the bottom of the essential score, but these types of treatments can be somewhat urgent as they are often related to pain management.

3.7 Forming the analysis sample

The final dataset used for our analysis includes all households we observe purchasing health insurance, along with a 20% random sample of CA households who never purchase individual market health insurance.[28] Our 20% sample of the never-purchasers has 13.8 million transactions associated with 289,481 households in California. For 104,233 households, approximately 12% of our California sample, we observe at least one premium transaction in the individual health insurance market at some point during the years of our sample.

For the members we identify paying an individual insurance market premium, we define *sign-up* to be the month we observe the consumer's first premium payment and *drop-out* to be the month of her last premium payment. Figure 3a shows the number of members who have signed up for or dropped out of their health plans over time. We see a noticeable increase in take-up around December 2013, corresponding to the start of the ACA open enrollment. We define the 2014 open enrollment group as those members who signed up between December 2013 and March 2014 and the 2014 off-cycle enrollment group to be members who signed up between April 2014 and November 2014; we use similar definitions for the 2015 open and off-cycle

[28] We use a 20% random sample to ease the computational burden.

enrollment groups. Figure 3b displays the counts of members by enrollment period. In our data, we observe 15 thousand members who signed up during the 2014 open enrollment window.

Covered California reports that 2.89% of California's population enrolled in the state's individual insurance marketplace during the 2014 open enrollment period. This share is greater than the 1.8% (15K / 850K) of Californians we observe signing up for a new plan during the same period. This difference largely results from our definition of open enrollment sign-up. Covered California's share includes all consumers who purchase a marketplace plan; we include only those households who did not enroll in individual market health insurance prior to the ACA 2014 open enrollment period. Thus, the difference in our enrollment number and Covered California's suggests that 38% of 2014 ACA open enrollment enrollees had previously purchased health insurance on the individual market before the ACA.

We can also compare our 1.8% figure to data from the American Community Survey, which reported a net increase in the California population share that purchased individual market health insurance of 0.83 percentage points from 2013 to 2014. We expect this number to understate the number of new health insurance enrollees, because the 0.83 reflects the total count of new enrollees in 2014 less the count of those 2013 incumbent enrollees who exited the individual market before 2014.[29] Our analysis will focus on the behavior of the new enrollees.

Figure 3c displays the median after-tax household income by enrollment group. The annual 2013 household income of those who never purchase individual health insurance is $52,000. In contrast, households who purchased individual health insurance before the ACA have a 2013 median income of $70,000. Thus, consumers in the California individual health insurance market prior to the ACA appear to be largely high income households. This selection changes dramatically under the ACA, where the median household income of 2014 open enrollees is around $50,000.

4 Individual Behavior

In this section, we use our constructed dataset to study the individual behavior of ACA enrollees in terms of their health care spending and attrition rates.

4.1 Measuring Health Care Spending and Consumption

One of the unique features of our data is that we can measure out-of-pocket health care spending within households, regardless of the type of health insurance they have. That said, one potential problem with our data is that we only can observe out-of-pocket spending, but not the entire cost of one's health care utilization. This could be problematic when studying health care consumption of ACA enrollees for two reasons. First, it is possible these enrollees may have consumed charity care in the prior years, leading to zero out-of-pocket spending. This would not show up in our bank and credit card data. This may be especially important for households that were uninsured prior to ACA enrollment. Second, if households

[29] Enrollees in 2013 may exit the market in 2014 for a number of reasons, including obtaining employer-sponsored insurance, Medicaid coverage, or choosing to remain uninsured.

are paying their medical bills and don't have health insurance, out-of-pocket spending in periods prior to ACA enrollment may reflect the full cost of their health care consumption, while post ACA enrollment their out-of-pocket would only reflect their co-pay. Both of these issues could create problems when mapping out-of-pocket spending to health care consumption pre versus post ACA coverage.

We examine both issues in the Medical Expenditure Panel Survey. This survey collects information on health care and health insurance information on households over a two year period. Indeed, it contains data on whether individuals signed up for health insurance through an ACA exchange. It also contains data on health care consumption, regardless of who paid for the health care, including charity care. We can thus study these individuals reporting ACA coverage to learn about the potential biases in our data. A key downside of the MEPS data is that it is a very small sample: We only have 201 individuals, nationwide, reporting ACA coverage in 2014.

First, we examine the types of health insurance 2014 ACA enrollees had in 2013. Figure 3d shows that the average 2014 ACA enrollee had no health insurance for 5.5 months, employer sponsored health insurance for 3.75 months, 1 month individual market health insurance, and 1.75 months of Medicaid and other public insurance in 2013. Overall, the average household has some form of health insurance for more than half of the year, suggesting that pervasive consumption of charity care would likely be low and that out-of-pocket payments would also be combined with some health insurance claims to pay for health care.

Next, we look directly at the changes in total health care consumption among ACA enrollees pre versus post ACA enrollment in the MEPS data. In these data, health charges reflect the total "retail cost" of health care consumed and do not reflect negotiated prices with insurers or with the payer (or lack thereof) of the bill. Using the panel nature of the MEPS data, we analyze the within-individual change in health charges before versus after ACA coverage. Panel B of Table 3 shows that health charges increased by 28% among these individuals when covered by ACA insurance, however our small sample leads to large standard errors. An alternative measure for health care consumption is the count of interactions with the health care system. MEPS also reports this. Indeed, this is likely very similar data to the count of health care transactions we observe on households' bank and credit card accounts. Panel B of Table 3 shows that the ACA coverage is associated with as 21% increase in the count of health care interactions in MEPS, but standard errors are still quite large. Despite the noise in the estimates, it is comforting to see that the increase in health care consumption looks quite similar regardless of whether we use health charges or a count of health care interactions. To further validate this measure, we use our large sample of transaction data and estimate the change in the count of health care transactions among 2014 ACA enrollees. Panel A of Table 3 shows that our transactions data indicates a 24% increase in the number of health care transactions, but now our standard errors are quite small. Since these estimates are so close to what we see in MEPS, it suggests that we can interpret the change in health care transactions as observed in bank and credit card data as a reasonable proxy of the quantity of health care consumed.

As further validation of our data, we measure out-of-pocket spending on health and drugs in both MEPS and our data. Table 3 shows that changes around ACA enrollment in out-of-pocket health and drug spending, as well as in the count of drug transactions, all show consistent patterns between our data and the MEPS data. In addition, we validate our measure of health care transactions as a measure of health care consumption by

combining the MEPS data and our data to impute changes in health charges in our transactions data. To do this, we regress health charges on out-of-pocket spending in the MEPS data among the ACA enrollees. We run separate regressions in 2013 and 2014 to allow the relationship between health charges and health out-of-pocket to differ between these periods. Appendix Table A4 reports these regression coefficients. We then take the observed out-of-pocket spending in our transactions data and use these regression estimates to impute total health charges and changes in charges over time. Panel A of Table 3 shows that we find a 28% increase in total health charges, with tight standard errors. Taken together, this evidence indicates that our data provide an accurate picture of out-of-pocket spending and that the count of health care transactions provides an accurate measure of the quantity of health care consumed[30]

4.2 Total Health Care Consumption of ACA Enrollees

We begin our analysis by studying the extent to which the take-up of individual insurance during the 2014 open enrollment period affects total enrollee-level health care spending. We focus specifically on the 2014 open enrollment group first, since the decision to sign up in this period is tied most closely to the policy change in the marketplace. We will examine the effects of enrollment and drop-out on consumption among all types of ACA enrollees (2014 and 2015 open and off-cycle enrollment) in the next subsection.

We perform this analysis first using a simple differences analysis, comparing the health care spending of ACA enrollees post enrollment to their health care spending prior to enrollment. A potential concern, of course, is that this simple differences estimation only captures overall general population time trends in health spending. To address this concern, we additionally perform a differences-in-differences analysis, in which we difference out the pre-post change in health care spending by a "reference group" of households who have never had individual insurance. In this analysis, we do not argue for random assignment of households into our "treatment group" of ACA enrollees and the reference group. Indeed, a key aspect of our study will be to understand the reasons why certain households sign up for, and potentially leave, coverage in ACA regulated individual insurance markets. Rather, we use this reference group to ensure that our results are not driven solely by time trends in the overall population. It will turn out that this second difference makes essentially no difference in our estimates.

Our full differences-in-differences specification is given by:

$$H_{it} = \alpha_i + \delta\, open_{it} + \lambda\, post_{it} + \gamma\, open * post_{it} + \varepsilon_{it}, \tag{1}$$

where H_{it} is a measure of health care spending by household i in year t. In our analyses, this measure alternately represents the total dollars spent on health-related expenses and the dollars spent on drugs. We also use the total *number* of health-related transactions and the number of drug transactions as dependent variables. As previously discussed, we view these counts of transactions as measures of the quantity of health and drugs consumed, respectively. The variable $post_{it}$ is a dummy equal to one during the post-ACA enrollment period, i.e., for the years 2014 and 2015, and zero otherwise. The variable $open_{it}$ is a dummy equal to one if household i is in the 2014 open enrollment group, and zero if it is in the reference group. Finally, α_i denotes household fixed effects capturing time-invariant characteristics of a household that affect

[30] Missing charity care and changes in the generosity of insurance thus do not appear to be first order problems in our data.

health care spending, such as chronic health conditions. Our coefficient of interest is γ, which measures the impact of enrolling in an ACA marketplace plan on health care spending. We estimate our regressions separately by income, with households sorted into the following groups: $<$ \$20K, \$20K - \$40K, \$40K - \$60K, \$60K - \$100K, \$100K - \$200K of annual household income. These incomes are measured as 2013 annual income, prior to enrollment. In all regressions, we cluster the standard errors at the household level. Further, in this analysis, we do not control for whether or when a household drops their insurance; here, we simply compare how health care consumption changes around enrollment in a new ACA plan.

We report our results for general health spending and drug spending in Table 4. We have scaled our estimates of γ by the pre-ACA mean to present our estimates in terms of percentage changes. We find that the take-up of ACA insurance plans by lower income households leads to significant increases in their health care spending relative to the pre-ACA period. As shown in Panel A of Table 4, the estimates suggest that ACA enrollees with income of less than \$20K increase their number of health transactions by a highly significant 51%. While 51% sounds very large, this low-income group consumes very little health care overall. They only increase their health care consumption by 0.64 health transactions per year, and they only consumed an average of 1.24 health care transactions per year in the pre-period. The impact of ACA enrollment then declines monotonically with income. Those with income between \$20K and \$40K increase their number of health transactions by 27%, those between \$40K and \$60K by 19%, and those between \$60 and \$100K by 13%. The estimated effect is small and statistically indistinguishable from zero for households with greater than \$100K of income. These effects are very similar if we use a simple pre-post analysis, dropping the comparison to our reference group.

These moral hazard estimates of the ACA are in the range of previous studies of the moral hazard effects of health insurance. Probably the most comparable research is The Kaiser Family Foundation, 2018 that used the random assignment of access to Medicaid. They conclude the elasticity of health care consumption to Medicaid coverage is about 0.2, although this number varies depending on the type of health care consumption, some of which is measured through survey evidence. We build on their work by being able to show that the moral hazard effect declines with income, possibly reflecting that lower income households are more price sensitive.

As Panel B of Table 4 shows, the increase in the number of transactions was so great that, despite facing lower costs due to insurance coverage, ACA enrollees spend more in dollar terms on health care post-enrollment than they did pre-enrollment. A possible explanation for this finding is that individuals are able to afford big-ticket procedures that were prohibitively costly prior to receiving coverage. Looking at the full differences-in-differences specification, we find that the impact once again declines monotonically with income. Households with less than \$20K of income increase their total dollars spent on health care by 45.0% post-enrollment relative to the reference group. Those with income between \$20K and \$40K increase their dollars spent by 24.8% and those between \$40K and \$60K by 23.2%, relative to the reference group. The estimated effects are small and statistically indistinguishable from zero for the two highest income categories.

Our results remain consistent when we examine drug spending instead of general health spending. As we show in Panel C of Table 4, ACA plan enrollment increases the number of drug transactions for the lowest

income households. According to the differences-in-differences specification, households with less than $20K of income increase their number of drug transactions by 11.4% post-enrollment relative to the reference group. For households with income between $20K and $40K, the estimated increase is 5.7%. The effects are again statistically indistinguishable from zero for higher income categories. As shown in Panel D, we also find that households in the lowest income category increase their total dollars spent on drugs by 16.6% post-enrollment relative to the reference group, despite facing lower prices.[31]

Finally, we use our classification of households with young children to examine the effect of ACA enrollment on the health spending of families. We report the results in Table 5. Again, we find that take-up of insurance leads to an increased number of health transactions and more dollars spent for lower income households. Interestingly, the effect of ACA enrollment on total dollars spent on health care is even more pronounced for lower income households with children than for the average lower income household. ACA plan enrollment increases dollar spending by 96.8% relative to the reference group for households with income less than $20K and by 37.8% for households with income between $20K and $40K. This suggests that the spending burden of ACA plans is higher for families than for individuals. Even though families increased their number of health care transactions by the same amount as all enrollees, they had to spend substantially more out of pocket. This is consistent with deductibles being higher for family insurance plans, relative to plans covering individuals. For drug spending, our estimated effects for families with children are noisier, but are consistent with the patterns seen on health spending.

4.3 Attrition Rates of ACA Enrollees

Although we observe a considerable increase in health spending in the period after sign-up, the health gains from this spending, particularly for chronic illnesses, may depend on continuing treatment for multiple months or years. Insurers' profitability also depends on collecting premium payments from enrollees each month. Thus, we use our bank transactions data to examine the extent to which ACA enrollees maintain insurance coverage throughout the year. We define drop-out as the point at which a household permanently stops paying premiums to the insurer it selected during open enrollment.

Figure 4b shows the share enrolled in each month, by income category, for consumers who signed up in the 2014 open enrollment window. That is, the curves indicate, at each point in time, the fraction of the initial enrollees who continue to pay their insurance premiums. The figure illustrates a striking fact: ACA enrollees across all income levels frequently drop coverage, with the strongest effects coming from the lowest income households. Among 2014 open enrollment households with less than $20K of income, approximately 35 percent drop coverage by July 2014. The rates of drop-out among high income households were also high, around 30 percent. By the end of 2014, approximately 50 percent of lower income households stop paying their premiums. The drop-out rates of the highest income households are only slightly lower.

Figure 4a examines the distribution of drop-outs across total months of coverage. We see that 15% of all 2014 open enrollment consumers stop paying their premiums after only one month. The remaining households,

[31] We illustrate the levels of annual health and drug spending and transactions underlying these difference-in-difference estimates in Appendix Figures A4,A4, A5, and A6.

however, are quite evenly distributed across total months of coverage. That is, while there is an initial surge in drop-outs among 2014 enrollees, subsequently, the rate of drop-out is relatively constant. Finally, 53% of enrollees do not drop out and remain covered into the following year (at least 13 months).

The pattern of attrition we observe in 2014, the first year of the marketplaces, persists in later years. In fact, the rate of early exit appears to increase over time relative to the pre-ACA period. Figure 4c reports the six month drop-out rate among consumers who purchased insurance prior to the ACA, as well as those in the 2014 open enrollment, 2014 off-cycle enrollment, and 2015 open enrollment. We find that the six month drop-out rate in the pre-ACA period was 25%. It increased to 29% among the 2014 open enrollees, and then jumped to 40% among the 2014 off-cycle enrollees. 2015 open enrollees exhibit a six month drop-out rate of 41%. Given this increasing pattern in attrition rates, the contribution of drop-out to insurer profitability and market stability may grow in importance as well.[32]

4.4 Causes of Sign-up and Drop-out

Why do households choose to enroll and/or drop insurance coverage? We can use our unique dataset to test possible drivers of sign-up and attrition. The most natural candidates for observed changes in insurance are losing or obtaining employer-sponsored coverage or changes in eligibility for state Medicaid coverage, such as Medi-Cal in California. Both changes are likely to occur when the household experiences a shock to its income. Increases in income may signal employment changes or a move away from the federal poverty line, which determines eligibility for Medi-Cal's income-based coverage. Negative income shocks may signal a change in job status and moves closer to the poverty line. In addition, a decrease in income may make individual insurance premiums less affordable for consumers who purchased coverage during a period of relatively higher income.

4.4.1 Income Shocks

We first examine the relationship between changes in income and consumers' decisions to sign up and drop out in the year prior to the ACA.[33] In Figure 5, the solid line shows the monthly income dynamics of consumers prior to, during, and after insurance coverage.[34] To harmonize the event study timing across

[32]Our examination of drop-outs using bank transaction data provides a new measurement of point-in-time enrollment and health spending for marketplaces consumers. We can compare our findings to aggregate enrollment statistics, which reflect both entry and exit from marketplace coverage over the year. In 2015, for example, 11.7 million Americans signed up for insurance through the individual marketplaces in the open enrollment period (HHS, 2015). On June 30th of the plan year, only 9.9 million had paid premiums and had an active policy (CMS, 2015). This drop of 15.4% is smaller than our finding that 29% of the sample exit coverage in the same period. Our measurement is larger both because we ignore consumers who signed up after the open enrollment window–the federal statistic includes consumers who signed up both on-cycle and off-cycle–and because our bank data allows us to identify those delinquent consumers who do not pay in future months. We treat the latter as dropouts in the month in which they stop payment.

[33]We focus on sign-ups and drop-outs that occur prior to July 2013, as the expected entry of ACA marketplaces in 2014 may drive drop-out at the end of 2013. We also focus on drop-outs in the calendar months before November or December, as the ACA penalty design allows a short coverage gap exemption of up to two months.

[34]We restrict our analysis to households with less than $60,000 of annualized average post-tax income in the 10 months leading up to sign-up. We analyze this subset so that income changes are not dominated by the highest income earners. Further, we

households who enroll in health insurance for different amounts of time, we code the event time as zero during all months of health insurance coverage. Thus, negative event times indicate months prior to sign-up, and strictly positive event times indicate months since drop-out. In our pre-ACA sample, income jumps 40% in the months of health insurance coverage, relative to the month prior to enrollment. The increase is even larger, 50%, when compared to the period six months prior to enrollment. In addition, income falls 20% over the three months after drop-out. Thus, prior to the ACA, the decision to enroll and drop out of individual health insurance was strongly related to income, pointing to employment changes or affordability as drivers.

In contrast, the dashed line in Figure 5 plots the income dynamics around sign-up and drop-out of ACA enrollees.[35] ACA enrollees exhibit a substantially smaller (20%) jump in income at the month of sign-up, and only exhibit a 30% increase in income relative to the period six months prior to enrollment. Further, there is no significant drop in income after drop-out. In Appendix Table A5 we further examine these fluctuations within income categories. We find similar effects within in each income group. The income spikes are more than twice as strong in the pre-ACA period than in the post period, and we never find income declines around drop-out in the post-ACA period. Overall, the ACA thus appears to have significantly lessened the role of income and affordability in the decision to obtain and maintain health insurance coverage.

4.4.2 Medi-Cal Eligibility

In addition to changes in income, we look explicitly at the role of Medi-Cal eligibility in the observed pattern of attrition. Our goal in this analysis is twofold. First, we measure the extent to which Medi-Cal eligibility can explain the observed share of consumers exiting coverage. Second, exploiting an event study design around take-up and drop-out, we examine the role that eligibility may play in the observed changes in health spending.

We start by predicting each family's Medi-Cal eligibility, for each month of our analysis period. We describe how we create this measure in Section 3.5 and in Appendix A.4. Appendix Table A6 presents summary statistics for the estimated probability of Medi-Cal eligibility by income group, where income is defined using pre-ACA income in 2013. In our total sample, the average probability of Medi-Cal eligibility by month is 21.8%. This probability varies widely across income groups.[36] For the lowest income group, with at most $20,000 in income, the average eligibility probability in each month equals 53.9%. This share drops monotonically in successively larger income groups. For households with income between $60,000 and $100,000, the likelihood of eligibility in a month falls to 5.9%, for example, and 37.0% of households never qualify for Medi-Cal in any month.

choose to estimate this effect in levels rather than logs to allows months with zero income. We study all income groups in an income-group-specific analysis in the appendix.

[35] In this analysis, we pool all ACA enrollees across all enrollment periods. We also restrict to households with less than $60K of annualized average post-tax income in the 10 months leading up to sign-up.

[36] This variation arises for two reasons. First, at lower levels of 2013 annual income, the likelihood of falling below the Federal Poverty Line cutoffs in any month is higher. Second, the typical family size varies by income, meaning the relevant poverty line thresholds also vary. For example, in 2015, the monthly income threshold for eligibility below 138% of the Federal Poverty Line equaled $1354 for a single individual, and $2789 for a family of four.

Using the likelihood of eligibility, we conduct an event study to test how Medi-Cal eligibility changes around the period of sign-up and drop-out. We find only small changes in eligibility. In Figure 6a, we plot the probability of Medi-Cal eligibility for all households. The zero period in the figure reflects an average across all periods in which the household maintained individual insurance coverage. The likelihood of eligibility in the overall sample equals approximately 25% six to ten months before sign-up. Around the period of sign-up, the probability drops to 19% and then rises by about one percentage point after drop-out. Thus, the increase in Medi-Cal eligibility coinciding with drop-out (as we move along the x-axis from 0 to 1 in the graph) is about one percentage point in our overall sample, suggesting that churn into Medi-Cal may not be an important driver of the observed drop-out in our data. In reality, we would only expect churn into Medi-Cal among the lower income groups, however. Panels (b) through (f) of Figure 6 therefore present the same event study, separately for our five income groups. In the lowest income group (panel b), predicted Medi-Cal eligibility in fact falls from the drop-out month to the first month after drop-out; it then slightly increases, and falls again, within the subsequent ten months. While the exact change in Medi-Cal eligibility post drop-out thus depends on the time interval over which we measure it, an upper bound on the change in the probability of eligibility is about three percentage points.[37] In the second lowest income group (panel c), the change in predicted Medi-Cal eligibility around drop-out is approximately one percentage point. Thus, in the two income groups that are at risk for leaving their ACA plan because they gain Medi-Cal coverage, we find that the predicted changes in eligibility are small relative to the share that drop ACA coverage. Churn into Medi-Cal – while likely present – does not appear to account for an economically important share of the attrition from ACA plans.[38]

As a second method of analysis, we investigate whether the dynamics of enrollees' health care consumption around sign-up and drop-out exhibit different patterns among those who also experience larger changes in the probability of Medi-Cal eligibility. Since Medi-Cal does not require copays on health care consumption, it is possible that switching into/out of ACA coverage from/towards Medi-Cal would create the appearance of increased out-of-pocket spending during the period of ACA coverage. In this case, our measure of health care consumption would be contaminated by the fact that we do not observe transactions without a co-pay. To estimate the importance of this effect we run the following regression:

$$\begin{aligned} Y_{it} =& \alpha_i + \beta_1(PostSignUp_{it}) + \beta_2(PostDropOut_{it}) + \beta_3 * (PostSignUp_{it} * PrMediCal_{it}) \\ &+ \beta_4 * (PostDropOut_{it} * PrMediCal_{it}) + \beta_5(PrMediCal_{it}) + \varepsilon_{it}. \end{aligned} \tag{2}$$

The dependent variable is either health or drug consumption. t indexes months and i indexes enrollees. β_1 measures the change health/drug consumption after an enrollee signs up, relative to their health/drug consumption in the ten months prior. β_3 measures how different this change in health/drug consumption is among enrollees who also have a high probability of Medi-Cal eligibility (at the same time). Similarly, β_2 measures the change in health/drug consumption post drop-out and β_4 measures how different this change

[37] This is the change from zero to four along the x-axis.

[38] Another feature that is striking in the figure is that there is a substantial decrease *before* sign-up, especially in the lowest income group (panel b). This pattern reflects our definition of annual income that uses pre-ACA income in year 2013. Those consumers with less than $20,000 in income in 2013 who subsequently enroll in individual insurance in 2014 or later are likely those whose income increased over that time period. This selected cohort with increased income could then purchase ACA coverage; with lower incomes, these consumers would have been referred to Medi-Cal during the enrollment process.

is among households with a high probability of Medi-Cal eligibility. We also control for the direct effect of the probability of Medi-Cal eligibility on health/drug consumption. If a substantial amount of the observed changes in health care consumption around sign-up and drop-out of ACA coverage are due to transitions to and from Medi-Cal, we would expect β_3 to be positive and β_4 to be negative.

Table 6 reports these results for health care consumption by household, across income groups. We place households into income groups based on annualized measures of their monthly income over the 10 months prior to ACA sign-up. For all households with income below $60,000, we see economically and statistically insignificant estimates of β_3 and β_4, for both health spending (panel A) and health transactions (panel B). For the higher income groups, we do find economically and statistically significant effects. However, since these are such high income households during the 10 months prior to sign-up, only a tiny share of them ever experience more than a modest probability of enrolling in Medi-Cal. Table 7 reports similar results for drug spending and transactions. Here we do find significantly larger changes in drug spending and transactions around ACA enrollment among those with high Medi-Cal eligibility probabilities, suggesting that churn into and out-of Medi-Cal may be driving some of our drug findings. We therefore focus on health care spending and transactions in the analysis that follows.

4.4.3 Re-timing of Health Care Expenditures

To the extent that income shocks and Medi-Cal eligibility are *not* the sole drivers of drop-out behavior of ACA enrollees, we explore an alternative hypothesis: consumers sign up for insurance to help defray the costs of non-chronic, potentially discretionary, health care needs, and then drop coverage once they have satisfied these needs. This behavior may be planned and strategic; consumers may also simply discover they are healthy after an initial visit and then decide to exit coverage.

The regulatory structure of the ACA may incentivizes this type of strategic behavior. A key feature of the law is that insurance companies cannot discriminate based on pre-existing conditions, either by raising premiums or denying coverage. Thus, insurers cannot reject applicants in subsequent years who have strategically dropped coverage in the current year. Prior to 2018, there was a tax penalty in each year for not maintaining coverage; however, as we discuss in Section 7, given the premiums charged in the marketplace and the relatively small size of the penalty, straightforward calculations reveal that many households would be better off incurring the penalty and only signing up for coverage when needed.

If households follow this pattern of strategic behavior, then we would expect to observe spikes in health consumption around the time of sign-up and steep declines in health spending immediately following drop-out. To look for this behavior, we estimate the change in consumers' health care consumption from the ten months prior to sign-up, during coverage, and during the ten months after drop-out. We control for contemporaneous monthly income, as well as the average monthly income during the prior three months to ensure any health care consumption changes are not driven by income shocks.[39] Panel A in Table 8 reports the monthly changes in health spending around sign-up and drop-out among individual market health insurance enrollees prior to the existence of the ACA. Apart from in the lowest income group, we

[39] Removing these controls has little effect on the estimates.

find no statistically significant change in health spending during enrollment or after drop-out. In contrast, Panel B shows that enrollees signing up an dropping out of coverage after the implementation of the ACA exhibit large increases in health spending when covered by insurance. Consumers with annual income less than 20K, for example, increased their health spending by 28% when covered and then cut their spending by 40% after dropping out. Those with annual income in the 20K-40K range increased spending by 54% when covered, and cut spending by 51% when dropping out. The 40K-60K income group increased spending by 11% when covered and cut spending by 5% after dropping out. The 60K-100K group increased their spending by 19% when covered and cut spending by 28% after dropping out. For the highest income group, the point estimates are smaller and cannot be statistically distinguished from zero. Repeating this analysis in on the count of health transactions shows that even during the period prior to the ACA, enrollees increase their health are consumption during coverage and lower it when dropping coverage. However, the magnitude of these effects are much larger among ACA enrollees, suggesting the ACA market made it much easier to strategically consume health care when covered by insurance and then subsequently drop coverage.

The estimates are noisier when studying prescription drug usage in Table 9, where we often can reject that spending dynamics are similar between dropouts pre- and post ACA. Enrollees in the pre-ACA period increase their drug spending and transactions when covered by insurance, but have smaller to zero declines when they drop coverage. Those covered during the ACA exhibit spikes in drug spending and transactions during coverage, which then reverts after drop-out. Further, as previously discussed, our measures of drug spending may be more contaminated by churn from Medi-Cal, making these estimates harder to interpret. Overall, the evidence on health spending and transactions is highly consistent with the hypothesis that households strategically time their entry into the ACA marketplace to cover health care needs, and then drop coverage.

4.5 How Dropouts use the ACA

To provide further support for our hypothesis that some fraction of the observed attrition reflects strategically-timed spending on health care, we use our MTurk classification of the essentialness and urgency of each health care transaction to investigate what types of health care enrollees consume before and after enrollment. Recall from Section 3.6 that urgency refers to the extent that a certain procedure can or cannot be postponed, with higher scores reflecting procedures that require greater immediacy. As an example, expenses related to a heart attack are extremely urgent, while regular check-ups can be postponed easily. Essentialness, on the other hand, refers to how important a procedure is to an individual's overall health and well-being. Diabetes and cancer treatments, for example, are highly essential, while minor skin treatments may be significantly less essential.

Given our hypothesis of strategic timing, we would expect to find that those individuals who end premium payments prior to the end of year use their coverage to defray the costs of discretionary, and therefore less urgent and likely less essential, health care needs. Moreover, we also would expect individuals with higher baseline levels of essential health care needs, i.e. individuals with more serious health conditions, to be less

likely to drop out. To test these hypotheses, we run the following specification:

$$share_{it} = \alpha_i + \delta\ dropout_i + \lambda\ post_t + \gamma\ dropout_i \times post_{it} + \varepsilon_{it},$$

where $share_{it}$ refers to either the percentage share of health spending that is classified as very essential or the percentage share that is classified as very urgent, respectively. We define very essential and very urgent health care as those expenses with text descriptions including words that score in the 95th percentile of the MTurk score distributions across words (for urgency and essentialness, respectively). The variable $dropout_i$ is a binary variable that takes the value of one if individual i signs up for individual coverage and then drops out within 8 months. We set $dropout_i = 0$ otherwise. The variable $post_t$ is a binary variable that takes the value of one in year 2014 and zero in year 2013. We again cluster the standard errors at the individual level. In this analysis we focus on the 2014 open enrollees, but separate them into those who drop out within 8 months and compare them to those who do not drop coverage.

Table 10 presents our results. The final row of regression estimates in the table illustrates the difference in the types of spending completed by those who drop coverage in the post-ACA period relative to 2013. We see that individuals who drop coverage do indeed devote a smaller share of their total health spending to very essential and very urgent health care. Specifically, the share of very essential health care drops by 13.86% during the coverage period, relative to the baseline mean. The share of very urgent health care drops by 29.77%. These results are again consistent with the hypothesis that a subset of individuals sign up for ACA insurance to cover discretionary health care needs, and then drop coverage once they satisfy those needs.

Table 10 also reports the full difference-in-differences specification, comparing the spending behavior of those who drop coverage to those who do not. The point estimates indicate that individuals who eventually drop out spend less on essential and urgent health care while covered, relative to individuals who maintain coverage for the entire plan year. Our diff-in-diff estimate for urgent health care is significant at the 5% level, while our estimate for essential health care is significant at the 10% level. Finally, the first row of Table 10 shows that those individuals who drop out are estimated to have lower levels of essential health care needs in the pre-period, before signing up for an ACA plan. This finding is consistent with the hypothesis that those who maintain coverage throughout the year have more serious health care needs. With a significance of 10.1%, the estimate just barely misses the 10% significance level. In sum, these results indicate that enrollees who drop out mid-year use their coverage to increase spending on less urgent and more discretionary health care, while those who maintain coverage have more serious and essential health care needs.

5 Model

We develop a conceptual framework to illustrate how the prevalence of drop-out affects premiums and plan availability in a competitive insurance market. In our setting, all households that are potential enrollees share the same expected spending over the course of a year. That is, we rule out the traditional channel for adverse selection in which enrollees of higher cost have higher willingness to pay for insurance. Instead, potential enrollees in our setting differ in their ability to re-time health consumption towards the beginning of the plan year, but all have the same annual health care cost.

In reality, there may be a correlation between the underlying health care cost and the willingness to drop out. If those willing to drop-out were especially low cost enrollees, then it is possible that their monthly health care consumption could be lower than that of long-term enrollees even if the drop-outs re-time their health care consumption to a short time period. Since our model will rule this out, we first empirically investigate the monthly number of health care transactions during the months of ACA coverage, separately for enrollees of different durations. We plot this in Figure 4c. We normalize the monthly health transactions to 1 for enrollees who stay enrolled for 13 months or more and then compare the percentage difference in the average number of monthly health transactions, across enrollees of different durations. Across all income groups, we find that enrollees who only stay covered for 1-4 months consume between 10 and 30 percent more health care per month when covered than those enrolled for 13 months or more. This number falls with duration of coverage, where those enrolled for 9 to 12 months actually consume between 10 and 40 percent *less* health care per month than the long-term enrollees.

These patterns are suggestive of a number of findings. First, there may be a limited amount of health care that enrollees can re-time. If enrollees can squeeze all re-timed care into a short window of coverage, then their monthly health care consumption spikes and becomes higher than that of the long-term enrollees. However, if this re-timed health care is spread over a larger number of months, then the monthly health care costs are lower. Indeed the long-term enrollees in-fact appear sicker those who drop out after 9 to 12 months, as the latter consume less health care than the long-term enrollees. However, the short-term enrollees can concentrate their spending such that they appear more costly than the long-term enrollees. The fact that short-term drop-outs are more costly (due to re-timing) than long-term enrollees will be important for out model.

5.1 Model Set-Up

Our framework is based on Klemperer (1987)'s model of duopoly with differentiated products and switching costs. We model product differentiation via a Hotelling spatial location model. There are two insurers, labeled A and B, located at 0 and t, respectively, along the Hotelling line. This differentiation, for instance, may reflect differences in the provider network. We model household preferences as a location $x \in (0, t)$. The cost of purchasing insurance from A is x and the cost of purchasing insurance from B is $(t - x)$. We further assume that reservation prices are sufficiently high such that a household always wants to purchase insurance from one of the two insurers. The cost of insuring any household is equal to c. By assuming costs are homogeneous across all households, we rule out the traditional adverse selection channel.

The model has two periods. In the first period, households are arrayed along the interval $(0, t)$ with uniform unit density. Thus, the total mass of households is equal to t. A household can then be one of two types. The first, which we denote as "non-dropout" (n) types, are unable to re-time their annual health expenditures and therefore pay the full annual premium. On the other hand, "dropout" (d) types are able to re-time all of their annual spending within the first $\phi \times 12$ months of the year and thus pay only a fraction ϕ of the annual premium. A household's type is private information. Let $\tilde{\nu}_1(x)$ denote the true probability that a household at location x is of the dropout type. Insurers do not know this true probability but instead believe that $E[\tilde{\nu}_1(x)] = \mu_1$ for all x. That is, their prior belief is that being a dropout type is independent

of location preference.[40]

Turning to the second period, suppose that a total fraction ν of the period 1 households have dropped out. Suppose further that a fraction σ^A of the households who do not drop out were enrolled with insurer A in period 1 and a fraction $\sigma^B = 1-\sigma^A$ were enrolled with insurer B. Then insurer A's existing customer base is equal to $(1-\nu)\sigma^A t$ and similarly for insurer B. These existing households face switching costs when moving from their current insurer to a competitor. Finally, the νt mass of households that dropped out in period 1 is replaced by new households. These new households' preferences are again uniformly distributed along the interval $(0,t)$. They again can be either a dropout type or a non-dropout type. We let $\tilde{\nu}_2(x)$ denote the period 2 probability that a household at location x is of the dropout type. For simplicity, we assume $\tilde{\nu}_2(x)$ is independent of $\tilde{\nu}_1(x)$ and that insurers believe $E[\tilde{\nu}_2(x)] = \mu_2$ for all x.[41]

Insurers seek to maximize profits and engage in price competition, setting prices p_1^A and p_1^B in period 1 and prices p_2^A and p_2^B in period 2. We solve for the subgame perfect Nash equilibrium of the game.

In Appendix B we develop all the results of the model, but we will summarize the key findings. First, drop-out types d will be less price sensitive when selecting an insurer than long-term enrollees at time of initial sign-up. Since long-term enrollees will pay 12 monthly premiums, but drop-outs only pay 1, the premium payment is a much more substantial cost for long-term enrollees than drop-outs. Second, since non-dropouts face switching costs in the second period, an insurer that captures a large share of long-term enrollees in period one will have a large, captive market in period two.

5.2 Empirical Predictions

We can use our model to develop a set of empirical predictions regarding the relationship between dropout and prices.

Proposition 5.1. *The insurer with the higher price increase between the first and second periods will have a higher dropout rate among its new enrollees.*

Proof. See Appendix B.2 □

This result is quite intuitive and illustrates the new form of adverse selection that we highlight. As an insurer raises the prices, increasingly the only customers who remain are those who are able to re-time their health expenditure and thus do not intend to pay the full year's worth of insurance premiums.

We also have an arguably more surprising result on the relationship between dropout rates in period 1 and the growth rate of prices between periods 1 and 2.

Proposition 5.2. *The insurer that experiences the lower (higher) dropout rate in period 1 will increase prices more (less) than its competitor.*

[40] These technical assumptions simply allow for the possibility that the two insurers will experience differential dropout rates in period 1.

[41] These assumptions simply mean there is no learning by insurers between periods 1 and 2 and that insurers once again believe there is no correlation between being a dropout type and location. These assumptions could be relaxed, but doing so would add significant complexity to the model without contributing to the economic message we wish to convey.

Proof. See Appendix B.2 □

Intuitively, a lower dropout rate in period 1 means an insurer enters into period 2 with a larger customer base. Due to switching costs, the insurer is able to monopolize these existing enrollees. The larger this existing customer base, the less aggressively the insurer looks to compete for new customers and the higher its period 2 price.

6 Empirical Evidence on Prices

We now test the predictions from our model about the relationship between drop-out rates and insurer pricing. We examine both the effects of insurer pricing on equilibrium drop-out as well as how drop-out rates affect the dynamics of the insurers' premium choices in future periods.

6.1 Empirical strategy

We use Covered California data to measure insurers' prices. While premium prices vary by income level and risk pool, this variation depends on formulas that adjust a standard price the insurer sets for each plan. Thus, we can summarize insurers' price by the unsubsidized premium set for a given plan, for a given age group and family size. We focus on prices of silver plans since they are (by far) the most popular metal tier purchased. Further, in California, all silver plans, by state regulation, have the same coverage characteristics, apart from the coverage network.[42]

To examine the effect of drop-out rates on prices, we look at the association between observed drop-out rates in year t and price changes from t to $t+1$. Since 2014 was the first year of the marketplace, it seems unlikely that insurers could forecast their drop-out rates when setting prices in the first year. After observing their drop-out rates in 2014, insurers may adjust prices for 2015. Similarly, 2015 drop-outs may influence price changes in 2016. Thus, we estimate:

$$\Delta \ln p_{s,t+1,c} = \alpha\, cheap1_{s,t,c} + \beta\, cheap2_{s,t,c} + \theta dropout_rate_{s,t,c} + \delta_t + \delta_s + \delta c + \varepsilon_{s,t,c} \tag{3}$$

where $\Delta \ln p_{s,t+1,c}$ is the average log change in the silver tier plan price from year t to year $t+1$ of company c in rating region s.

Our main object of interest is how $dropout_rate_{s,t,c}$, the total drop-out rate among enrollees of insurance company c in rating region s during year t, affects the log price change from t to $t+1$. In our specification, we also control for $cheap1_{s,t,c}$ and $cheap2_{s,t,c}$, which are dummies set equal to one if the company offers the cheapest or second cheapest plan, respectively, among silver tier plans in rating region s and year t. These indicators serve as important controls, since being the cheapest or second cheapest plan in the area directly

[42] Typically, each insurer offers one silver plan in each pricing region of the ACA marketplaces. Occasionally insurers may offer separate HMO, EPO, and PPO silver plans. In these instances, we average the prices of each plan within a given insurer, region, metal tier, and year.

influences the level of the government's premium subsidies in the area. Finally, we include δ_t, δ_s, and δc, which are year, region, and insurer fixed effects, respectively.

Column 1 of Table 11 shows that a 1 percentage point increase in last year's drop-out rate leads to 0.14 percent *lower* prices next year. Column 2 of Table 11 shows that this result is robust to adding controls for the quantity of health and drug spending of enrollees, as well as the share of enrollees who signed up off-cycle. Column 3 of Table 11 adds region-year and region-insurer fixed effects. These additional controls have little effect on the estimated relationship between lagged dropout rates and price changes.

Consistent with the model's prediction the results suggest dynamics related to consumers' inertia: If a plan experienced a low drop-out rate last year, it will have a larger number of incumbent enrollees in the subsequent year and can raise prices. In contrast, a plan with a high drop-out rate will have few incumbent enrollees in the following year, and must compete for new enrollees who are likely to be more price sensitive than the inertial incumbents.

Our findings emphasize the very undesirable effects of drop-out in the market: despite drop-outs raising average costs, the enrollees that end up paying the highest premiums are those who follow the rules of the insurance markets and do not drop out. These consumers may be the more chronically sick enrollees, the most risk averse enrollees, or both.

Next, we turn to analyzing how pricing affects the drop-out rate of new enrollees. Here, our model predicts that since drop-out types will not pay for a full year of insurance, they are less price sensitive than those who plan to enroll for the full year. Thus, insurers who set high prices will disproportionately discourage new full year enrollees from choosing their plans.

To test this hypothesis in the data, we examine a difference-in-differences regression of insurer prices on contemporaneous drop-out. Specifically, we estimate:

$$\begin{aligned} ln\ (dropout_rate_{s,t,c}) \ &= \alpha\ cheap1_{s,t,c} + \ \beta\ cheap2_{s,t,c} + \\ &\quad \theta\ ln\ (price_{s,t,c}) + \ \delta_{t,s} + \ \delta_{s,c} + \varepsilon_{s,t,c}. \end{aligned} \tag{4}$$

By including region-year and insurer-region fixed effects, our estimated effect of prices on drop-out is identified by comparing changes in plan prices over time with within-plan changes in drop-out rates over time.

Column 4 of Table 11 shows that a one percent increase in insurer prices leads to a 1.8 percent increase in drop-out rates, consistent with our model's predictions. This result highlights why, despite high drop-out rates leading to higher insurer average costs, insurers cannot recoup these costs through higher prices. By increasing prices, the drop-out rate only worsens. We observe the classic adverse selection unraveling, in which higher prices push out low-cost enrollees, leading to higher average costs. However, in our setting, selection is driven by heterogeneity in the willingness to drop out mid-year, rather than by heterogeneity in unobserved underlying health status. It appears that the follow-the-rules type helps prevent market unraveling. Insurers can recoup losses from dropouts through higher mark-ups on the inertial follow-the-rules types, after their first year of enrollment. This suggests dynamics where insurers would want to price low in the first year of the market to attract follow-the-rules types who are not yet inertial, but then raise prices in future years.

7 Mandated Insurance and the Tax Penalty

When the ACA exchanges were implemented, the only incentive to not drop out early was the tax penalty households would pay if they were not covered by some kind of health insurance for at least 9 months out of the year.[43] However, this tax penalty did not differentiate between households who remained uninsured because they never purchased health insurance through ACA exchanges versus those who purchased ACA insurance for a portion of the year, but then dropped coverage for the remainder of the year. This indicates that the functioning of the market could be improved by allowing the tax penalty for never enrolling in health insurance to differ from the penalty for dropping coverage. Indeed, these two different scenarios leading to paying the tax penalty have very different cost impacts.

First, consider those who choose to never purchase ACA insurance and pay the tax penalty. Since this group prefers to pay the tax penalty over signing up for any ACA coverage, if there were a tax penalty for dropping out mid-year that was higher than the current non-enrollment penalty, then this group would not be impacted. In contrast, those who choose to sign-up and then drop-out after only a few months would now face substantially higher costs. Indeed, since these short-term enrollees are the costliest group to insure, this drop-out penalty would either push these enrollees to enroll for a longer period or to never sign-up in the first place. Under either one of these scenarios, the average cost of enrollees per month would fall, making ACA insurance more affordable (for all enrollees). The optimal penalty levels for never enrolling in health insurance versus the penalty for dropping coverage mid-year are beyond the scope of this paper. However, this example suggests that a higher penalty for dropping out than for never enrolling could help improve the efficiency of these markets.

8 Conclusion

A crucial component of social insurance design is to ensure the programs *de facto* "reach" the intended beneficiaries. Empirical evidence suggests that the coverage dimension is highly relevant, with a sizable literature documenting low take-up of social insurance. While both the academic literature and policy responses have focused on take-up, we contribute a new analysis of attrition, or drop-out, from social insurance. This is especially important in the context of health insurance where the cost of providing insurance depends on the quantity of health care enrollees consume. If only the sick sign-up, the cost of the insurance per enrollee could be large.

We document a sharp increase in new enrollees in the first year of the ACA marketplaces' open enrollment period. We find that these new insurance enrollees are often of lower income and receive insurance premium subsidies. Moreover, following enrollment, we observe dramatic spikes in their health consumption, with more transactions and greater out-of-pocket health care and drug spending. These enrollees rarely pay consistent monthly premiums, however. Instead, attrition is widespread among new enrollees, even among the poorest newly enrolling households that are receiving government premium subsidies. In the 2014 and 2015 open enrollment years, roughly half of all enrollees in California drop out before the end of the plan

[43] As part of the 2018 tax changes, the tax penalty for not purchasing health insurance was completely removed.

year, with the sharpest drop-out observed after only one month of payment. Our data offers suggestive evidence that this attrition is chiefly driven by strategic re-timing of discretionary health care spending – a form of moral hazard – rather than by a lack of affordability. Finally, dropout rates appear to rising over time across each enrollment period of the ACA, suggesting the consequences of dropout may become more severe in the future.

Such attrition can have fundamental effects on market stability in the individual market. Our model illustrates that drop-out behavior generates a distinct adverse selection or moral hazard problem and can affect premium setting, even absent differences in enrollees' underlying health costs. Thus, measures that limit the extent of attrition could have a substantial effect on market stability.

Attrition also entails costs that are borne not exclusively by the households that *de facto* discontinue coverage before the end of the plan year, but also by the households that "play by the rules." Indeed, we show that non-drop-out enrollees, who are more inelastic than drop-out enrollees once inertial in plan choices, bear the brunt of the premium costs associated with attrition. Therefore, on top of strengthening the stability of the individual market, measures that contain attrition would protect households that follow the rules against the premium increases that result from the strategic behavior of households that do not.

References

Antwi, Yaa Akosa, Asako S. Moriya, and Kosali Simon (2015). "Access to health insurance and the use of inpatient medical care: Evidence from the Affordable Care Act young adult mandate". *Journal of Health Economics* 39, pp. 171–187.

Bhargava, Saurabh and Dayanand Manoli (2015). "Psychological Frictions and the Incomplete Take-Up of Social Benefits: Evidence from an IRS Field Experiment". *American Economic Review* 105 (11).

Cabral, Marika (2017). "Claim Timing and Ex Post Adverse Selection". *The Review of Economic Studies* 84 (1), pp. 1–44.

CMS, HHS (2015). *June 30, 2015 Effectuated Enrollment Snapshot*. URL: https://www.cms.gov/Newsroom/MediaReleaseDatabase/Fact-sheets/2015-Fact-sheets-items/2015-09-08.html.

Currie, Janet (2006). "The Take-up of Social Benefits". *Poverty, the Distribution of Income, and Public Policy*. Ed. by David Card Alan Auerbach and John Quigly. New York: Russel sage. Chap. 3.

Dafny, Leemore, Jonathan Gruber, and Cristopher Ody (2015). "More Insureres Lower Premiums: Evidence from Initial Pricing in the Health Insurance Marketplaces". *American Journal of Health Economics* 1 (1), pp. 53–81.

Dickstein, Michael J. et al. (2015). "The Impact of Market Size and Composition on Health Insurance Premiums: Evidence from the First Year of the Affordable Care Act". *American Economic Review* 105 (5), pp. 120–25.

Einav, Liran, Amy Finkelstein, and Mark Cullen (2010). "Estimating Welfare in Insurance Markets using Variation in Prices". *Quarterly Journal of Economics* 125 (3), pp. 877–921.

Ericson, Keith M Marzilli and Amanda Starc (2016). "How Product Standardization Affects Choice: Evidence from the Massachusetts Health Insurance Exchange". *Journal of Health Economics* 50.

Ericson, Keith Marzilli (2014). "Consumer Inertia and Firm Pricing in the Medicare Part D Prescription Drug Insurance Exchange". *American Economic Journal: Economic Policy* 6 (1), pp. 38–64.

Fang, Hanming, Michael P. Keane, and Dan Silverman (2008). "Sources of advantageous selection: Evidence from the medigap insurance market". English (US). *Journal of Political Economy* 116 (2), pp. 303–350. ISSN: 0022-3808. DOI: 10.1086/587623.

Finkelstein, Amy, Nathaniel Hendren, and Mark Shepard (2017). "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts". *NBER Working Paper No. 23668*.

Finkelstein, Amy and Kathleen McGarry (2006). "Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market". *American Economic Review* 96 (4), pp. 938–958. DOI: 10.1257/aer.96.4.938.

Handel, Benjamin R. (2013). "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts". *The American Economic Review* 103 (7), pp. 2643–2682. ISSN: 00028282.

HHS, aspe (2015). *Health Insurance Marketplaces 2015 Open Enrollment Period: March Enrollment Report*. URL: https://aspe.hhs.gov/system/files/pdf/83656/ib_2015mar_enrollment.pdf.

IRS Government, Report (2017a). *Individual Shared Responsibility Provision – Reporting and Calculating the Payment*. URL: https://www.irs.gov/affordable-care-act/individuals-and-families/aca-individual-shared-responsibility-provision-calculating-the-payment#Determining%20if%20You%20Need%20to%20Make%20a%20Payment.

IRS Government, Report (2017b). *Letter from IRS Commissioner John Koskinen to Congress*. URL: https://www.irs.gov/pub/newsroom/commissionerletteracafilingseason.pdf.

Klemperer, Paul (1987). "The competitiveness of markets with switching costs". *The RAND Journal of Economics*, pp. 138–150.

Kleven, Henrik and Wojciech Kopczuk (2011). "Transfer Program Complexity and the Take Up of Social Benefit". *American Economic Journal: Economic Policy* 3 (1), pp. 55–65.

Kowalski, Amanda E. (2014). "The Early Impact of the Affordable Care Act State-By-State". *Brookings Papers on Economic Activity*, pp. 277–333.

Patient Protection and Affordable Care Act 1302 (2010). Section 1302.

Patient Protection and Affordable Care Act 156 (2010). Title 45, Section 156.155.

Shepard, Mark (2016). "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange". Working Paper Series (22600). DOI: 10.3386/w22600.

Simon, Kosali, Aparna Soni, and John Cawley (2017). "The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions". *Journal of Policy Analysis and Management* 36 (2), pp. 390–417.

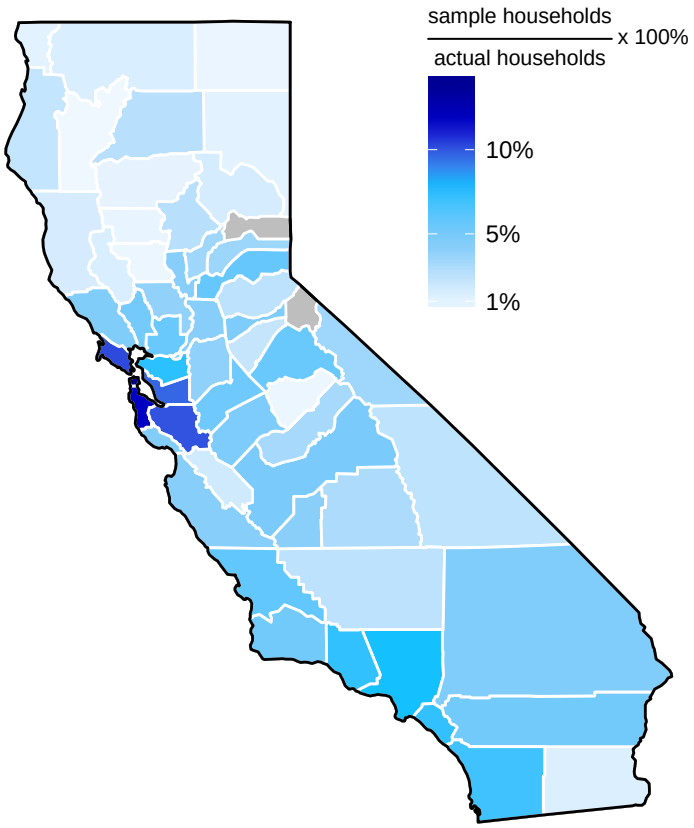
Tebaldi, Pietro (2017). "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA". Becker Friedman Institute for Research in Economics Working paper No 2017-05.

The Kaiser Family Foundation, State Health Facts (2012). *Health Insurance Market Reforms*.

– (2016). *Data Source: Census Bureau's March Current Population Survey (CPS: Annual Social and Economic Supplements), 2014-2017. Health Insurance Coverage of the Total Population*.

– (2018). *State Decisions on Health Insurance Marketplaces and the Medicaid Expansion*.

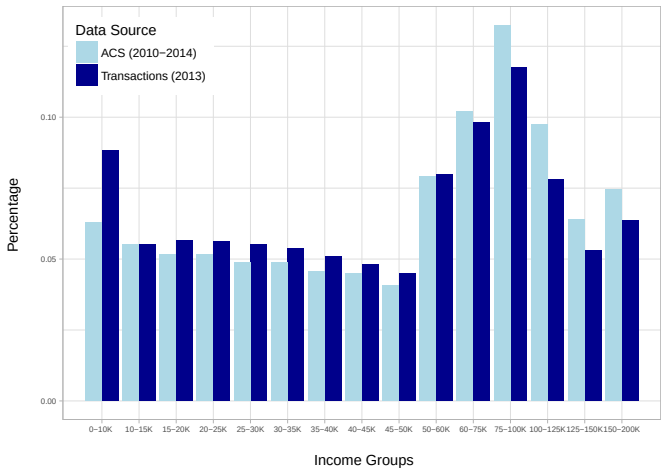
Figure 1: Geographic Distribution of Financial Account Members in California



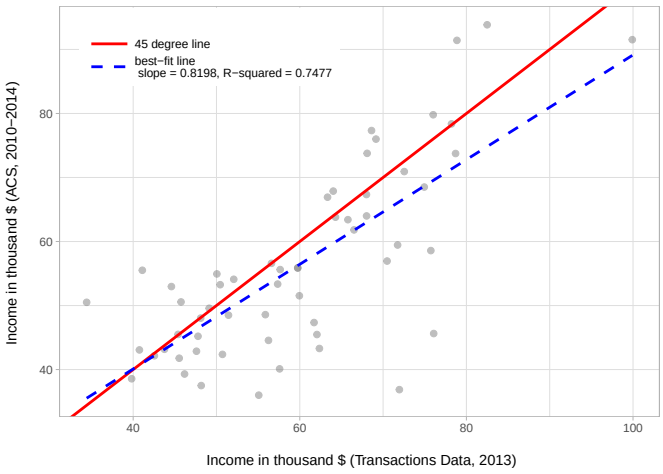
Notes: The map illustrates the geographic distribution of the financial account members in our data relative to the actual distribution of households in California. Our data contains approximately 850,000 active users, which corresponds to 6.7% of the California population. We use data available from the State of California's Department of Finance as our measure of the full distribution of households in the state. As shown in the figure, the counties in the Bay Area, such as Alameda, San Francisco, and Santa Clara counties, are over-represented in our sample: we have more than a 10% sample of the population in the Bay Area. In contrast, in many rural counties, our sample represents less than 3% of the total population.

Figure 2: Data Representativeness

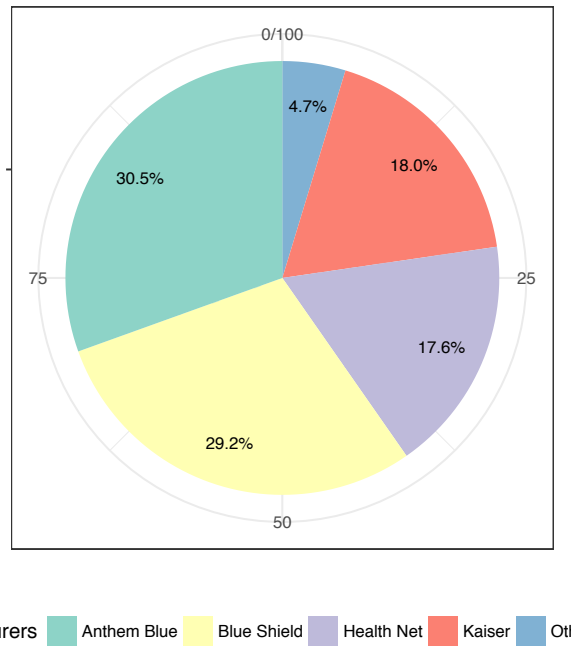
(a) Individual level Income



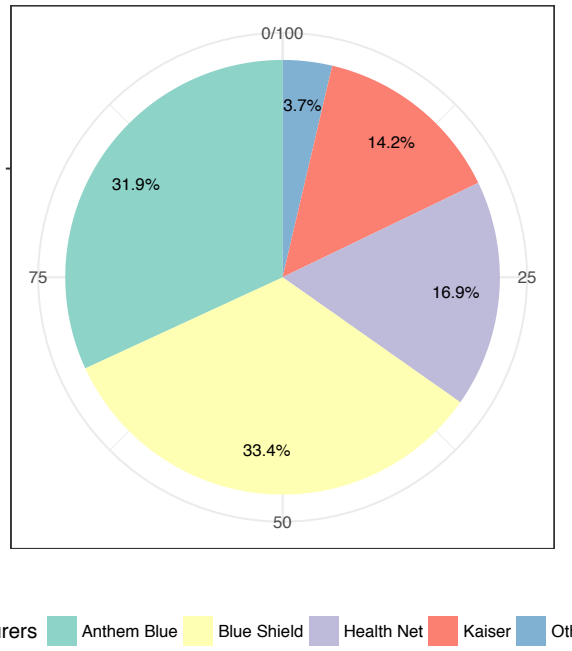
(b) County level Median Income

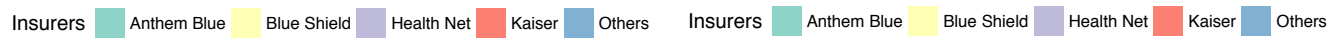


(c) Market Shares (Covered California data)



(d) Market shares (our transactions data)

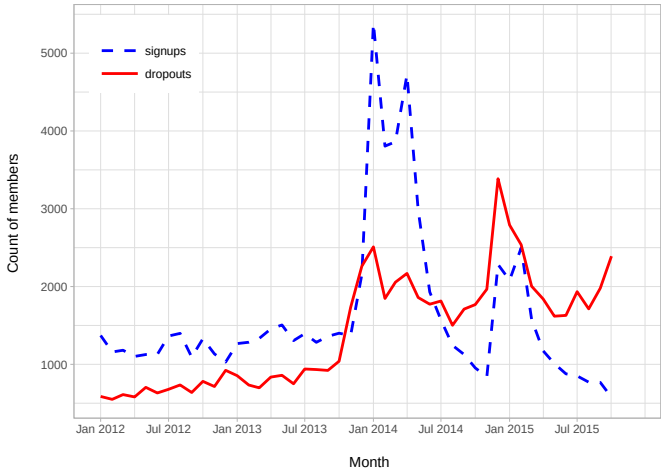




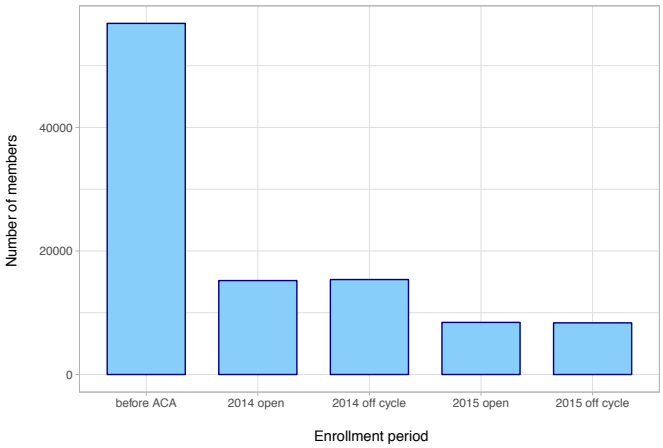
Notes: Panel (a) compares the 2014 income distribution of financial account members to the 2014 California household income distribution as reported by the American Community Survey (ACS). As shown in the graph, consumers in our data have incomes with a distribution slightly skewed to the left, with a larger mass of very low income users. Panel (b) compares the geographic variation in median incomes across counties between our data and the ACS. The scatter plot shows a correlation of 0.75, with many counties falling close to the 45-degree line. Panel (c) shows the market share by insurance carrier participating in the Covered California marketplace in 2014. Panel (d) shows the market share by insurance carrier as recovered in our transaction data. We slightly under-represent Kaiser Permanente in our sample due to difficulties distinguishing insurance premium payments from payments for other types of health services in Kaiser's billing records.

Figure 3: Summary statistics of consumers who participate in the individual insurance marketplaces, 2012-2015

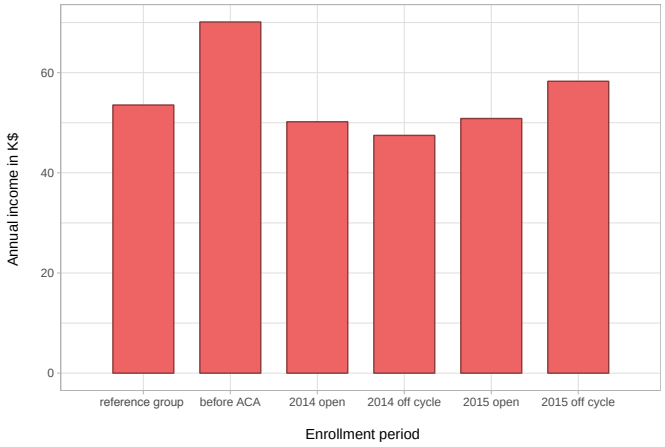
(a) Monthly sign-ups and drop-outs



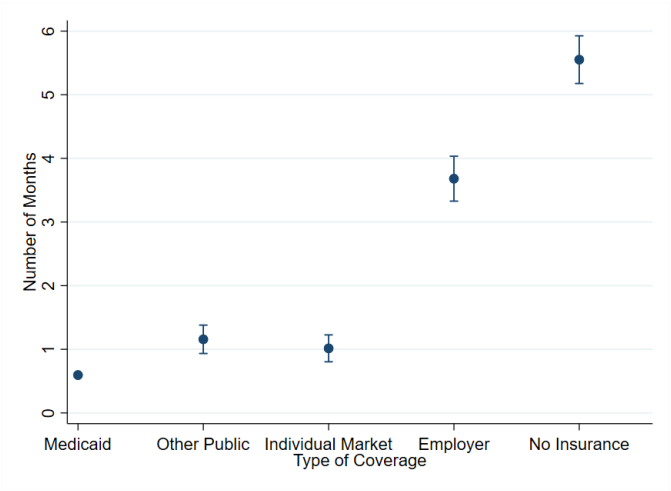
(b) Counts of members by enrollment period



(c) Median Household Income by Enrollment Groups



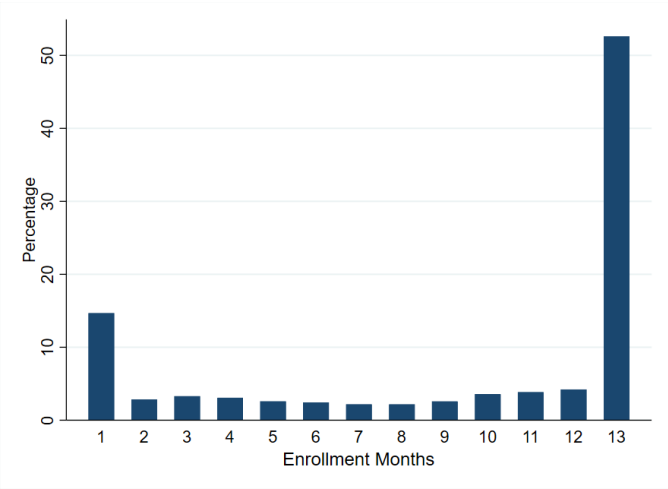
(d) 2013 Insurance of 2014 ACA Open Enrollees



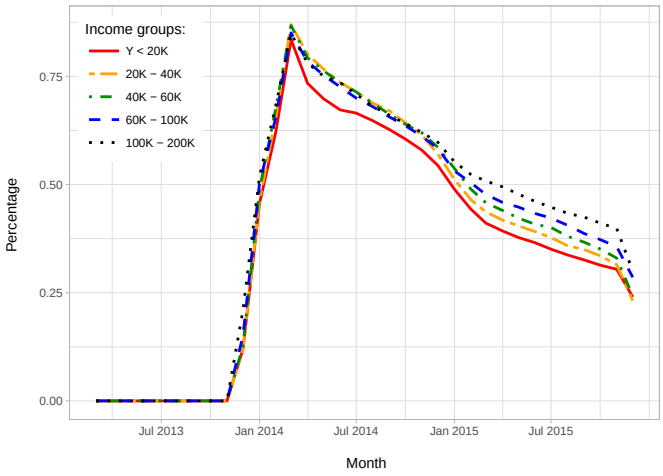
Notes: Panel (a) shows the number of new enrollee sign-ups and drop-outs in each month from Jan 2011 to Oct 2015. A new sign-up or drop-out is defined based on the observed first and last monthly premium payment. We label an enrollee as a "dropout" only if we observe his/her last premium payment at least two months prior to that consumer's exit from the bank account record data. Panel (b) shows the absolute count of members in each enrollment period. The "open enrollment" group includes households who pay their first premium between December of the prior year and March of the enrollment year; the "off-cycle" group includes households who pay their first premium between April and November of the enrollment year. Panel (c) plots the median 2013 household income of ACA enrollees by ACA enrollment cycle. Panel (d) plots the mean number of months of health insurance 2014 ACA enrollees had in 2013. Data come from the 2013-2014 MEPS Longitudinal file. Plot reports that average number of months of coverage across different types of health insurance in 2013, among those who enroll is ACA coverage during 2014 open enrollment. Sample includes all respondents across the US reporting ACA coverage at some point from January 2014 through May 2014. Sample size is 201 respondents.

Figure 4: Summary Statistics Describing Dropout Behavior

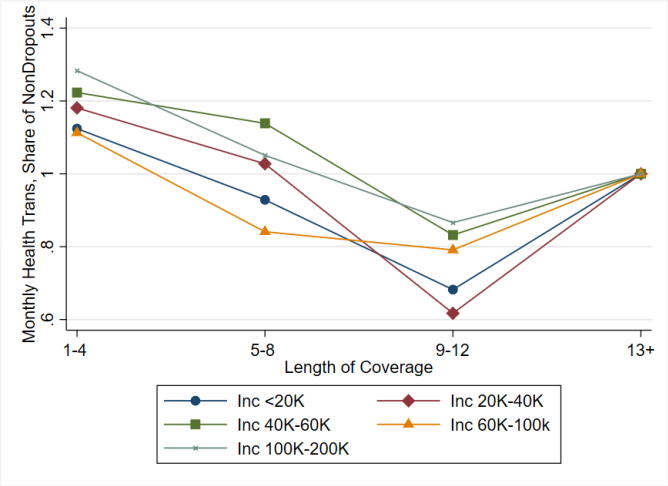
(a) Distribution of Drop-outs Across Months, for 2014 Open Enrollment Participants



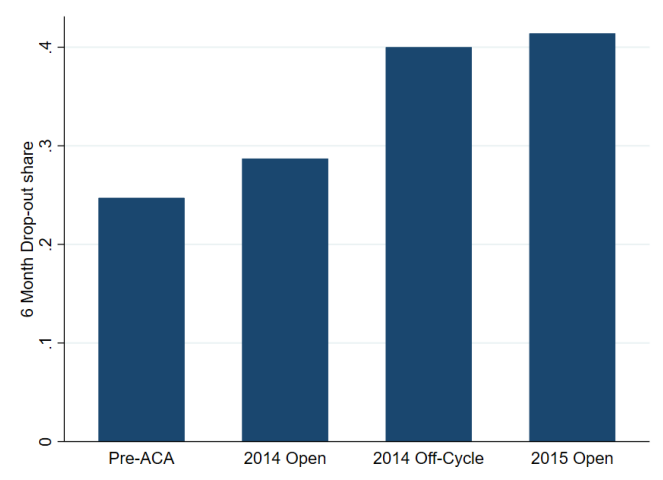
(b) Drop-out Rates by Income Groups



(c) Monthly Health Transactions when Covered by ACA, by Length of Coverage

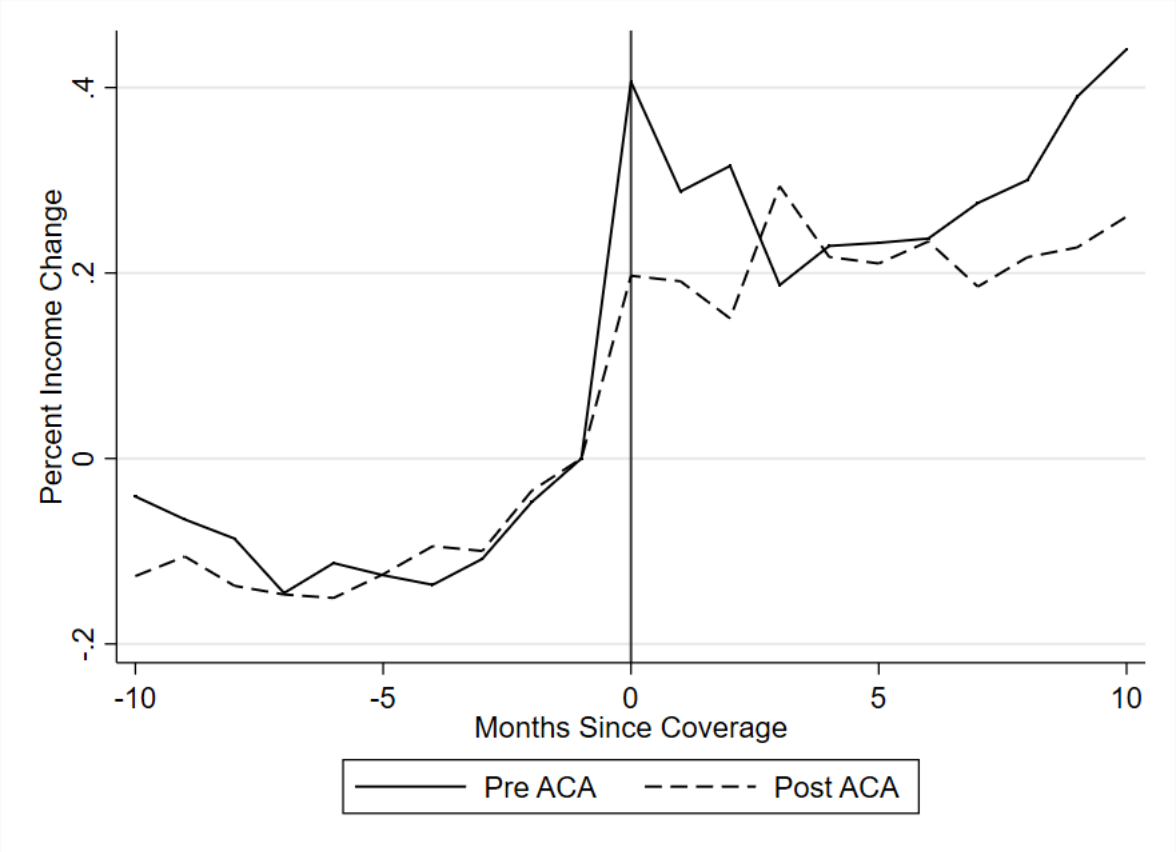


(d) Six Month Drop-out Rates by Enrollment Group



Notes: Panel (a) plots the distribution of enrollment months within 2014 open enrollees. Panel (b) plots the net share of households who continue to pay insurance premiums in each month, conditional on participating in the 2014 open enrollment period (N = 15,207). We break out this share by 2013 income. Panel (c) compares the count of health transactions during the period of ACA health insurance coverage across households of varying length of coverage. These estimates are recovered from a regression of a county of monthly health transactions on dummies for coverage length (1-4 months, 5-8 months, 9-12 months, and 13+ months). The regression is run separately for each income group and monthly and lagged monthly income are included as controls. Quantities of monthly health care usage are reported relative to the amount of the 13+ months of coverage group. These regression only include data during the period of healthy insurance coverage for each household.

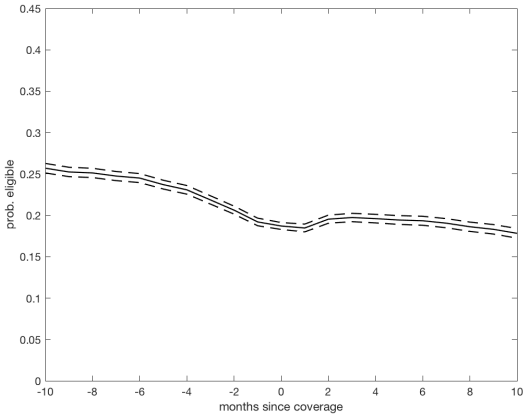
Figure 5: Monthly Income Around Drop-out



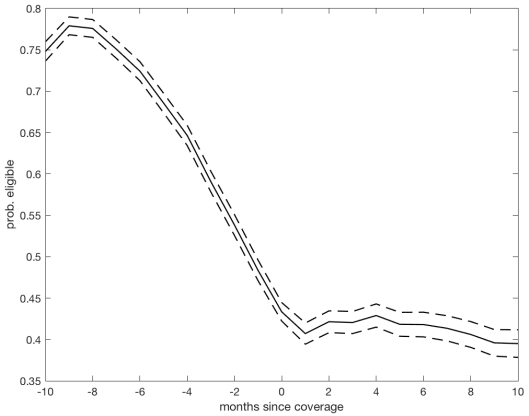
Notes: This figure compares the income dynamics around sign-up and drop-out, separately for the pre-ACA and post-ACA period. The x-axis measures months since coverage, where negative values indicate months since enrollment and strictly positive values indicate months since dropout. All months of coverage are collapsed into the event time period of 0. Pre-ACA indicates households who both sign up and drop out prior to July 2013, while post ACA indicate sign-ups that begin as of the 2014 ACA open enrollment period and drop-outs beginning in January 2014. We restrict the sample to balanced panels where at least 10 months of data is available prior to sign up and after drop out. Units are measured as a percentage change in monthly income, relative to average monthly income in the event month equal to -1.

Figure 6: **Probability Eligible for Medi-Cal by Pre-ACA Income Category**

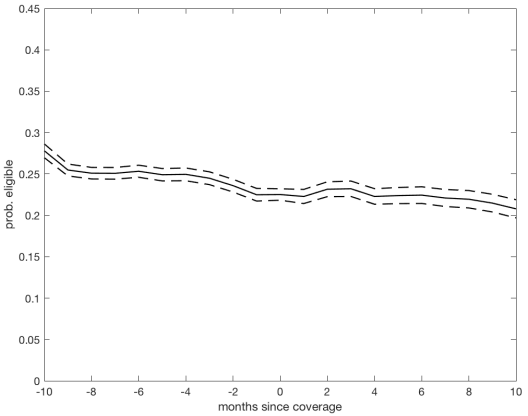
(a) All income levels



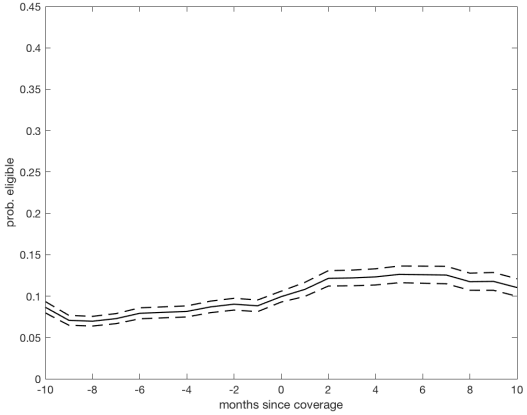
(b) Income at or below $20,0000



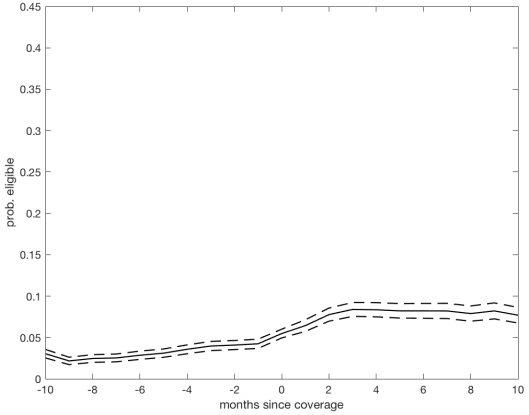
(c) Income above $20,000 and at or below $40,000



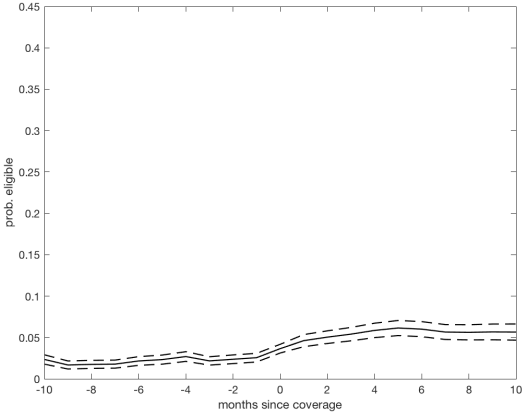
(d) Income above $40,000 and at or below $60,000



(e) Income above $60,000 and at or below $100,000



(f) Income above $100,000 and at or below $200,000



Notes: The figure illustrates the probability of Medi-Cal eligibility among households in a given annual income category. We assign income categories using pre-ACA annual income from the year 2013. The x-axis contains the months surrounding the coverage period under individual health insurance; all months of coverage are contained in the "0" period. The solid line represents the average among individual households in the income bin. The dashed lines represent the 95% confidence interval around this average at each month in the event study. The figures contain all households that (a) signed up for individual insurance coverage under the ACA, including both on and off-cycle enrollees from 2014-2016, and (b) subsequently dropped coverage before the end of the enrollment period.

Table 1: Comparison of annual spending with MEPS data

	MEPS			Transactions Data		
	drug	health	total	drug	health	total
2012	147.33	214.31	361.64	231.93	220.85	452.78
2013	121.45	262.74	384.19	272.65	231.91	504.55
2014	113.36	250.97	364.33	296.98	257.43	554.42
2015	110.30	264.76	375.06	295.08	274.22	569.30

Notes: MEPS data covers the Western region of the US, while our transactions data cover California. MEPS does not report more detailed geography. Drug spending in MEPS covers true prescription purchases, while our transactions data for drug spending covers all spending at drug stores, including non prescription purchases.

Table 2: **Health Spending Words Ranked by Urgency and Essentialness**

	"Essential" Words	"Urgent" Words
Top 5 Words		
	cancer	cancer
	tumor	trauma
	cardiology	tumor
	heart	cardiology
	ambulance	brain
Bottom 5 Words		
	acupuncture	supplement
	holistic	healthfirst
	healthfirst	allergy
	chiropractic	orthosynetics
	allergy	clinics
Total number of words	137	137
Mean reviewer score	6.855	5.62
Std dev. of revenue score	1.38	1.43
Correlation of the two metrics	0.8466	

Notes: Among the unique terms appearing in health spending records, 137 met our "relevance" criteria for inclusion (Section 3.6). We report the top and bottom five ranked words based on their mean essential score and mean urgency score, as ranked by Mechanical Turk workers.

Table 3: **Healthcare Consumption for 2014 Open Enrollees: Transactions vs. MEPS Data**

	Health Charges	Health OOP	Health Transactions	Drug OOP	Drug Transactions
Panel A: Transactions Data					
Post ACA	1295.646***	32.781*	0.680***	25.543***	0.874***
	(74.322)	(14.826)	(0.064)	(4.294)	(0.103)
Pre-Period	4549.484***	335.369***	2.839***	286.934***	9.856***
Mean	(66.725)	(11.831)	(0.062)	(5.259)	(0.157)
% Change	0.285***	0.098*	0.239***	0.089***	0.089***
	(0.019)	(0.046)	(0.025)	(0.016)	(0.011)
Observations	29494	29494	29494	29494	29494
Panel B: MEPS Data					
Post ACA	1212.748	3.977	1.187	-11.697	2.679*
	(1751.883)	(58.460)	(1.005)	(48.165)	(1.235)
Pre-Period	4292.207***	289.758***	5.614***	189.831***	8.484***
Mean	(962.012)	(39.111)	(0.800)	(44.622)	(1.291)
% Change	0.283	0.014	0.211	-0.062	0.316*
	(0.445)	(0.203)	(0.193)	(0.248)	(0.142)
Observations	402	402	402	402	402

Notes: Robust standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 4: **Health care Consumption for 2014 Open Enrollees**

	Y < 20K	20K - 40K	40K - 60K	60K - 100K	100K - 200K
			Panel A: Change in health transactions		
Treat x Post	0.635***	0.549***	0.525***	0.465***	-0.000
	(0.119)	(0.124)	(0.177)	(0.169)	(0.265)
Post ACA	0.368***	0.348***	0.482***	0.427***	0.559***
	(0.026)	(0.036)	(0.054)	(0.045)	(0.055)
Pre Period Treatment Mean	1.24	2.03	2.71	3.47	4.88
% Change	0.514***	0.270***	0.193***	0.134***	-0.000
	(0.096)	(0.061)	(0.065)	(0.049)	(0.054)
			Panel B: Change in health spending		
Treat x Post	50.085***	45.913**	52.063**	18.810	-1.480
	(15.831)	(22.683)	(23.104)	(26.282)	(48.379)
Post ACA	11.983***	12.319***	17.230***	21.489***	32.783***
	(3.388)	(3.771)	(4.691)	(5.837)	(8.172)
Pre Period Treatment Mean	111.24	185.17	224.44	340.59	565.47
% Change	0.450***	0.248**	0.232**	0.055	-0.003
	(0.142)	(0.122)	(0.103)	(0.077)	(0.086)
			Panel C: Change in drug transactions		
Treat x Post	0.640*	0.467**	0.445	0.118	0.073
	(0.334)	(0.226)	(0.303)	(0.287)	(0.333)
Post ACA	1.392***	1.087***	0.842***	0.840***	1.021***
	(0.073)	(0.083)	(0.090)	(0.078)	(0.084)
Pre Period Treatment Mean	5.60	8.23	10.81	11.63	12.48
% Change	0.114*	0.057**	0.041	0.010	0.006
	(0.060)	(0.028)	(0.028)	(0.025)	(0.027)
			Panel D: Change in drug spending		
Treat x Post	20.681**	10.125	-8.756	4.095	12.891
	(8.282)	(7.281)	(9.945)	(12.260)	(14.903)
Post ACA	36.355***	32.074***	27.655***	31.411***	36.833***
	(1.879)	(2.023)	(2.711)	(2.995)	(3.707)
Pre Period Treatment Mean	124.25	199.72	285.44	331.66	403.96
% Change	0.166**	0.051	-0.031	0.012	0.032
	(0.067)	(0.036)	(0.035)	(0.037)	(0.037)
No. of members in treatment	2627	3200	2337	2568	2084
No. of members in reference	35643	36599	28586	35411	32004

Notes: Standard errors clustered by household * $p<0.1$, ** $p<0.05$, *** $p<0.01$

The treatment group includes individuals who enrolled during the 2014 open enrollment period. The reference group includes individuals who never enrolled in any individual insurance market plan. Y represents total annual post-tax income, as computed from individual bank account records.

Table 5: **Healthcare Consumption for 2014 Open Enrollees with Children**

	Y < 20K	20K - 40K	40K - 60K	60K - 100K	100K - 200K
			Panel A: Change in health transactions		
Treat x Post	1.003**	0.748**	0.637	0.145	-0.020
	(0.421)	(0.352)	(0.784)	(0.487)	(1.053)
Post ACA	0.403***	0.459***	1.151***	0.821***	0.525***
	(0.063)	(0.117)	(0.215)	(0.167)	(0.178)
Pre Period Treatment Mean	1.52	3.06	3.47	5.09*	7.64
% Change	0.658**	0.244**	0.184	0.029	-0.003
	(0.277)	(0.115)	(0.226)	(0.096)	(0.138)
			Panel B: Change in health spending		
Treat x Post	112.546***	71.573*	55.696	-53.357	-191.555
	(41.444)	(42.300)	(62.390)	(84.827)	(172.066)
Post ACA	12.593	31.373**	21.847	15.099	23.985
	(8.726)	(15.378)	(14.880)	(17.371)	(24.544)
Pre Period Treatment Mean	116.33	189.44	274.96	513.26	920.49
% Change	0.967***	0.378*	0.203	-0.104	-0.208
	(0.356)	(0.223)	(0.227)	(0.165)	(0.187)
			Panel C: Change in drug transactions		
Treat x Post	0.630	0.245	-0.145	0.505	-0.969
	(0.962)	(0.731)	(1.012)	(0.825)	(1.238)
Post ACA	1.390***	1.218***	1.253***	1.532***	0.915***
	(0.207)	(0.304)	(0.286)	(0.239)	(0.275)
Pre Period Treatment Mean	6.91	9.39*	12.11*	12.00*	16.86*
% Change	0.091	0.026	-0.012	0.042	-0.057
	(0.139)	(0.078)	(0.084)	(0.069)	(0.073)
			Panel D: Change in drug spending		
Treat x Post	4.552	4.436	-46.798	10.940	-9.309
	(27.332)	(22.434)	(34.289)	(33.265)	(42.926)
Post ACA	43.318***	36.581***	35.467***	53.240***	19.085*
	(6.062)	(6.741)	(7.130)	(9.143)	(10.829)
Pre Period Treatment Mean	148.66	208.78	287.24	257.32	456.95*
% Change	0.031	0.021	-0.163	0.043	-0.020
	(0.184)	(0.107)	(0.119)	(0.129)	(0.094)
No. of members in treatment	389	358	218	273	191
No. of members in control	4930	3958	3133	4160	3865

Notes: Standard error clustered by household. * $p<0.1$, ** $p<0.05$, *** $p<0.01$

The treatment group includes individuals who enrolled during the 2014 open enrollment period. The reference group includes individuals who never enrolled in any individual insurance market plan. We define "changes" as the difference in expenditures or transactions from the pre-treatment period (before 2014) to the post-treatment period (starting in January 2014). Y represents total annual post-tax income, as computed from individual bank account records. We identify households with children as those whose classification score exceeds the threshold defined in the Data Appendix.

Table 6: Health Spending Around Sign-up, Drop-Out, and Medicaid Eligibility

Panel A: Health spending

	(1) <20K	(2) 20K-40K	(3) 40K-60K	(4) 60K-100K	(5) 100K-200K
Post Signup	6.504*	9.695***	2.650	4.225	1.145
	(3.355)	(3.481)	(2.527)	(3.151)	(5.400)
Post Dropout	-9.849***	-9.189***	0.0829	-6.992**	-5.773
	(2.754)	(3.413)	(2.611)	(3.213)	(5.192)
Pr Medicaid Eligible x	-3.837	-5.410	2.127	39.02**	114.1*
Post Signup	(4.494)	(8.666)	(13.01)	(17.93)	(69.07)
Pr Medcaid Eligible x	3.655	5.698	-11.36	-25.67	-86.23
Post Signup	(4.129)	(7.991)	(10.33)	(18.96)	(61.14)
Prob Medicaid Eligible	-6.977**	-8.251*	-8.082	-18.80***	-24.29
	(3.173)	(4.389)	(10.30)	(5.730)	(15.57)
Sign Up (% Change)	0.304**	0.611***	0.110	0.143	0.024
	(0.157)	(0.220)	(0.105)	(0.106)	(0.115)
Dropout (%Change)	-0.460***	-0.579**	0.003	-0.236**	-0.123
	(0.129)	(0.215)	(0.109)	(0.108)	(0.110)
Pre-Signup Mean Health Spending	21.42	15.86	24.05	29.65	47.11
Number of observations	72677	55702	36951	38203	28006

Panel B: Health transactions

Post Signup	0.112***	0.0761***	0.0588***	0.0502**	0.0330
	(0.0236)	(0.0182)	(0.0177)	(0.0218)	(0.0254)
Post Dropout	-0.0723***	-0.0476***	-0.0542***	-0.0446**	-0.0431
	(0.0221)	(0.0167)	(0.0181)	(0.0220)	(0.0273)
Pr Medicaid Eligible x	-0.0340	0.00529	-0.0460	0.126	0.158
Post Signup	(0.0331)	(0.0451)	(0.0810)	(0.0770)	(0.199)
Pr Medcaid Eligible x	0.00146	-0.0170	-0.000539	-0.144*	-0.343*
Post Signup	(0.0306)	(0.0431)	(0.0641)	(0.0760)	(0.185)
Prob Medicaid Eligible	-0.0893***	-0.117***	-0.0728	-0.104***	-0.0321
	(0.0202)	(0.0276)	(0.0585)	(0.0387)	(0.0852)
Sign Up (% Change)	0.440***	0.351***	0.191***	0.134**	0.064
	(0.093)	(0.084)	(0.057)	(0.058)	(0.049)
Dropout (%Change)	-0.285***	-0.219***	-0.176***	-0.119**	-0.084
	(0.087)	(0.077)	(0.059)	(0.059)	(0.053)
Pre-Signup Mean Health Transactions	0.25	0.22	0.31	0.38	0.51
Number of observations	72677	55702	36951	38203	28006

Notes: Standard errors are clustered at the individual level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. signup is a dummy indicating the month is at or post month of sign-up. dropout is a dummy indicating the month is post drop-out. All regressions include individual fixed effects. All data is restricted to balanced panels where at least 10 months of data is available prior to sign up and after drop out. Units are measured as a percentage change, relative to average consumption amount in the ten months leading up to sign-up. All regressions control for monthly income and average monthly income from prior three months.

Table 7: Drug Spending Around Sign-up, Drop-Out, and Medicaid Eligibility

Panel A: Drug spending					
	(1)	(2)	(3)	(4)	(5)
	<20K	20K-40K	40K-60K	60K-100K	100K-200K
Post Signup	-0.944	0.141	1.770	2.512	-1.667
	(1.318)	(1.311)	(1.304)	(1.823)	(2.120)
Post Dropout	-0.318	-1.086	-4.118***	-5.696***	-3.380*
	(1.163)	(1.312)	(1.355)	(1.920)	(1.994)
Pr Medicaid Eligible x	4.400**	5.577*	5.618	3.055	12.03
Post Signup	(2.004)	(3.027)	(4.810)	(6.276)	(10.33)
Pr Medcaid Eligible x	-4.309***	-5.806*	-1.243	1.515	-11.24
Post Signup	(1.542)	(3.073)	(4.771)	(6.545)	(7.135)
Prob Medicaid Eligible	-12.32***	-12.96***	-15.52***	-18.86***	-20.10**
	(1.237)	(2.131)	(3.362)	(3.835)	(10.20)
Sign Up (% Change)	-0.042	0.006	0.063	0.077	-0.038
	(0.059)	(0.060)	(0.047)	(0.056)	(0.049)
Dropout (%Change)	-0.014	-0.050	-0.148***	-0.176***	-0.077
	(0.052)	(0.060)	(0.049)	(0.059)	(0.046)
Pre-Signup Mean Drug Spending	22.41	21.91	27.91	32.43	43.66
Number of observations	72677	55702	36951	38203	28006
Panel B: Drug transactions					
Post Signup	-0.00307	0.0420	0.0635*	-0.00432	0.0228
	(0.0312)	(0.0314)	(0.0356)	(0.0299)	(0.0355)
Post Dropout	-0.0312	-0.0506*	-0.0997***	-0.106***	-0.0933**
	(0.0294)	(0.0278)	(0.0342)	(0.0350)	(0.0385)
Pr Medicaid Eligible x	0.136***	0.0689	0.244*	0.122	0.159
Post Signup	(0.0457)	(0.0786)	(0.145)	(0.127)	(0.163)
Pr Medcaid Eligible x	-0.104**	-0.169**	-0.212	-0.0495	-0.364***
Post Signup	(0.0482)	(0.0797)	(0.134)	(0.130)	(0.136)
Prob Medicaid Eligible	-0.442***	-0.398***	-0.508***	-0.531***	-0.328**
	(0.0374)	(0.0502)	(0.103)	(0.0828)	(0.136)
Sign Up (% Change)	-0.003	0.046	0.060**	-0.004	0.017
	(0.035)	(0.034)	(0.033)	(0.025)	(0.026)
Dropout (%Change)	-0.035	-0.055*	-0.094***	-0.090***	-0.069***
	(0.033)	(0.031)	(0.032)	(0.030)	(0.028)
Pre-Signup Mean Drug Spending	0.89	0.91	1.07	1.18	1.35
Number of observations	72677	55702	36951	38203	28006

Notes: Standard errors are clustered at the individual level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. signup is a dummy indicating the month is at or post month of sign-up. dropout is a dummy indicating the month is post drop-out. All regressions include individual fixed effects. All data is restricted to balanced panels where at least 10 months of data is available prior to sign up and after drop out. Units are measured as a percentage change, relative to average consumption amount in the ten months leading up to sign-up. All regressions control for monthly income and average monthly income from prior three months.

Table 8: Health Spending Around Sign-up and Drop-Out

	(1) <20K	(2) 20K-40K	(3) 40K-60K	(4) 60K-100K	(5) 100K-200K
Panel A: Change in health spending: Pre ACA Sign-up/Drop-out					
Post Signup	23.93***	2.670	5.133	3.060	12.26
	(7.855)	(3.156)	(4.060)	(3.351)	(7.904)
Post Dropout	-19.36**	-0.625	-2.100	-3.890	-2.766
	(7.532)	(3.347)	(4.119)	(2.960)	(8.065)
Sign Up (% Change)	1.170***	0.133	0.176	0.099	0.339
	(0.384)	(0.158)	(0.139)	(0.108)	(0.218)
Dropout (%Change)	-0.947***	-0.031	-0.072	-0.126	-0.076
	(0.368)	(0.167)	(0.141)	(0.096)	(0.223)
Pre-Signup Mean Health Spending	21.36	20.82	25.91	30.57	40.01
Number of observations	25641	17879	15311	19255	15296
Panel B: Change in health spending: Post ACA Sign-up/Drop-out					
Post Signup	5.879***	8.602***	2.649	5.494*	4.401
	(2.074)	(2.082)	(2.413)	(3.048)	(6.072)
Post Dropout	-8.620***	-8.004***	-1.273	-8.148***	-8.367
	(1.683)	(2.316)	(2.377)	(3.078)	(5.864)
Sign Up (% Change)	0.275***	0.542***	0.110	0.185**	0.093
	(0.097)	(0.131)	(0.100)	(0.103)	(0.129)
Dropout (%Change)	-0.403***	-0.505***	-0.053	-0.275**	-0.178
	(0.079)	(0.146)	(0.099)	(0.104)	(0.124)
Pre-Signup Mean Health Spending	21.42	15.86	24.05	29.65	47.11
Number of observations	72677	55702	36951	38203	28006
Panel C: Change in health transactions: Pre ACA Sign-up/Drop-out					
Post Signup	0.0680***	0.0449*	0.0131	0.0199	0.0980***
	(0.0201)	(0.0256)	(0.0273)	(0.0327)	(0.0309)
Post Dropout	-0.0181	-0.0142	-0.0234	-0.0474*	-0.0416
	(0.0199)	(0.0240)	(0.0275)	(0.0256)	(0.0333)
Sign Up (% Change)	0.318***	0.180***	0.039	0.050***	0.279***
	(0.094)	(0.103)	(0.082)	(0.083)	(0.088)
Dropout (%Change)	-0.085***	-0.057***	-0.070	-0.120***	-0.118***
	(0.093)	(0.096)	(0.082)	(0.065)	(0.095)
Pre-Signup Mean Health Transactions	0.24	0.26	0.31	0.36	0.40
Number of observations	25641	17879	15311	19255	15296
Panel D: Change in health transactions: Post ACA Sign-up/Drop-out					
Post Signup	0.108***	0.0777***	0.0537***	0.0538**	0.0374
	(0.0147)	(0.0130)	(0.0163)	(0.0210)	(0.0250)
Post Dropout	-0.0711***	-0.0512***	-0.0574***	-0.0535**	-0.0573**
	(0.0144)	(0.0132)	(0.0162)	(0.0210)	(0.0267)
Sign Up (% Change)	0.426***	0.359***	0.175***	0.143***	0.073
	(0.058)	(0.060)	(0.053)	(0.056)	(0.048)
Dropout (%Change)	-0.280***	-0.236***	-0.187***	-0.142***	-0.111
	(0.057)	(0.061)	(0.053)	(0.056)	(0.052)
Pre-Signup Mean Health Transactions	0.25	0.22	0.31	0.38	0.51
Number of observations	72677	55702	36951	38203	28006

Notes: Standard errors are clustered at the individual level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. signup is a dummy indicating the month is at or post month of sign-up. dropout is a dummy indicating the month is post drop-out. All regressions include individual fixed effects. Pre-ACA indicates households who both sign up and drop-out prior to July 2013, while post ACA indicate sign-ups that begin as of the 2014 ACA open enrollment period and drop-outs beginning in January 2014. All data is restricted to balanced panels where at least 10 months of data is available prior to sign up and after drop out. Units are measured as a percentage change, relative to average consumption amount in the ten months leading up to sign-up. All regressions control for monthly income and average monthly income from prior three months.

Table 9: Drug Spending Around Sign-up and Drop-Out

Panel A: Change in drug spending: Pre ACA Sign-up/Drop-out					
	(1)	(2)	(3)	(4)	(5)
	<20K	20K-40K	40K-60K	60K-100K	100K-200K
Post Signup	11.66***	5.370***	6.733***	13.20**	15.11***
	(1.368)	(1.328)	(1.730)	(5.200)	(3.167)
Post Dropout	-2.947**	1.744	-1.115	-2.906	-0.0310
	(1.452)	(2.368)	(1.844)	(2.852)	(3.062)
Sign Up (% Change)	0.773***	0.291***	0.283***	0.533***	0.560***
	(0.091)	(0.072)	(0.073)	(0.210)	(0.117)
Dropout (%Change)	-0.195***	0.094***	-0.047***	-0.117***	-0.001***
	(0.096)	(0.128)	(0.077)	(0.115)	(0.113)
Pre-Signup Mean Drug Spending	16.15	20.20	23.34	26.23	33.46
Number of observations	25641	17879	15311	19255	15296
Panel B: Change in drug spending: Post ACA Sign-up/Drop-out					
Post Signup	1.626**	1.381	1.973*	2.377	-1.540
	(0.740)	(1.083)	(1.182)	(1.739)	(2.065)
Post Dropout	-1.677*	-2.304**	-4.515***	-6.016***	-4.024**
	(0.861)	(1.085)	(1.246)	(1.821)	(1.927)
Sign Up (% Change)	0.073**	0.063**	0.071***	0.073***	-0.035***
	(0.033)	(0.049)	(0.042)	(0.054)	(0.047)
Dropout (%Change)	-0.075**	-0.105**	-0.162***	-0.186***	-0.092***
	(0.038)	(0.050)	(0.045)	(0.056)	(0.044)
Pre-Signup Mean Drug Spending	21.97	21.68	28.81	32.60	41.37
Number of observations	72677	55702	36951	38203	28006
Panel C: Change in drug transactions: Pre ACA Sign-up/Drop-out					
Post Signup	0.412***	0.279***	0.301***	0.337***	0.431***
	(0.0405)	(0.0425)	(0.0547)	(0.0444)	(0.0706)
Post Dropout	-0.111***	-0.117***	-0.0622	-0.0467	-0.0354
	(0.0387)	(0.0430)	(0.0537)	(0.0445)	(0.0598)
Sign Up (% Change)	0.718***	0.374***	0.346***	0.457***	0.481***
	(0.070)	(0.057)	(0.063)	(0.060)	(0.079)
Dropout (%Change)	-0.194***	-0.156***	-0.071***	-0.063***	-0.040***
	(0.067)	(0.058)	(0.062)	(0.060)	(0.067)
Pre-Signup Mean Drug Transactions	0.57	0.75	0.87	0.74	0.90
Number of observations	25641	17879	15311	19255	15296
Panel D: Change in drug transactions: Post ACA Sign-up/Drop-out					
Post Signup	0.0825***	0.0580**	0.0745**	-0.00662	0.0236
	(0.0214)	(0.0236)	(0.0324)	(0.0290)	(0.0345)
Post Dropout	-0.0637***	-0.0861***	-0.129***	-0.120***	-0.113***
	(0.0210)	(0.0227)	(0.0326)	(0.0329)	(0.0374)
Sign Up (% Change)	0.093***	0.064***	0.070***	-0.006***	0.017***
	(0.024)	(0.026)	(0.030)	(0.025)	(0.026)
Dropout (%Change)	-0.072***	-0.094***	-0.121***	-0.102***	-0.084***
	(0.024)	(0.025)	(0.031)	(0.028)	(0.028)
Pre-Signup Mean Drug Transactions	0.88	0.89	1.07	1.19	1.29
Number of observations	72677	55702	36951	38203	28006

Notes: Standard errors are clustered at the individual level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. signup is a dummy indicating the month is at or post month of sign-up. dropout is a dummy indicating the month is post drop-out. All regressions include individual fixed effects. Pre-ACA indicates households who both sign up and drop-out prior to July 2013, while post ACA indicate sign-ups that begin as of the 2014 ACA open enrollment period and drop-outs beginning in January 2014. All data is restricted to balanced panels where at least 10 months of data is available prior to sign up and after drop out. Units are measured as a percentage change, relative to average consumption amount in the ten months leading up to sign-up. All regressions control

Table 10: Urgent and Essential Health Spending, for Continuing Enrollees vs. Dropouts

	% Very Essential	% Very Urgent
Drop-out consumers	-1.296	0.085
	(0.808)	(0.489)
Enrolled in Insurance	-0.228	0.344
	(0.620)	(0.361)
Drop-out consumer × Enrolled in Insurance	-1.208	-1.353**
	(1.060)	(0.624)
Enrolled in Insurance +	-1.435*	-1.008**
Drop-out consumer × Enrolled in Insurance	(0.859)	(0.509)
Mean	10.35	3.42
Std. dev.	25.07	15.12
Observations	11,641	11,641

Notes: Robust standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. We cluster the standard errors at the individual level. For this regression, we use a subset of our analysis dataset that includes individuals who participated in the 2014 open enrollment and completed at least one health transaction containing a relevant health word. 7,855 members out of the 15,207 open enrollment members appear in this sample.

The two outcome variables, "% Very Essential" and "% Very Urgent" measure the percentage of total health transactions that are considered very essential and very urgent, respectively, by their relevance score (Section 3.6). Drop-out consumers are those enrollees who discontinue premium payments within 8 months of sign-up. The Post-ACA period includes months in 2014 in which the enrollees maintain coverage. We test the null hypothesis that the drop-out consumers maintain the same health consumption in 2014 during their period of coverage, relative to continuing enrollees.

Table 11: **Pricing Regressions**

	(1)	(2)	(3)	(4)
	Price Change (%)	**Price Change (%)**	**Price Change (%)**	**Drop-out Rate (log)**
Drop-out Rate	-0.142***	-0.154**	-0.188**	
	(0.054)	(0.064)	(0.089)	
Mean Price (log)				1.790**
				(0.724)
Observations	145	143	145	147
R^2	0.403	0.412	0.639	0.822
Year FE x Region FE			YES	YES
Region FE X Company FE			YES	YES
Controls		YES		

Notes: Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. All models include: indicators for whether a plan was the cheapest or second cheapest plan in the area, insurance carrier fixed effects, year fixed effects, and region fixed effects. Models (3) and (4) include carrier-region and region-year fixed effects. We label plans as "cheapest" and "second cheapest" according to their premium levels in a given year and in a given geographic rating region. We focus only on silver tier plans in a rating region. The "Drop-out Rate" represents the share of enrollees who discontinue premium payments within 8 months of sign-up, for a specific insurance carrier, rating region, and year t. The *Price Change*, the outcome in Columns 1-3, is the average percentage change in the silver tier plan price from year t to year $t+1$ offered by an insurance carrier in a rating region. Column 2 includes controls for the average number of monthly health transactions for members enrolled in silver tier plans, and the share of enrollees who enrolled in a plan off-cycle–i.e. in the months of April through December.

A Data Appendix

A.1 Measuring Income

We measure income as all deposits into a user's bank accounts, after excluding transfers between accounts, withdrawals from brokerage accounts, wire transfers, tax refunds, and loan disbursements.

A.2 Classifying Premium Payments

We manually searched for premium payments for each insurer participating in the Covered California market. We first pull all transactions that contain the word 'KAISER' or 'KP' and conduct some data and text analysis. Table A7 reports the keywords we use for each insurer and which words we exclude for each insurer. The excluded key words ensure that we do not include non-ACA insurance purchases in our sample or mis-classify insurers with similar names (such as Blue Cross and Blue Shield, who are different firms in California). We selected the key words by trial and error until we found the correct phrase each insurer uses in their billing description.

The most complicated insurer in our data is Kaiser, which is both an insurer and health care provider. Here, we had to sort out the Kaiser transactions into premium payments, health care payments and drug payments. Our classification of words and phrases into premium, drug and health categories are mainly based on two criteria: frequency and dollar amount of the transaction. For example, premium transactions, such as 'KAISER DIRECT PAY' has a median of $262.1 and mean of $364.3 and often appear once every month at the individual level. On the other hand, drug transactions, such as 'KAISER CPP' has a median of $11.65 and a mean of $21.47 and often appear several times every month at the individual level. Another important consideration is the exact time a certain word/phrases starts to appear or disappear. For example, transactions with the phrase "KAISER HEALTH PLAN", which are clearly premium payments, remain roughly at 6% of the total KAISER transactions until they completely disappear from the dataset in Nov 2013. Right around this time, a new phrase 'CSC DIRECT PAY' starts to emerge and the dollar amounts of these transactions are almost identical to the previous "KAISER HEALTH PLAN" at the individual level. We thus conclude that 'CSC DIRECT PAY' are also premium transactions although they don't contain the word 'KAISER' at all.

Because we identify enrollees by their premium payments, we cannot capture enrollees who pay zero premiums. This can arise if an enrollee choses to not buy the benchmark plan, but to use the subsidy toward a cheaper (bronze) plan.

A.3 Classifying Transactions

While our data provider has developed an internal classification for each transaction into one of many categories, their definition of a health transaction is not quite appropriate for our analyses. For example, the data provider's definition included some items which we would not consider health care spending, such as gyms, spas, and other well-being consumption. We therefore developed our own unique classifier based off of the internal measure. First, we started with a sample of health care transactions, as classified by the data provider. We manually purged this sample of items that did not fit a definition of health care useful for our insurance market examination.[44] This sample formed our "training sample". Our goal was to narrow the definition of health spending, as previously defined, to include only transactions pertinent to medical payments such as doctor visits, medical procedures and lab testing. For drug transactions, we manually selected transactions that contain the names of major nationwide drugstore chains such as CVS, Walgreens, Rite Aid, as well as some local drugstores such as Duane Reade in New York, and Navarro in Florida, as the training sample. One caveat with this approach is that it is hard to separate general grocery spending from drug spending at these retail outlets. Next, we counted the frequency of each word appearing in the training samples of health, drug and 'other', respectively. To reduce the size of look-up table, words which appeared less then 5 times in total were dropped, resulting in a word bank of 84,263 words, of which 11,387 also appeared in the health category and 7,347 appeared in the drug category. We then applied a modified version of Laplacian correction by expanding the total word-counts of each category by 1% and equally distributing them to all words. For example, each word on the health list, existing or missing, received an additional count of 324846 * 0.01 / 84263 = 0.038 and each word on the drug list received an additional count of 745416 * 0.01 / 84263 = 0.0884. Finally, we calculated the probability of each word in each category as well as in total.

A.4 Determining Medi-Cal Eligibility

To determine the probability that a household might be eligible for Medi-Cal in a given month, we develop a three part procedure. First, we convert our observed post-tax income into a pre-tax measure. This conversion will depend on family size, which we do not observe in the data. We therefore make the conversion under a range of family size assumptions. Second, we collect monthly income eligibility thresholds based on the federal poverty level set for each family size and for each year. We compare our pre-tax income by family size against the thresholds to determine Medi-Cal eligibility in each family size bin for each household in our data. Finally, third, we collect data on the empirical distribution of family sizes by income within the population of California residents who purchase individual insurance. We then multiply the likelihood that a household falls into each family size bin by the eligibility determination. We sum these values to find a single weighted measure of the probability of Medi-Cal eligibility in each month for each household in our bank data. We describe this procedure below in more detail.

[44] We removed transactions such as gym membership, wellness supplements, pet services, drugstore spending and premium payments. We had already classified premiums manually. Similarly, we removed drug spending, since we created a separate category of spending containing only drug transactions.

A.4.1 Calculating Pre-Tax income

To convert our post-tax income measures into pre-tax measures, we need to assign average tax rates for federal, California, and federal (FICA) payroll taxes to each household in our dataset. To do this, we use the National Bureau of Economic Research's TAXSIM software package. Specifically, we use the following procedure:

1. Using our observed post-tax annual income measure, create a grid of pre-tax income around the observed post-tax amounts. We set the end points of the grid at 0 and the max of our post-tax income divided by $(1 - .4)$, to approximate a 40% average tax rate. We fill in the grid with approximately 10,000 points, with a larger spacing between grid points as the grid approaches the maximum value.
2. Create a matrix of incomes and family sizes to feed into NBER's TAXSIM (v27) software. We assign the state to California and consider five alternative family sizes:
   (a) single adult
   (b) two adults
   (c) two adults plus one child under 13
   (d) two adults plus two children under 13
   (e) two adults plus 4 children, all under 13
3. Feed the income and family size-specific matrices into the TAXSIM software. The program produces output that includes federal, state, and FICA payroll taxes in dollars. We convert these dollars of tax to an average tax rate by dividing the dollars by the total pre-tax income. We then sum the state, federal and fica tax rates into a single rate, τ_i.
4. Convert the income in this simulated dataset to be post-tax by multiplying the pre-tax income by $1 - \tau_i$.
5. To be able to match to the observed post-tax income in the bank dataset, create bins of the hypothetical post-tax income containing 10 consecutive entries in the post-tax income grid. Within each bin, we find the average of τ_i across all i's in the bin and label it $\bar{\tau}_i$.
6. For each observed post-tax income in our bank data, match the income to one of the bins created above. We assign to that observation the $\bar{\tau}_i$ for the bin. We then calculate pre-tax income by dividing the observed post-tax income by $(1 - \bar{\tau}_i)$.
7. Smooth over fluctuations around the borders between months by calculating a rolling average of pre-tax income by averaging periods
$$(t - 1, t, t + 1)$$
income together. In doing so, we lose the first and last period's observed income for each individual.

A.4.2 Eligibility determinations

Given a level of pre-tax income for each hypothetical family size, we compare these income measures to the monthly Federal Poverty Line (FPL) thresholds appropriate for adults aged 18 to 64 under the ACA. We use 138% of the FPL as our threshold. We use this threshold in 2013 as well, even though eligibility under the ACA guidelines didn't come into effect until 2014. Using the year 2013 data allows us to examine how hypothetical eligibility changed for open enrollment participants in 2014 in the period before sign-up compared to the period after drop-out.

A.4.3 Family Size Weighting

With eligibility determinations for each household at different hypothetical family sizes, we move next to find a single statistic for each household. To do so, we merge in the family size distribution amongst purchasers of ACA insurance from the California Health Interview Survey (CHIS).[45]. We observe family size shares (weighted to be representative for the state of California) for 2013-2016 for five different income bins. The family sizes we exploit include 1-4 family members, as well as 6 members to represent the bin of 5+ family members. We match the five dimensional family size weight vector to each family's five dimensional eligibility vector by income. That is, we use the household's annual income in a year and match it to the CHIS income bins. We compute a single probability of eligibility by multiplying each of the 5 eligibility indicators for an individual-month (for 5 family sizes) by the probability of being in each family size in a given year from the CHIS. We report summary statistics on these probabilities in Table A6. In Figures **??** and 6, we plot the probability of Medi-Cal eligibility around the period of individual insurance coverage.

B Model Details

B.1 Second Period Equilibrium

We begin our analysis by determining the Nash pricing equilibrium in the second period. We start by analyzing the enrollment decisions of the mass νt of new consumers. A new type n non-dropout consumer buys insurance from A if $p_2^A + x < p_2^B + (t - x)$. A new type d dropout consumer buys insurance from A if $\phi p_2^A + x < \phi p_2^B + (t - x)$. From this, we can immediately see that the decisions of dropout consumers are less sensitive to price changes than the decisions of non-dropout consumers. This is, of course, intuitive since dropout consumers only have to absorb a fraction of the increase in the annual premium.

Also in the market are the existing customer bases of insurers A and B who remain from period 1, with masses $(1 - \nu)\sigma^A t$ and $(1 - \nu)\sigma^B t$, respectively. As noted previously, these consumers face switching costs when moving from their current insurer to a competitor. As in Klemperer (1987), we assume that these switching costs are sufficiently high such that existing consumers who enrolled in period 1 do not switch firms in the second period.

[45] More detailed on the CHIS is available here: https://healthpolicy.ucla.edu/chis/data/Pages/GetCHISData.aspx.

From the logic above, given a set of second period prices, we can calculate each insurer's total expected sales to each type of consumer shopping in the market for insurance in period 2. First, insurer A's expected sales to non-dropout type consumers is given by:

$$q_2^{A,n}\left(p_2^A, p_2^B\right) = \nu(1-\mu_2)\left(\frac{t + p_2^B - p_2^A}{2}\right)$$

and similarly for insurer B. Total expected sales to dropout types is similar, but reflects the lower price sensitivity of these consumers:

$$q_2^{A,d}\left(p_2^A, p_2^B\right) = \nu\mu_2\left(\frac{t + \phi\left(p_2^B - p_2^A\right)}{2}\right).$$

Again, the expression is symmetric for insurer B. Finally, due to the high switching costs, each insurer continues to sell to its existing customer base, with total sales equal to:

$$q_2^{A,e}\left(p_2^A, p_2^B\right) = (1-\nu)\,\sigma^A t$$

and likewise for insurer B.

From these expressions, we can easily calculate the own-price sensitivity of each component of the demand function an insurer faces. In particular, we find that $\partial q_2^{A,n}/\partial p_2^A = -\nu(1-\mu_2)/2$, $\partial q_2^{A,d}/\partial p_2^A = -\nu\mu_2\phi/2$ and $\partial q_2^{A,e}/\partial p_2^A = 0$, with symmetric expressions for insurer B. Letting π_2^A and π_2^B denote the period 2 profits of insurers A and B, we calculate the own-price sensitivity of A's profits to be:

$$\begin{aligned}
\frac{\partial \pi_2^A}{\partial p_2^A} &= \left[q_2^{A,n} + q_2^{A,e} + \left(p_2^A - c\right)\frac{\partial\left(q_2^{A,n} + q_2^{A,e}\right)}{\partial p_2^A}\right] + \left[\phi q_2^{A,d} + \left(\phi p_2^A - c\right)\frac{\partial q_2^{A,d}}{\partial p_2^A}\right] \\
&= 1/2\left\{\left(\sigma^A - \sigma^B\right)(1-\nu)\,t + t\left[(1-\nu) + \nu(1-\mu_2) + \nu\mu_2\phi\right]\right\} \\
&\quad +1/2\left\{\nu\left(1-\mu_2\right)\left(p_2^B - 2p_2^A + c\right) + \nu\mu_2\phi^2\left(p_2^B - 2p_2^A + c/\phi\right)\right\}
\end{aligned}$$

with a similar expression for B.

Nash equilibrium requires that each insurer's price is a best-response to the price set by its competitor. That is, equilibrium requires $\partial\pi_2^A/\partial p_2^A = \partial\pi_2^B/\partial p_2^B = 0$. Solving this system of equations for p_2^A and p_2^B yields:

$$p_2^A = \frac{\nu[(1-\mu_2) + \mu_2\phi](c+t) + 2/3\left(\sigma^A + 1\right)(1-\nu)\,t}{\nu\left[(1-\mu_2) + \mu_2\phi^2\right]} \tag{5}$$

$$p_2^B = \frac{\nu[(1-\mu_2) + \mu_2\phi](c+t) + 2/3\left(\sigma^B + 1\right)(1-\nu)\,t}{\nu\left[(1-\mu_2) + \mu_2\phi^2\right]} \tag{6}$$

as the second period equilibrium prices.

B.2 First Period Equilibrium

For the purpose of generating the empirical predictions that we are interested in, we do not need to explicitly solve for the first period price equilibrium. It is sufficient to note that, given the uniform distribution of

consumer preferences and insurer beliefs that being a dropout type is uncorrelated with location, the first period equilibrium will by symmetric in prices ($p_1^A = p_1^B$). This symmetry in turn implies that all consumers with $x < t/2$ will chose insurer A and all consumers with $x > t/2$ will choose insurer B.

In general, *for a given first-period demand*, the presence of switching costs in the second period makes firms compete more aggressively in the first period. Intuitively, firms want to maximize market share in the first period since they know they will be able to monopolize their remaining first-period consumers in the second period. All else equal, this increased competition leads to lower equilibrium prices in the first period than would otherwise be the case if firms instead were myopic and neglected the impact of their current actions on future outcomes.

The exact structure of first-period demand and thus the first-period price equilibrium, however, depends on consumer beliefs. If consumers are naive and do not take into account the impact of current market share on future prices, then first-period prices are unambiguously lower than they would be in a world without switching costs. If, however, consumers rationally anticipate the impact of current market share on future prices, then they are less attracted by a price cut, since they know that a higher market share today means higher future prices which they will be locked into. Demand thus becomes less elastic than would otherwise be the case, which can lead to situations where the first period is actually less competitive than a world without switching costs. Regardless of how consumer beliefs are specified, though, the price equilibrium is symmetric. See Klemperer (1987) for a detailed discussion of all these issues.

Proposition B.1. *The insurer with the higher price increase between the first and second periods will have a higher dropout rate among its new enrollees.*

Proof. This result simply reflects that dropout types are less price sensitive than non-dropout types. First, let us assume without loss of generality that it is insurer A with the higher price increase. Since first-period prices are symmetric, this implies that $p_2^A > p_2^B$. For non-dropout types all consumers with $x < x^n = (t+p_2^B-p_2^A)/2$ will choose insurer A and those with $x > x^n$ will choose insurer B. For dropout types, all consumers with $x < x^d = (t+\phi(p_2^B - p_2^A))/2$ will choose insurer A and those with $x > x^d$ will choose insurer B. Since $\phi < 1$, it is immediately obvious that $x^n < x^d$. Thus, a greater fraction of insurer A's period 2 new enrollees will be dropout types compared to insurer B. □

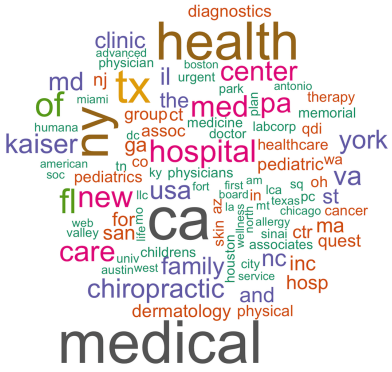
Proposition B.2. *The insurer that experiences the lower (higher) dropout rate in period 1 will increase prices more (less) than its competitor.*

Proof. The proof follows directly from equations (5)-(6). Without loss of generality, suppose insurer A experienced the lower dropout rate in period 1. Then $\sigma_A > \sigma_B$, i.e. insurer A has the larger customer base going into period 2. By inspection, this implies $p_2^A > p_2^B$. Since prices are symmetric in period 1, this means insurer A raises prices more than insurer B. □

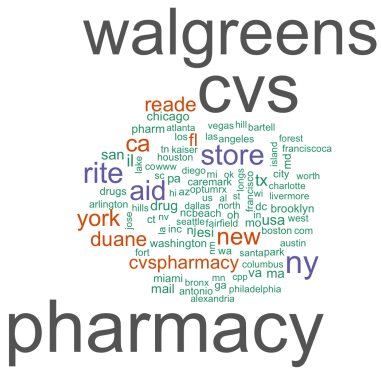
C Appendix Figures and Tables

Figure A1: Word clouds for the machine learning predictions, health and drug spending

(a) health



(b) drug



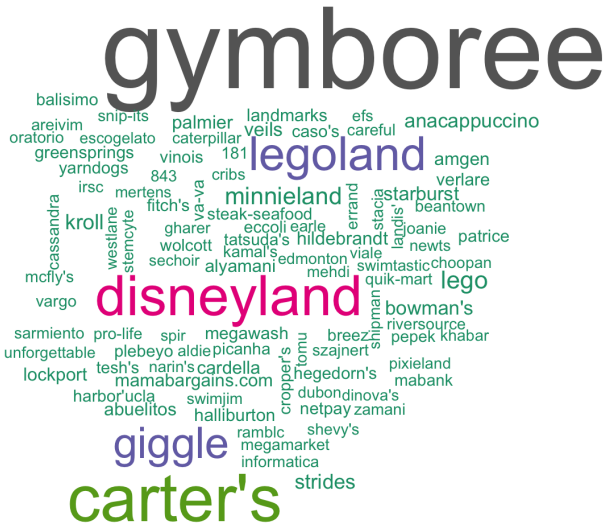
(c) Others



Notes: In this figure, we report the top 100 words classified by our machine learning algorithm as health spending, drug spending, or other non-health spending. The font size indicates the frequency with which the word appears in the transactions dataset.

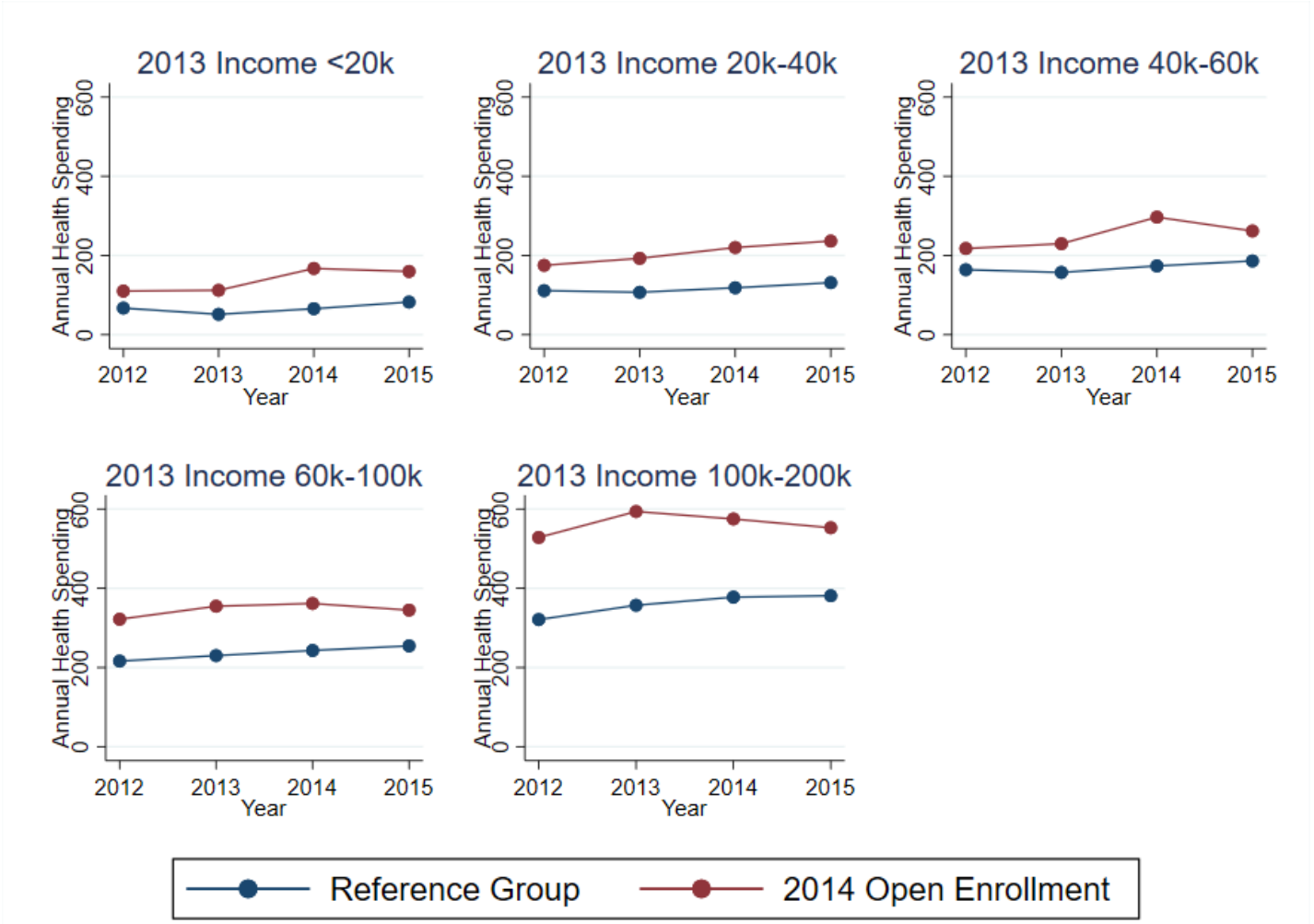
Figure A2: Word clouds for the machine learning prediction, families with children

(a) children



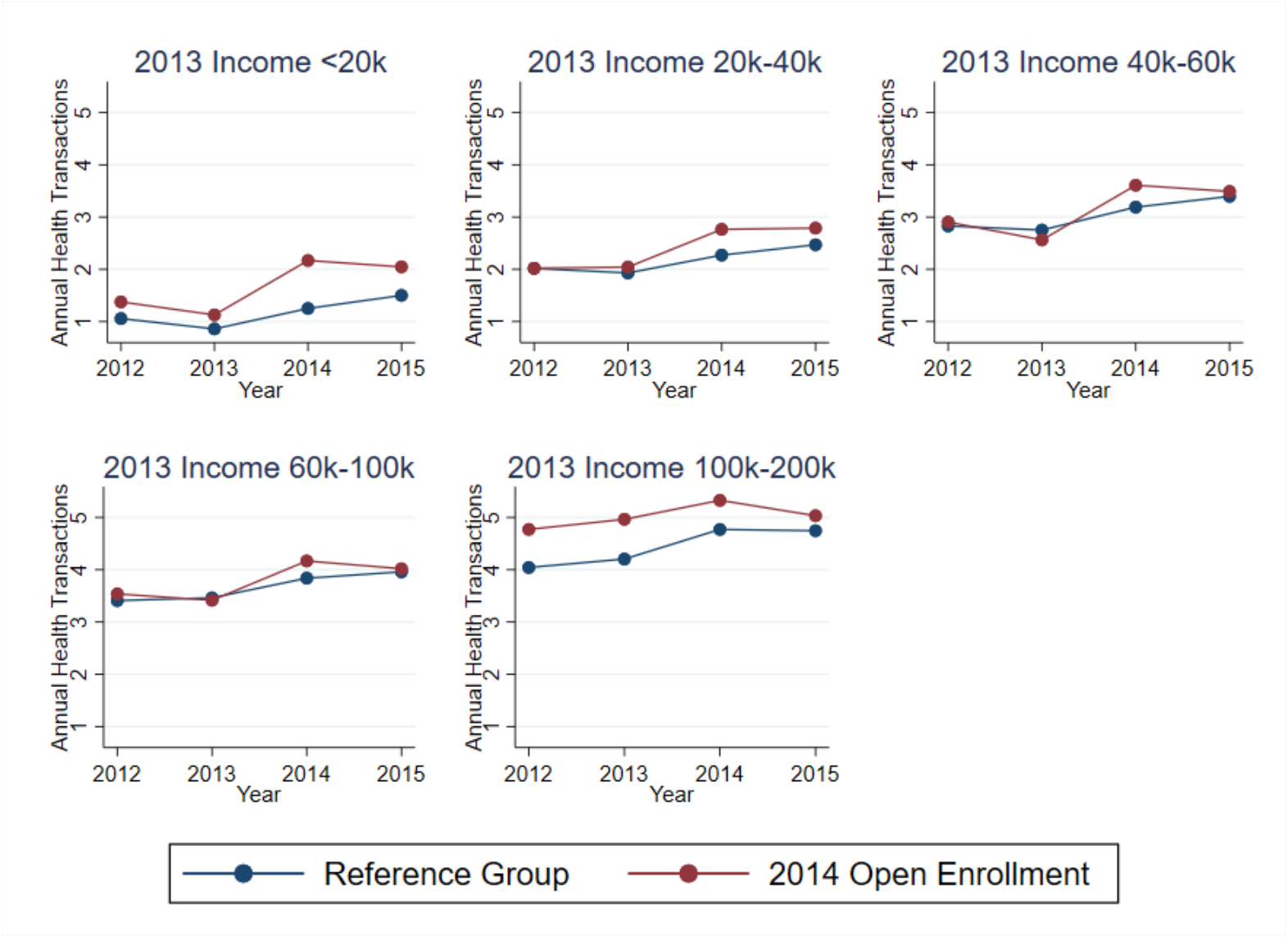
Notes: In this figure, we report the top 100 words classified by our machine learning algorithm as related to households with children. The font size indicates the frequency with which the word appears in the transactions dataset.

Figure A3: Annual Health Spending: 2014 Open Enrollees vs Reference Group



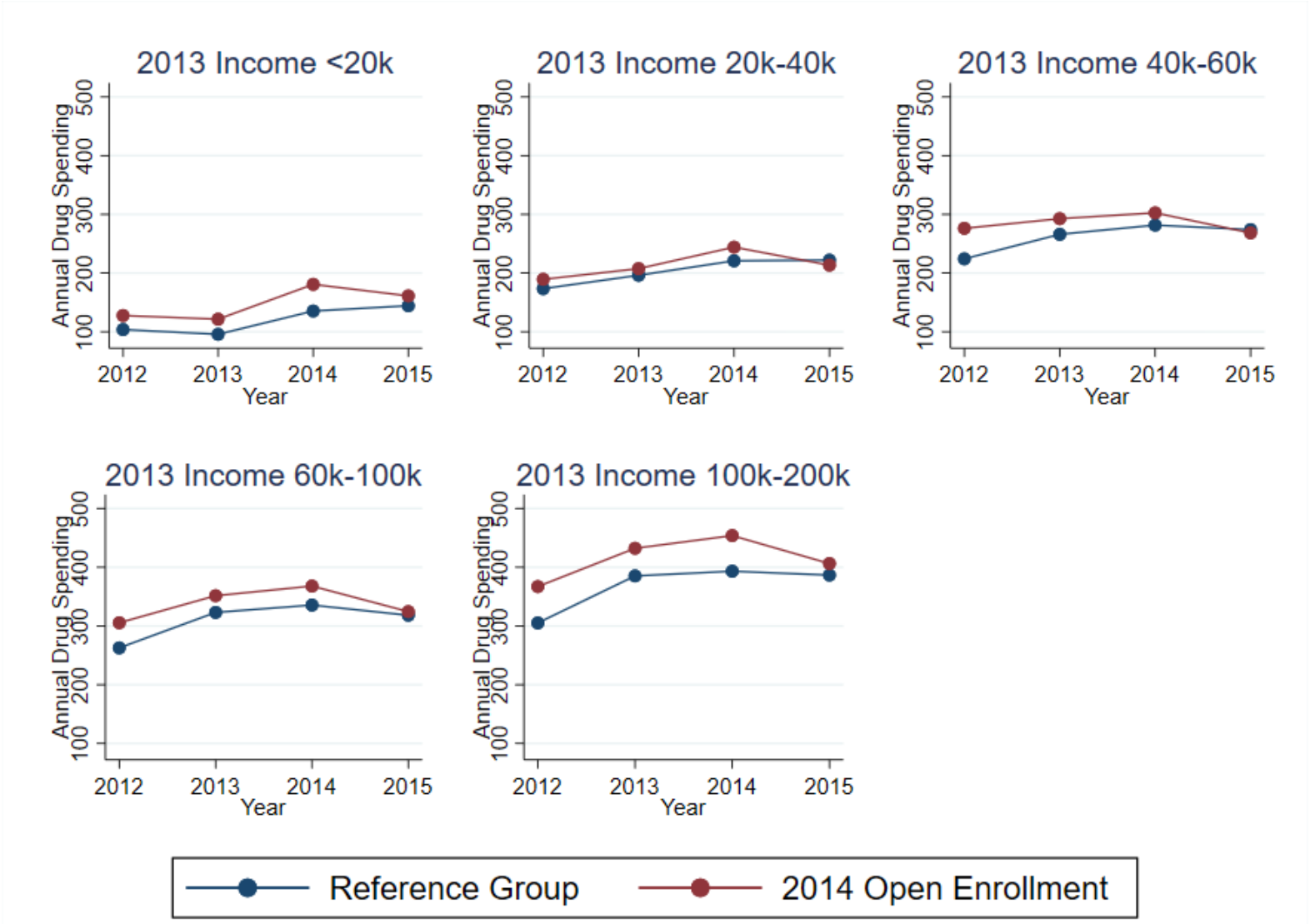
Notes: Figures show average annual health care spending by income group for 2014 open enrollees and a "reference group" of those who never purchase individual health insurance.

Figure A4: Annual Health Transactions: 2014 Open Enrollees vs Reference Group



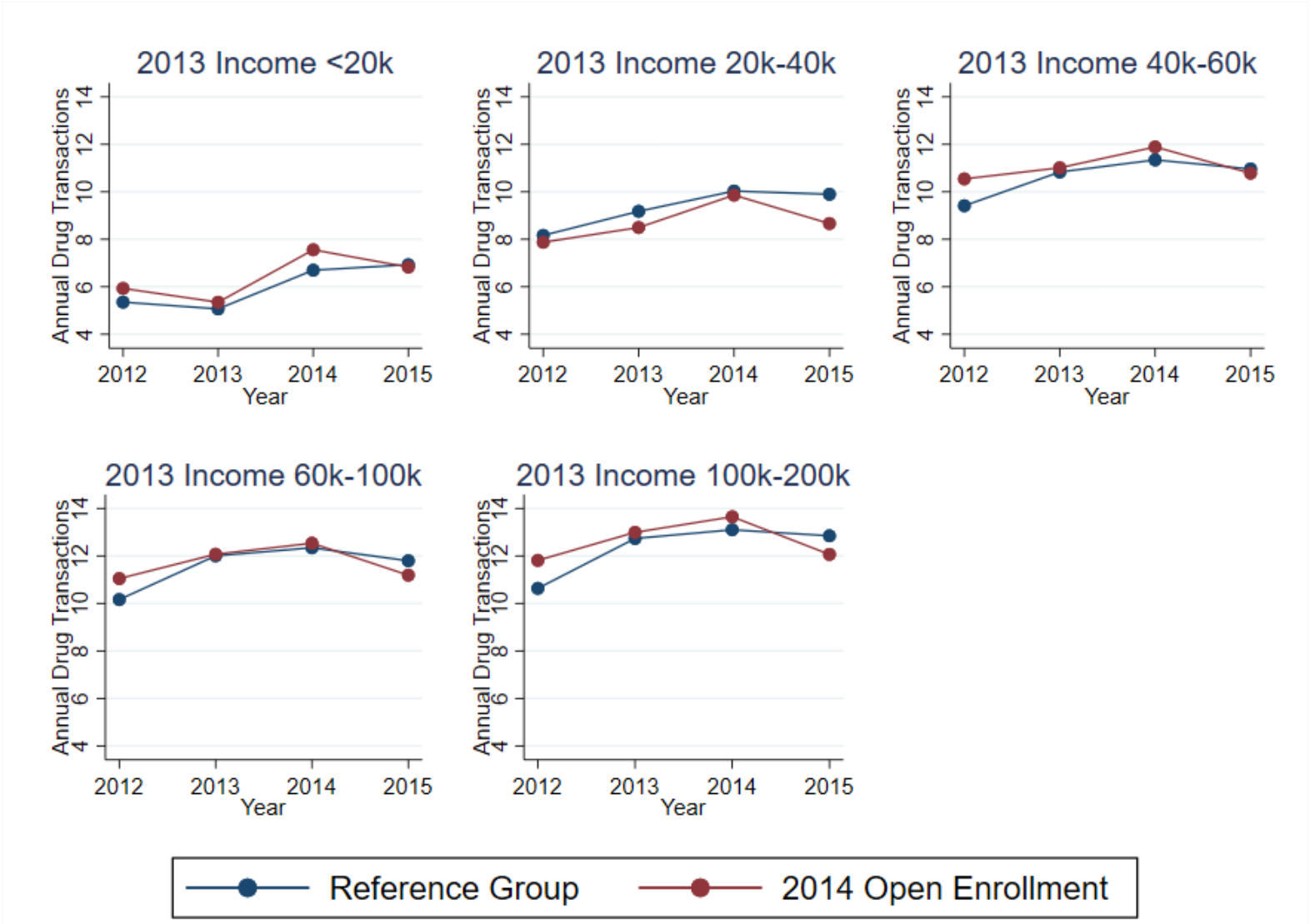
Notes: Figures show average annual health care transactions by income group for 2014 open enrollees and a "reference group" of those who never purchase individual health insurance.

Figure A5: Annual Drug Spending: 2014 Open Enrollees vs Reference Group



Notes: Figures show average annual drug spending by income group for 2014 open enrollees and a "reference group" of those who never purchase individual health insurance.

Figure A6: Annual Drug Transactions: 2014 Open Enrollees vs Reference Group



Notes: Figures show average annual drug transactions by income group for 2014 open enrollees and a "reference group" of those who never purchase individual health insurance.

Table A1: **Standard Benefit Design by Metal Tier, 2014**

Coverage Category	Bronze	Silver	Gold	Platinum
Percent of cost coverage	Covers **60**%	Covers **70**%	Covers **80**%	Covers **90**%
Preventive Care Copay	No cost	No cost	No cost	No cost
Primary Care Visit Copay	$60 for 3 visits	$45	$30	$20
Specialty Care Visit Copay	$70	$65	$50	$40
Urgent Care Visit Copay	$120	$90	$60	$40
Emergency Room Copay	$300	$250	$250	$150
Lab Testing Copay	30%	$45	$30	$20
X-Ray Copay	30%	$65	$50	$40
Generic Medicine Copay	$19 or less	$19 or less	$19 or less	$5 or less
Annual Out-of-Pocket, Maximum Individual and Family	$6,350 individual and $12,700 family	$6,350 individual and $12,700 family	$6,350 individual and $12,700 family	$4,000 individual and $8,000 family

Notes: This table reports the standard benefit design required for plans offered in the Covered California insurance marketplace in 2014, by metal tier. The percent of cost coverage represents a share of average annual cost. In most situations, the free preventative care copay holds only for one visit per year.

Table A2: **Standard Benefit Design by Metal Tier, 2015**

Coverage Category	Bronze	Silver	Gold	Platinum
Percent of cost coverage	Covers **60**%	Covers **70**%	Covers **80**%	Covers **90**%
Annual Wellness Exam	$0	$0	$0	$0
Primary Care Visit	$60	$45	$30	$20
Specialist Visit	$70	$65	$50	$40
Emergency Room	$300	$250	$250	$150
Laboratory Tests	30%	$45	$30	$20
X-Ray	30%	$65	$50	$40
Imaging	30%	20%	20%	10%
Preferred Drugs	50%	$50	$50	$15
Generic Drugs	$15 or less	$15 or less	$15 or less	$5 or less
Deductible	$5,000	$2,000 medical $250 brand drugs	$0	$0
Annual Out-of-Pocket, Maximum Individual and Family	$6,250 individual $12,500 family	$6,250 individual $12,500 family	$6,250 individual $12,500 family	$4,000 individual and $8,000 family

Notes: This table reports the standard benefit design required for plans offered in the Covered California insurance marketplace in 2015, by metal tier. The percent of cost coverage represents a share of average annual cost. In most situations, the free preventative care copay holds only for one visit per year.

Table A3: Examples of Financial Transaction Records

Description	Date	Amount	Merchant	Street	City	State
MCDONALD'S FXXXXX	12/4/12	2.16	McDonald's	3737 S Soto St	Vernon	CA
KROGER #969	12/4/12	72.79	The Kroger Co.	11700 Olio Rd	Fishers	IN
WALGREENS #6355	12/4/12	2.12	Walgreens	4201 Dale Rd	Modesto	CA
WM SUPERCENTER#2883	12/4/12	116.90	Walmart	8801 Ohio Dr	Plano	TX
PARTY TIME PLAZA LIQUO	12/4/12	29.99	Party Time Plaza Liquors	5520 Lake Otis Pkwy	Anchorage	AK
THE HOME DEPOT 2550	12/4/12	36.48	The Home Depot		Gaithersburg	MD
PEOPLES PIZZA	12/5/12	6.15	Peoples Pizza	1500 Route 38	Cherry Hill	NJ
ADVANCE AUTO PARTS #91	4/11/13	5.44	Advance Auto Parts	969 N Daleville Ave	Daleville	AL
FERRIS ORTHODONTICS-LE	4/11/13	715.00	Britton and Ferris Orthodontics	1130 E Sonterra Blvd	San Antonio	TX

Notes: The rows above provide an example of the financial transaction data we observe. In addition to the columns reported, we also observe a brief description of the transaction, which provides additional text for us to use in our machine learning algorithms. We use the transaction descriptions to classify each entry as either health spending, drug spending, or other non-health spending. We use the physical address of the transaction locations to help identify the consumer's home geographic region.

Table A4: Imputation of Healthcare Charges from Health OOP in MEPS

	2013 Health Charges	2014 Health Charges
Health OOP	5.640**	4.580*
	(1.714)	(2.122)
Constant	2658.086**	4159.501**
	(996.514)	(1294.454)
Observations	201	201

Notes: * $p<0.1$, ** $p<0.05$, *** $p<0.01$. Data include MEPS respondents in the 2013-2014 longitudinal file who report ACA exchange coverage at some point from January 2014 through May 2014.

Table A5: Income Changes Around Sign-up and Drop-Out

Panel A: Change in income: Pre ACA Sign-up/Drop-out

	(1)	(2)	(3)	(4)	(5)
Controls	<20K	20K-40K	40K-60K	60K-100K	100K-200K
signup	2.195***	0.427***	0.212***	0.122***	0.0396
	(0.267)	(0.0758)	(0.0643)	(0.0421)	(0.0480)
dropout	-0.0285	-0.143*	-0.140**	-0.119***	-0.0824
	(0.306)	(0.0744)	(0.0657)	(0.0439)	(0.0520)
Pre-Signup Monthly Income	601.39	2453.48	4140.59	6473.61	11501.53
Number of observations	26063	18164	15580	19585	15565

Panel B: Change in income: Post ACA Sign-up/Drop-out

	(1)	(2)	(3)	(4)	(5)
Controls	<20K	20K-40K	40K-60K	60K-100K	100K-200K
signup	1.264***	0.276***	0.0806***	0.0593**	0.0851**
	(0.134)	(0.0377)	(0.0264)	(0.0271)	(0.0382)
dropout	0.150	-0.0277	0.0288	-0.0169	-0.0465
	(0.129)	(0.0401)	(0.0283)	(0.0278)	(0.0382)
Pre-Signup Monthly Income	669.80	2458.56	4115.61	6432.88	11362.29
Number of observations	73064	55935	37106	38308	28078

Notes: Standard errors are clustered at the individual level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$. signup is a dummy indicating the month is at or post month of sign-up. dropout is a dummy indicating the month is post drop-out. All regressions include individual fixed effects. Pre-ACA indicates households who both sign up and drop-out prior to July 2013, while post ACA indicate sign-ups that begin as of the 2014 ACA open enrollment period and drop-outs beginning in January 2014. All data is restricted to balanced panels where at least 10 months of data is available prior to sign up and after drop out. Units are measured as a percentage change, relative to average consumption amount in the ten months leading up to sign-up.

Table A6: Probability of Medi-Cal Eligibility by Income Group

Income Bin	No. of households	Share never eligible in any month	Share with >0 eligibility in every month	Average probability eligible by month
All	17,113	0.177	0.137	0.218
$20,000 or less	3,776	0.004	0.343	0.539
$20,000-$40,000	4,355	0.011	0.230	0.242
$40,000-$60,0000	3,135	0.075	0.015	0.107
$60,000-$100,000	3,297	0.370	0.000	0.059
$100,000-$200,000	2,550	0.595	0.000	0.041

Notes: We examine Medi-Cal eligibility in our dataset by income group. We define income groups using pre-ACA annual income in year 2013. For each income bin, we report the number of households in that bin, the share of those households that are never eligible for Medi-Cal in any month, the share of households that have a strictly positive probability of being eligible in all months observed, and the average probability eligible by month. The probability of eligibility is not closer to 100% in the poorest income bin because we define the bin based on 2013 income. Households that subsequently enrolled in ACA coverage likely had higher income in later periods, making the average probability eligible lower in later months.

Table A7: Words and phrases for identifying premium payments

Panel A: Insurers other than Kaiser

Insurer	**Include any**	**Exclude any**
Anthem Blue Cross of California	ANTHEM B, ANTHEMBLU	BCBS
Blue Shield of California	BLUE SH, BC OF CA, BSC HTL	CROSS
Chinese Community Health Plan	CCHP	
Health Net	HEALTH NET, HEALTHNET	TRI CARE, TRICARE, MEDCO, MEDICARE, AZ, ARIZO , FED, SEMO, OF NE, MYBALANCE
LA Care Health Plan	L.A. CARE, LA CARE HEALTH	
Molina	MOLINA HEALTH	
Sharp Health Plan	SHARP HEALTH	
Contra Costa Health Plan	CONTRA COSTA HEALTH	
Western Health Advantage	WESTERN HEALTH	

Panel B: Words and phrases in classifying Kaiser transactions

Premium	**Drug**	**Health**
DUE, HEALTH PL, HPS, DIRECR PAY, BILL PAY , ONLINE PAY,	PHAR, CPP, MAIL, RX, DOWNEY, LIVERMORE	Everything else such as PERM, PRMNT, PRNTE

Notes: We use the keywords reported above to identify all insurer premium payments in our financial transactions data in California from 2012 through 2015. We identify premiums paid to all insurance carriers participating in the Covered California marketplace.

C.1 MTurk Survey Questions

We include below the survey questions we asked MTurk workers.

There are lots of types of health care spending.

First, some visits to the doctor **are essential**: A patient with a diabetes needs to receive a treatment, else he might die. On the opposite side of the spectrum, there are visits to the doctor that are **much less essential**, say, acne removal, for example.

Further, some visits to the doctor are urgent and **cannot be postponed**: A patient with a heart attack needs to go to the ER immediately, else he might die. On the opposite side of the spectrum, there are visits to the doctor that **can be substantially postponed,** like a regular check-up.

Some of the words to classify may not be related to health care; however, a majority of the words should be related.

WORD TO CATEGORIZE:

"sample health word"

1 How much do you associate this word with **health care that is essential**?

Not essential — **Extremely essential**

1	2	3	4	5	6	7	8	9	10

2 How much do you associate this word with **health care that can be postponed**?

Not urgent — **Extremely urgent**

1	2	3	4	5	6	7	8	9	10

3 If you believe this word is **irrelevant to health care spending**, check the following box:

- o **irrelevant (non-related to health care)**
- o relevant (related to health care)