NBER WORKING PAPER SERIES

# TAKE-UP, DROP-OUT, AND SPENDING IN ACA MARKETPLACES

Rebecca Diamond
Michael J. Dickstein
Timothy McQuade
Petra Persson

Working Paper 24668
http://www.nber.org/papers/w24668

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2018

We would like to thank Anais Galdin, Ziao Ju, Jiayu Lou, Lucienne Oyer, and Ryan Shyu for outstanding research assistance. We would also like to thank the seminar participants at the 2017 AEA Annual Meeting, NYU's Friday Micro Workshop, and Stanford's Opportunity Lab Lunch for helpful comments and suggestions. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Take-Up, Drop-Out, and Spending in ACA Marketplaces
Rebecca Diamond, Michael J. Dickstein, Timothy McQuade, and Petra Persson
NBER Working Paper No. 24668
May 2018
JEL No. H4,H51,I13,L1

**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 more than half of all new enrollees dropping coverage before the end of the plan year. Enrollees who drop out re-time health spending to the months of insurance coverage. This drop-out behavior generates a new type of adverse selection: insurers face high costs relative to the premiums collected when they enroll strategic consumers. We show that the pattern of attrition undermines market stability and can drive insurers to exit, 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 efficient when they jointly target take-up and attrition.

Rebecca Diamond
Graduate School of Business
Stanford University
655 Knight Way
Stanford, CA 94305
and NBER
diamondr@stanford.edu

Michael J. Dickstein
New York University
Stern School of Business
Kaufman Management Center,7-78
44 West 4th Street
New York, NY 10012
and NBER
michael.dickstein@nyu.edu

Timothy McQuade
Stanford Graduate School of Business School
655 Knight Way
Stanford, CA 94305
tmcquade@stanford.edu

Petra Persson
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
perssonp@stanford.edu

# 1 Introduction

Social insurance programs are large, government-provided 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 that the coverage dimension is highly relevant: take-up of social insurance is often low, and a sizable literature has documented various factors driving low take-up, ranging from stigma and a lack of information to 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 take-up.

This academic and policy focus on increasing take-up, however, misses a crucial ingredient to the success of social insurance: individuals who take up insurance must *stay* enrolled for the program to be effective. 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 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.

[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 TV shows.

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. Using a theoretical framework, we show that this drop-out behavior undermines the stability of the market for individual insurance and has the potential to cause market unraveling. 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 individual 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 (CC) 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

[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 CC.

health care spending and out-of-pocket drug spending. 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]

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 and the dollar value of out-of-pocket health care and drug spending increases. 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 suggests that, once covered by insurance, ACA enrollees increase their health transactions by 51.4% and their out-of-pocket health spending by 45.0%.[7] These increases are concentrated among households that are poor (but not poor enough to qualify for Medicaid); in fact, our estimated effects fall monotonically with household income. The increases in health spending are especially pronounced among households with young children.

While new enrollees 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.

To gauge the drivers of this drop-out behavior, we first assess whether affordability explains the pattern we observe. Specifically, we analyze whether households experience income losses 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

[6]According to the National Health Interview Survey published by the Centers for Disease Control, 20.4% of U.S. households were uninsured in 2013.

[7]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 simply ensures that the changes to ACA enrollees' health care consumption do not simply reflect time trends in the overall population.

post-ACA. This change suggests the ACA lessens the need for households that suffer temporary economic hardship to drop individual market coverage. We also conduct text analysis by using Mechanical Turk workers to score the words we observe in the transaction descriptions based on how essential and/or urgent they perceive each health care description to be. 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. Our analysis suggests that households that ultimately drop coverage strategically 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 illustrate how the presence of drop-out enrollees affects market stability. We employ a standard model of adverse selection, as in Einav, Finkelstein, and Cullen (2010), with the innovation that our framework allows for two types of enrollees: "follow-the-rules" types and "drop-outs." Follow-the-rules types consume health care and pay premiums over the entire year. Drop-outs, in contrast, are able to re-time consumption of health care across the plan year and drop their insurance coverage before the end of the year by discontinuing premium payments. Intuitively, drop-outs are costly to the insurer, as they, in equilibrium, display a high consumption of health care for a subset of the year, and then cease to pay premiums. The model illustrates that the presence of drop-outs can generate adverse selection and potential market unraveling. This selection problem arises even in a health insurance market in which all consumers are *ex ante* identical in their expected annual health care expenditures; that is, we find adverse selection even when we eliminate the traditional adverse selection channel.

Our description of consumers as multi-dimensional– with both a risk type and drop-out type– 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. We then consider alternative penalty designs and discuss how changes in the penalties for failing to maintain coverage might combat possible unraveling when drop-out types represent a high proportion of the initial enrollment pool. In this examination, our calculation is similar to Kowalski (2014), who examines how the creation of individual insurance marketplaces affects adverse selection and markups state-by-state, using 2014 data. We focus on the role of dropouts in generating selection, and relate the penalty size to the premium savings from drop-out.

Our paper contributes to an emerging literature on the consequences of the ACA.[8] Antwi, Moriya, and 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, the insight that drop-out behavior can generate market-wide unraveling and undermine the stability of the insurance market is related to the literature on adverse selection 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 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, consumers' strategically rush their treatments, 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 potential market unraveling. Section 6 examines the pass-through of the costs associated with the presence of drop-out types through premium increases, and Section

[8]Outside of the literature on the ACA but also related is work by Finkelstein, Hendren, and Shepard (2017), who analyze low-income households' willingness to pay for health insurance using administrative data from Massachusetts' subsidized insurance exchange. They do not document drop-out behavior, however.

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 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.[9] 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.[10] 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.[11] Fourth, the ACA required employers

[9] 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.

[10] 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.

[11] 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).

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.[12] 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.[13] 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).[14]

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.[15] Premium subsidies, which depend on both the individual's income

[12] 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.

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

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

[15] 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

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

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.

this financial institution, we 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 1 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.[16] 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.[17] 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.[18] 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.

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

[17] 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.

[18] 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.

### 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.[19] For each insurer, we manually searched for premium payments from both bank and credit card accounts based on certain keywords or phrases appearing in the description of the transactions. We were able to identify all of the participating insurers in the data.[20] Based on the dates of the first insurance payment, we identify those consumers who select an insurance plan during open enrollment.

Figure 3 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.[21] 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.[22] 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

[19] 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.

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

[21] 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.

[22] 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.

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.[23]

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 4 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. [24] 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,

[23] 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 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.

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

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 kids. This value is much lower than the number reported by the Census Bureau because our definition of kids is mainly driven by purchases indicative of young children. As shown in Figure 5, the most popular merchant names such as Gymboree and Carter's are retailers of baby clothing and toddler essentials.

### 3.5 Measuring the types of health care consumed

After classifying transactions into health care, we extracted all words observed in the health care transactions. Our sample of health-related words contained 1,522 different root words. Since we will study how dropouts and long-term enrollees differ in their health care consumption, we also want to measure 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.

Defining which health care transactions are more discretionary or urgent requires an outside data source to measure these characteristics. 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 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 define 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, some visits to the doctor patients can postpone, like an annual check-up. Similarly, some visits to the doctor are essential: a patient with diabetes needs to receive treatment on a regular basis. 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 3 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.

### 3.6 Summary statistics

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.[25] Our sample has 13.8 million transactions associated with 289,481 households in California. For 104,233 households, we observe at least one premium transaction in the individual health insurance market.

For the 104,233 members in the individual insurance market, we define *sign-up* to be the

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

month we observe the consumer's first premium payment and *drop-out* to be the month of her last premium payment. Figure 6a 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 enrollment groups. Figure 6b 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.[26] Our analysis will focus on the behavior of the new enrollees.

Figure 7 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.

[26]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.

# 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 Total Health care Spending of ACA Enrollees

We begin our analysis by studying the extent to which the take-up of insurance through the 2014 open enrollment on the ACA exchanges affects total enrollee health care spending. 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. 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, $\alpha_i$ denotes household fixed effects capturing time-invariant characteristics of a household that affect health care spending, such as chronic health conditions. Our coefficient of interest is $\gamma$, which measures the impact of enrolling in an ACA marketplace plan on health care spending. We estimate our regressions separately by income, with individuals 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 individual 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 focus specifically on the 2014 open enrollment group since the decision to sign up at this point is mostly driven by the policy change in the health insurance markets. 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 report our results for general health spending and drug spending in Table 4. We have scaled our estimates of $\gamma$ 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 simple differences estimates suggest that ACA enrollees with income of less than \$20K increase their number of health transactions by a highly significant 70.8%. The impact of ACA enrollment then declines monotonically with income. Those with income between \$20K and \$40K increase their number of health transactions by 36.5%, those between \$40K and \$60K by 30.9%, and those between \$60 and \$100K by 18.1%. The estimated effect is small and statistically indistinguishable from zero for households with greater than \$100K of income. These effects remain when we move to the full differences-in-differences specification, indicating that the results are not driven solely by time trends. Using the full specification, we find increases of 51.4%, 27.0%, 19.3%, 13.4%, and 0.0% relative to the reference group as we move from the lowest income category to the highest.

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.[27]

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.

### 4.2 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 8 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

[27] We illustrate the levels of annual health and drug spending and transactions underlying these difference-in-difference estimates in Appendix Figures A2,A2, A3, and A4.

figure illustrates a striking fact: ACA enrollees frequently drop coverage across all income levels, 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 similarly 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 9 examines the distribution of drop-outs across total months of coverage. We first see that 15% of all 2014 open enrollment consumers stop paying their premiums after only one month. The remaining households, 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 for the 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 10 reports the six month dropout 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 market instability may grow in importance as well.[28]

### 4.3 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. One natural candidate is income

[28] 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.

shocks. Premiums are likely to be more affordable when households receive positive income shocks, and less affordable when they receive negative 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.[29] In Figure 11, the solid line shows the monthly income dynamics of consumers prior to, during, and after insurance coverage.[30] To harmonize the event study timing across 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. 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, at 50%, when compared to the period six months prior to enrollment. In addition, income falls 20% over the three months after drop-out. Thus, affordability appeared to play an important role in the decision to enroll and drop out of individual health insurance prior to the ACA.

In contrast, the dashed line in Figure 11 plots the income dynamics around sign-up and drop-out of ACA enrollees.[31] ACA enrollees exhibit a 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 drop in income after drop-out. In Appendix Table A4 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 and we never find income declines around drop-out in the post-ACA period. Overall, the ACA appears to have significantly lessened the role of affordability in the decision to obtain and maintain health insurance coverage.

To the extent that income shocks are *not* driving the drop-out behavior of ACA enrollees, we explore an alternative hypothesis: consumers strategically 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. Indeed, the regulatory structure of the ACA

[29] 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 dropout at the end of 2013. We also focus on drop-outs that do not occur in November or December, as ceasing to pay for insurance in these months would be an allowed drop during open enrollment.

[30] We restrict our analysis to households with less than $60,000 of annualized average income in the 10 months leading up to sign-up. This allows the income changes to not be dominated by the highest income earnings. This is needed since our estimates are done in levels, not logs to allows months with zero income. We study all income groups in income group specific analysis in the appendix.

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

may incentivize exactly this 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. There is 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 spending 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. Panel A in Table 6 reports the percentage change in health spending among pre-ACA enrollees. Apart from in the lowest income group, we find no statistically significant change in health spending during enrollment or after drop-out. In contrast, Panel B shows that enrollees under 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. The highest income group show smaller point estimates, which cannot be statistically distinguished from zero. We observe similar patterns when looking at transaction counts in Panels C and D in Table 6, where results are more precisely estimated.

The estimates are noisier when studying prescription drug usage in Table 7, 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. This evidence 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.

### 4.4 How Dropouts use the ACA

To provide further support for our hypothesis that 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 3.5 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 8 contains 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 8 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 8 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, potential enrollees all share the same expected spending over the course of a year. That is, we rule out the traditional channel for adverse selection, in which consumers of higher cost have higher willingness to pay for insurance. Instead, consumers in our setting differ in their ability to re-time health consumption towards the beginning of the plan year, which – as we will see – will translate into differential drop-out rates and differential willingness to pay for insurance.

## 5.1 Model set-up

The model set-up below follows the simplest case in Einav, Finkelstein, and Cullen (2010), adding in the potential for drop-out in each month.

A consumer faces two possible coverage options. The first option is to enroll in a single, standardized insurance plan once a year, with coverage beginning on January 1 of the plan year.[32] Once consumers enroll in a plan, they may choose each month whether to continue paying for insurance or lapse and lose coverage for the remaining months of the year. The second option is to remain uninsured, which entails no premium payment but may involve a penalty paid at the end of the year.

We label the two options "with insurance" (W) and "with no insurance" (N). A consumer's utility over these two options depends on her expected health costs over the year and the share of those costs that can be re-timed. Intuitively, potential enrollees who can carry out all of their health spending in a short time horizon will have an incentive to drop coverage before the end of the plan year. To highlight the drop-out motivation, we assume that the total expected health costs over the year are equal for all consumers. What differs between consumers is the share $\theta_i$ of these costs that may be re-timed. We assume a distribution over this share, $G(\theta_i)$. In the population, the share $\theta_i$ may differ across $i$ depending upon buyer characteristics, including whether the individual chooses a plan with coverage for dependents and whether the buyer is new to insurance and the health care system.

In this environment, we define utility for the two coverage choices:

$$\text{Utility with insurance: } U^W(\theta_i, p)$$
$$\text{Utility with no insurance: } U^N(\theta_i, 0)$$

Here, $p$ denotes the insurance premium.[33] Consumers will buy coverage if $U^W(\theta_i, p) > U^N(\theta_i, 0)$. As in Einav, Finkelstein, and Cullen (2010), we let $\pi(\theta_i) = max\{p : U^W(\theta_i, p) \geq U^N(\theta_i, 0)\}$ be the highest price at which an individual is willing to buy insurance. Then, demand is:

$$D(p) = \int 1\{\pi(\theta_i) \geq p\} dG(\theta) \tag{2}$$

The average costs and marginal costs of the population who choose coverage are:

$$AC(p) = \frac{1}{D(p)} \int costs(\theta_i) 1\{\pi(\theta_i) \geq p\} dG(\theta) \tag{3}$$

$$MC(p) = E(costs(\theta_i) | \pi(\theta_i) = p) \tag{4}$$

[32] The assumption that the premium is fixed for a year matches our empirical setting. We abstract in our model from the possibility of mid-year enrollment, which can occur in the insurance marketplaces when enrollees fall into one of several special categories, including, for example, the sudden loss of employer coverage.

[33] In the presence of a penalty for choosing to be uninsured, $p$ denotes the incremental premium relative to the penalty.

In our graphical illustration below, we do not take a stance on the nature of competition between insurers and therefore leave open more than one possible equilibrium outcome. However, if we assume Bertrand competition, we can find a unique equilibrium in which firms set price equal to average cost, $AC(p)$. In this equilibrium, all consumers whose willingness to pay exceeds the average cost, $AC(p)$, choose to enroll. Uniqueness is guaranteed as long as it is profitable to provide insurance to those consumers with the highest willingness to pay, and if the marginal cost curve crosses demand at most once ( Einav, Finkelstein, and Cullen (2010)).

To illustrate the possible equilibrium outcomes in this market, including the possibility of full unraveling, we plot demand and average costs. To simplify the illustration, we assume that $\theta_i$ takes one of two values: $\theta_L$ and $\theta_H$. Consumers with $\theta = \theta_L$ have a low share of expenditures that are re-timable, and so have less incentive to drop coverage. We label these the "follow-the-rules" type; because a large share of their expenditure reflects urgent care or other chronic expenses that occur throughout the year, these consumers will not be as willing to drop coverage. Families with children or consumers with longer term care needs are likely to fall into this population.[34]

Consumers with $\theta = \theta_H$ have a higher share of expenditures that are re-timable. We label these consumers the "drop-out" type, because they have an incentive to re-time their care into the first months of the year, and then drop coverage. Because they are less likely to remain in care for multiple months, these consumers will also exhibit relatively inelastic demand for insurance that is priced at an annual rate to the average consumer. Drop-out type consumers will be those whose expenditure represents, for example, annual specialty visits that can be re-timed to a specific month during the year. A relatively smaller share of their expenditure goes toward urgent care needs.

### 5.2 Graphical Illustration

We present equilibrium outcomes when (a) the insurer can distinguish potential enrollees by type $\theta = \{\theta_L, \theta_H\}$ and (b) the insurer can only set a single premium priced to the average consumer. The equilibrium outcomes differ depending on the relative proportions of the two types in the population. As we will see, full unraveling can occur in this market with sufficiently few "follow-the-rules" types – i.e. a sufficiently small share of $\theta = \theta_L$ – in the

[34] For simplicity in this conceptual framework, we have ruled out differences in ex ante risk between consumers of different types, $\theta_i$. We could incorporate heterogeneity in both $\theta_i$ and underlying costs. In our setting, it is likely that those consumers with $\theta = \theta_L$ would also have higher ex ante risks or higher costs of coverage.

population.

To provide an illustration, we consider a linear functional form for demand and a particular cost function. An insurer's monthly revenue from insuring individual $i$, during the months of coverage chosen by $i$, is given by

$$revenue(\theta_i) = \frac{p}{12} - (12\theta_i) * \alpha c_i,$$

where $c_i$ represents individual $i$'s underlying costs per month and $\alpha$ represents the fraction of costs that the insurer covers. Again, we assume that $c_i$ is the same across both consumer types. The $12 * \theta_i$ reflects that consumers with a high ability to re-time can spend more of their annual costs, $12 * \alpha c_i$, in the first months of coverage. In particular, $\theta_i$ represents the share of annual expenditures that can be re-timed to one month, which implies that individual $i$ will choose to stay covered for $1/\theta_i$ months. Consumers pay 1/12 of annual premium dollars, $p$, each month.

We illustrate a case in which the drop-out enrollee incurs all 12 months of health care costs in one month and thus pays only one month of premium, whereas the follow-the-rules enrollee does no re-timing and pays 12 months of premiums. That is, $\theta_L = 1/12$ and $\theta_H = 1$. At the same level of underlying health costs, the drop-out types will be far less responsive to price, since consumers who drop out pay only one month of premiums. We assume linear demand functions for consumers, in member months.[35]

We plot these example demand functions and average cost functions for both the "drop-out" and "follow-the-rules" type consumers in Figure 12. The demand functions in Figure 12a illustrate that the drop-out consumer will change his quantity of coverage demanded, in member months, very little in response to changes in premiums, and much less so relative to the follow-the-rules consumer. The average costs in Figure 12b illustrate how the cost per enrollee to the insurer changes as the premium changes. With higher premiums, the quantity demanded by drop-out types falls less than the quantity demanded by follow-the-rules types, leading to higher average costs. Average costs equal the drop-out type cost at sufficiently high premiums (above $p = \frac{1}{3}$ in our example), when only drop-out types choose to enroll.

In Figure 13, we plot average costs and market demand against premiums, in a form more commonly seen in the literature on selection markets. Figure 13a shows evidence of adverse selection: as premiums rise, the average cost in the enrolled population increases. We observe a range of premiums for which demand exceeds average costs, and thus insurers will find it

[35] In our example, we assume the drop-out type has demand $Q^H = 1 - p$, while the follow-the-rules type has demand $Q^L = 4 - 12p$. We further set $c_i = 15$ and assume the fraction of costs the insurer covers, $\alpha$, equals 0.8.

profitable to enter the market. At those levels of premium, both follow-the-rules types and drop-out types pool in the market. In Figure 13b, we show how the share of drop-out types enrolling varies with changes in the premium level.

Finally, we illustrate unraveling as a function of the share of drop-out types in the potential enrollee pool. Using our example, in Figure 14 we plot the demand and average cost curves assuming two different shares of drop-out types among potential enrollees – 10% and 20%, respectively. With the drop-out type representing 20% of the potential enrollees in Figure 14b, we continue to find adverse selection, with a downward sloping average cost function. However, the demand curve always lies below the average cost curve, suggesting full unraveling: no insurer would enter the market at any premium level. In this setting, firms that enter would attract an enrollment pool with costs exceeding the premium dollars earned, once we account for the level of drop-outs in that enrollment pool. In our later discussion in Section 7, we consider how changes in the tax penalties for dropping insurance coverage might combat possible unraveling when drop-out consumers represent a large share of potential marketplace enrollees.

# 6 Empirical Evidence on Prices

We now explore the empirical relationships 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 choice 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.[36]

To examine the effect of drop-out rates on prices, we look at the association between observed

[36] 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.

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:

$$\Delta \ln p_{s,t+1,c} = \alpha \, cheap1_{s,t,c} + \beta \, cheap2_{s,t,c} + \theta dropout\_rate_{s,t,c} + \delta_t + \delta_s + \delta c + \varepsilon_{s,t,c} \quad (5)$$

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 influences the level of the government's premium subsidies in the area. Finally, we include $\delta_t, \delta_s$, and $\delta c$, which are year, region, and insurer fixed effects, respectively.

Column 1 of Table 9 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 9 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 9 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.

The model presented in the previous section showed how higher drop-out rates lead to higher average costs, which could lead to higher prices. The results in Columns 1-3 of Table 9 suggest that, in response to high drop-out rates, insurers may lower prices to try to increase demand from the pool of consumers who do not drop out, because these consumers are more price sensitive. The results could also reflect 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. 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 makes a clear prediction: 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:

$$\begin{aligned} ln\ (dropout\_rate_{s,t,c}) \ &= \alpha\ cheap1_{s,t,c} + \ \beta\ cheap2_{s,t,c} + \\ &\theta\ ln\ (price_{s,t,c}) + \ \delta_{t,s} + \ \delta_{s,c} + \varepsilon_{s,t,c}. \end{aligned} \tag{6}$$

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 9 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 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 deciding whether to drop coverage, enrollees trade off the savings in premium coverage with the penalty costs imposed under the insurance mandate. In this section, we discuss changes in the tax penalty levied on consumers that might encourage adherence to the mandate.

## 7.1 Penalty Design

For years 2014 and 2015, the federal government determined tax penalties for lack of insurance coverage in two different ways: as a percentage of household income and per person. Those consumers who fail to show qualified coverage must pay whichever penalty is higher. In 2014, the per-person annual penalty equaled $95 per adult with an additional $47.50 per child under 18. The penalty as a percentage of income used the basis of 1% of household income above a threshold, roughly $20,000 for a couple. In total, the maximum amount this penalty could reach as a percentage of income equaled the national average premium of a bronze plan sold through the marketplace. In 2015 and 2016, both the per person and income-based penalties rose, with the income share rising to 2.5% by 2016.

Consumers could also qualify for an exemption from the requirement to maintain minimum essential coverage and thus avoid the tax penalty. The federal government grants these exemptions in the case of financial hardship, such as a recent eviction, when premiums are not affordable in a geographic region, and under several other circumstances.

## 7.2 Back of the envelope calculation of penalty strength

To assess whether the penalty design effectively discourages consumers from dropping coverage, we compute estimates of the average penalty and the sum of premiums a consumer must pay if she drops insurance coverage after exactly $n$ months. We estimate monthly premiums as a function of household size and annual income. We estimate a household's net share of premium costs after federal subsidies using the income qualifications established in the ACA.[37]

We illustrate our results in Figure 15. We provide estimates for a household whose annual income is $40,000. We take two baseline families: (1) one household of two people, one adult and one child, and (2) one household of four people, two adults and two children. All penalty and premium estimates are computed over 12 months, but penalties always equal zero in months 11 and 12 due to the "short gap" exemption of two months.[38] In the figure, we illustrate cumulative penalties and premiums if household were to drop coverage at each month in the year. We define savings as the cumulative difference between the monthly premium and the penalty at each month, compared to the amount an individual would pay

[37] We use the 2013 federal poverty line to determine the 2014 estimated costs, and the 2014 federal poverty line to determine the 2015 estimated costs.

[38] If a consumer drops coverage after 10 months, she is exempted from penalties (See IRS Government (2017a)).

if she never dropped coverage.

Figure 15 illustrates that, in terms of premiums and penalties, dropping out before the last month of the tax year always yields savings relative to staying enrolled for a full year. The amount of the savings is due to the fact that penalties are small relative to the typical premium levels. We find this level of savings consistently across years, despite the gradual increase in penalties observed in 2015 and 2016. However, the penalty increases, particularly in 2016 for larger households, did substantially reduce the profitability of drop-out. Comparing the upper and lower panels of the figure shows that the smallest households benefit more from dropping out early.[39] Thus, in sum, Figure 15 illustrates that larger penalties can mitigate drop-out – and, conversely, a reduction (or elimination) of penalties may encourage consumers to drop coverage.

More broadly, there are several other measures, beyond explicit financial penalties, that may lower attrition and thereby enhance the stability of the individual market. First, pre-payment–that is, making individuals pay a larger share of their annual premium upon sign-up–could mitigate drop-out. Such a measure, however, might discourage sign-up among liquidity-constrained individuals.[40] Second, higher deductibles would mitigate drop-out through a similar channel. Intuitively, with a higher deductible, an individual who signs up and strategically consumes a large amount of care in the early months of coverage would bear more of those costs upfront. Thus, both the first and second measures *de facto* front-load more of the costs associated with health care, reducing the profitability of strategic re-timing of expenses to the beginning of the plan year. Third, any restrictions placed on the possibility to re-enroll post drop-out would reduce the incentive to drop coverage mid-year.

# 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.

[39] In addition, were we to examine families of different income levels, we would observe low income families having both smaller penalties and smaller premium savings. The smaller premium savings arise due to the larger premium subsidies these poorer households receive in a geographic market.

[40] The poorest consumers, who may be most likely to be liquidity constrained, would be partly shielded from large up-front costs because these same consumers receive the largest premium subsidies.

We document a sharp increase in new enrollees (new premium payments) in the end of 2013, corresponding to the start 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 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 problem and potential market unraveling, 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.

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

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.

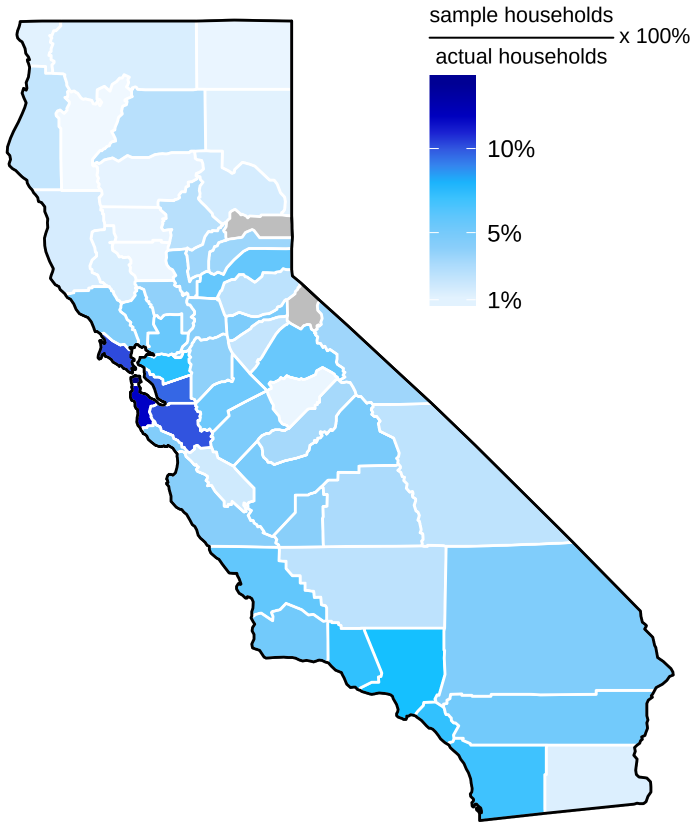
Tebaldi, Pietro (2017). "Estimating Equilibrium in Health Insurance Exchanges: Price Competition and Subsidy Design under the ACA". Becker Friedman Institute for Research in Economics Working paper No 2017-05.

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*.

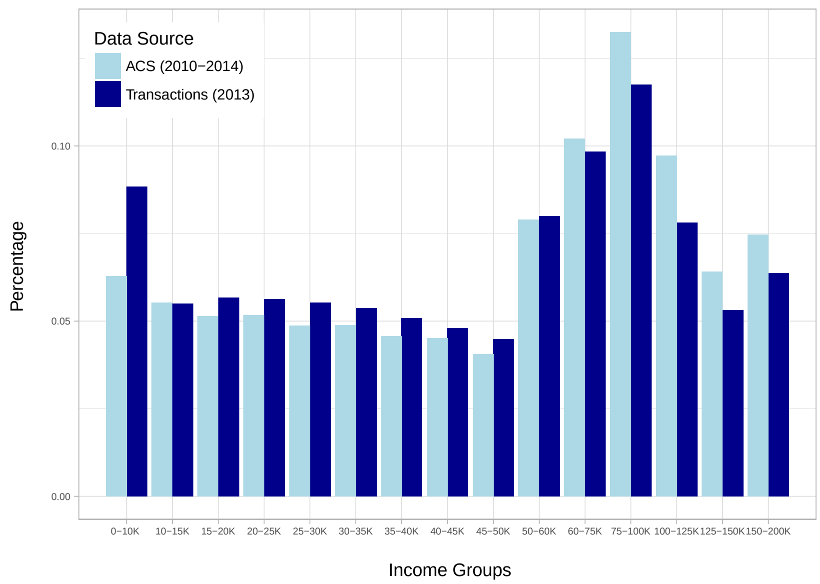
Figure 1: Geographic Distribution of Financial Account Members in California



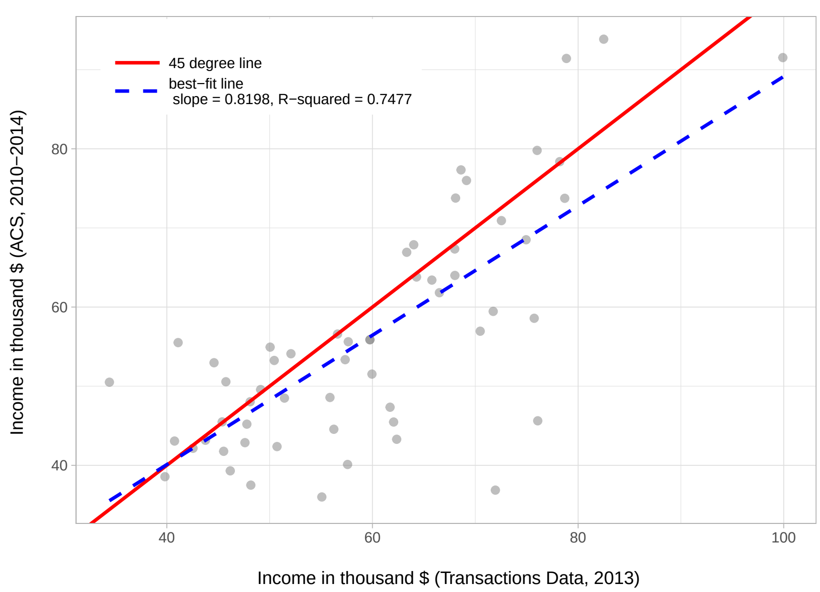

**Notes:** The map illustrates the geographic distribution of the financial account members in our data relative to the actual distribution of households in California. Our data contains approximately 850,000 active users, which corresponds to 6.7% of the California population. We use data available 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: Income Representativeness

(a) Individual level



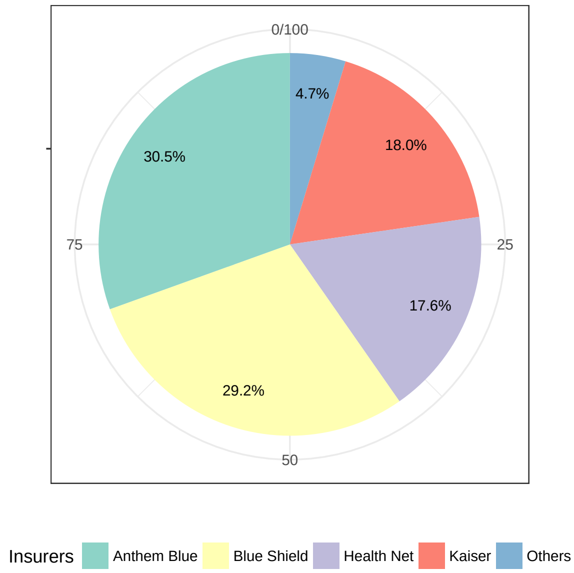

(b) County level



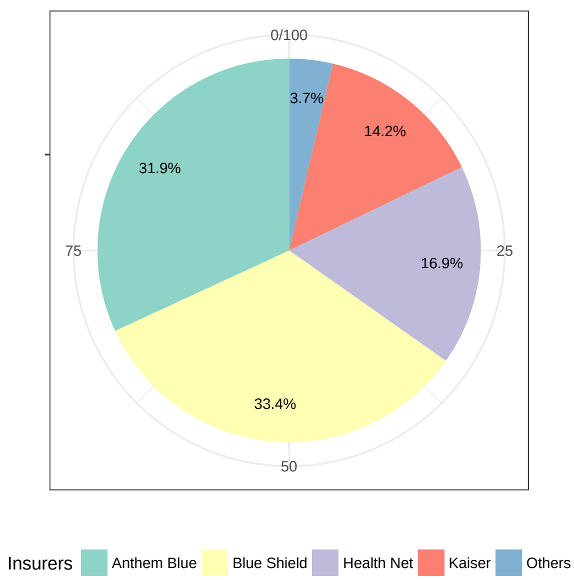

**Notes:** Panel (a) compares the 2014 income distribution of financial account members to the 2014 California household income distribution as reported by the American Community Survey (ACS). As shown in the graph, consumers in our data have incomes with a distribution slightly skewed to the left, with 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.

Figure 3: 2014 Market Shares by Insurance Carrier

(a) As reported by Covered California




(b) As recovered in our transactions data




**Notes:** Panel (a) shows the market share by insurance carrier participating in the Covered California marketplace in 2014. Panel (b) shows the market share by insurance carrier as recovered in our transaction data. As shown in the figure, the two major carriers, Anthem Blue Cross of California and Blue Shield of California, each take approximately one third of the market; the remaining 10 insurance carriers cover the balance of marketplace enrollees. 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.

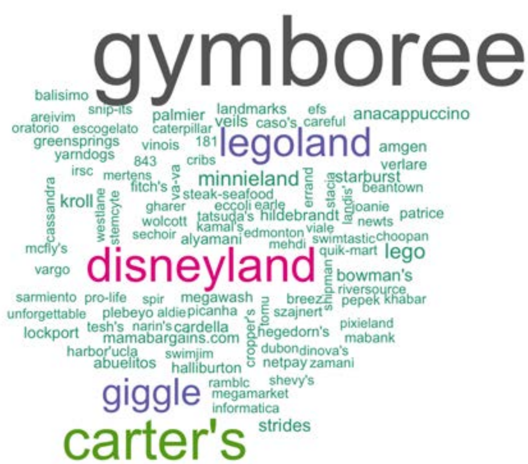
Figure 4: Word clouds for the machine learning predictions, health and drug spending

(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 5: Word clouds for the machine learning prediction, families with children

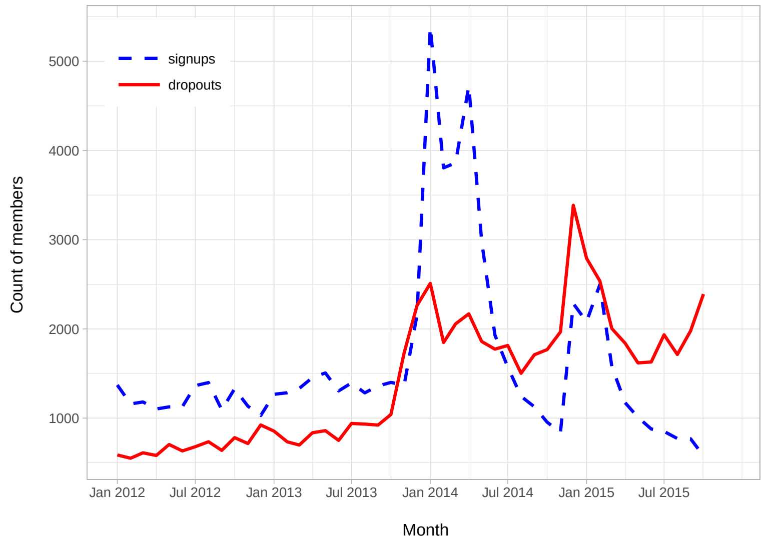
(a) children



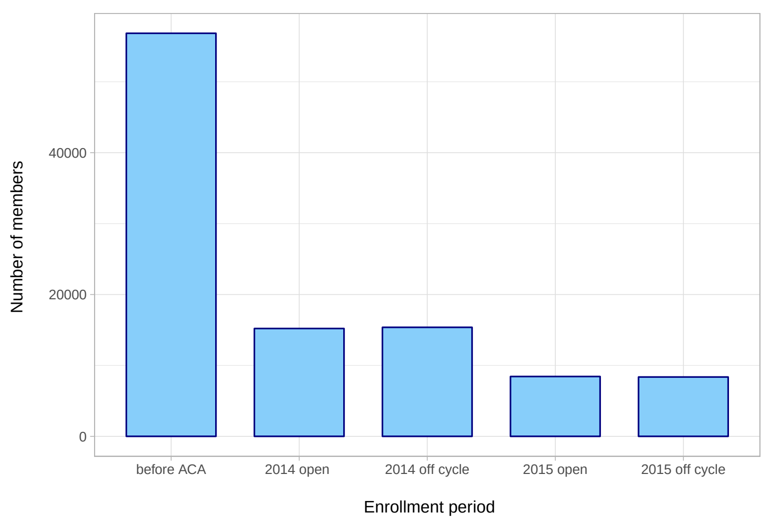

**Notes:** In this figure, we report the top 100 words classified by our machine learning algorithm as related to households with children. The font size indicates the frequency with which the word appears in the transactions dataset.

Figure 6: Summary statistics of consumers who participate in the individual insurance marketplaces, 2012-2015

(a) Monthly sign-ups and drop-outs

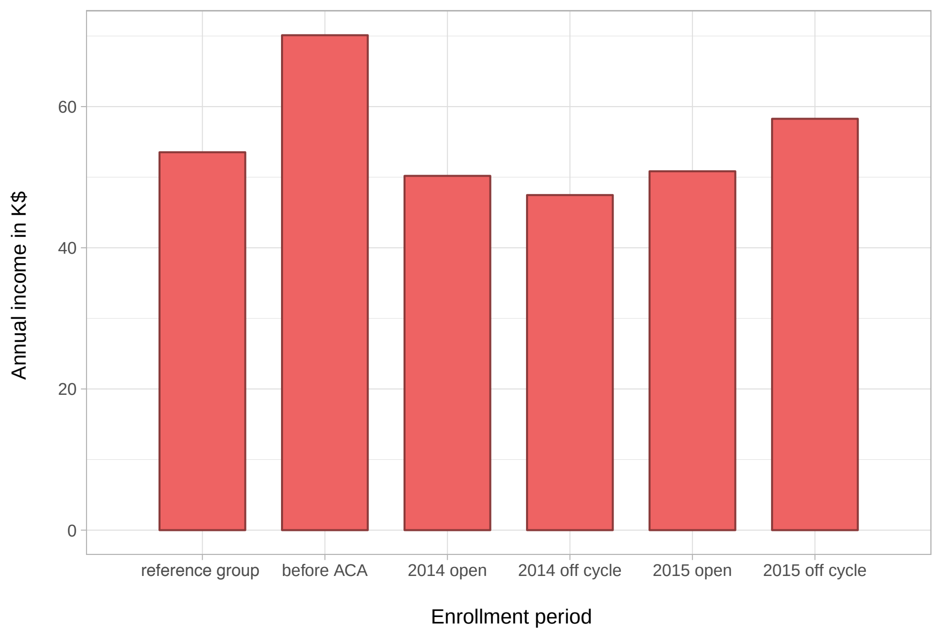



(b) Counts of members by enrollment period



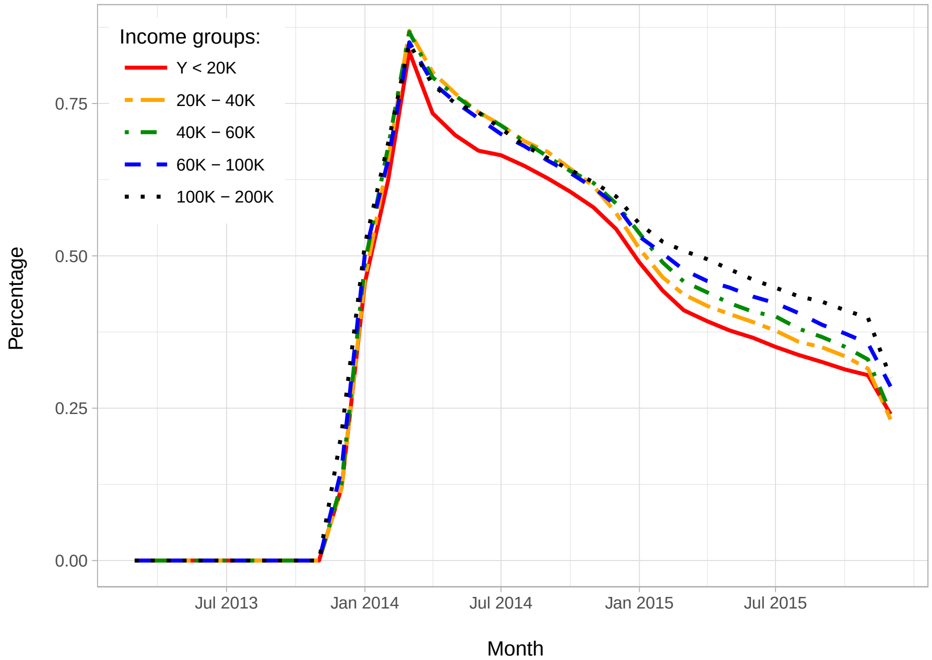

**Notes:** Panel (a) shows the number of new enrollee sign-ups and drop-outs in each month from Jan 2011 to Oct 2015. A new sign-up or drop-out is defined based on the observed first and last monthly premium payment. We label an enrollee as a “dropout" only if we observe his/her last premium payment 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.

Figure 7: Median Household Income by Enrollment Groups



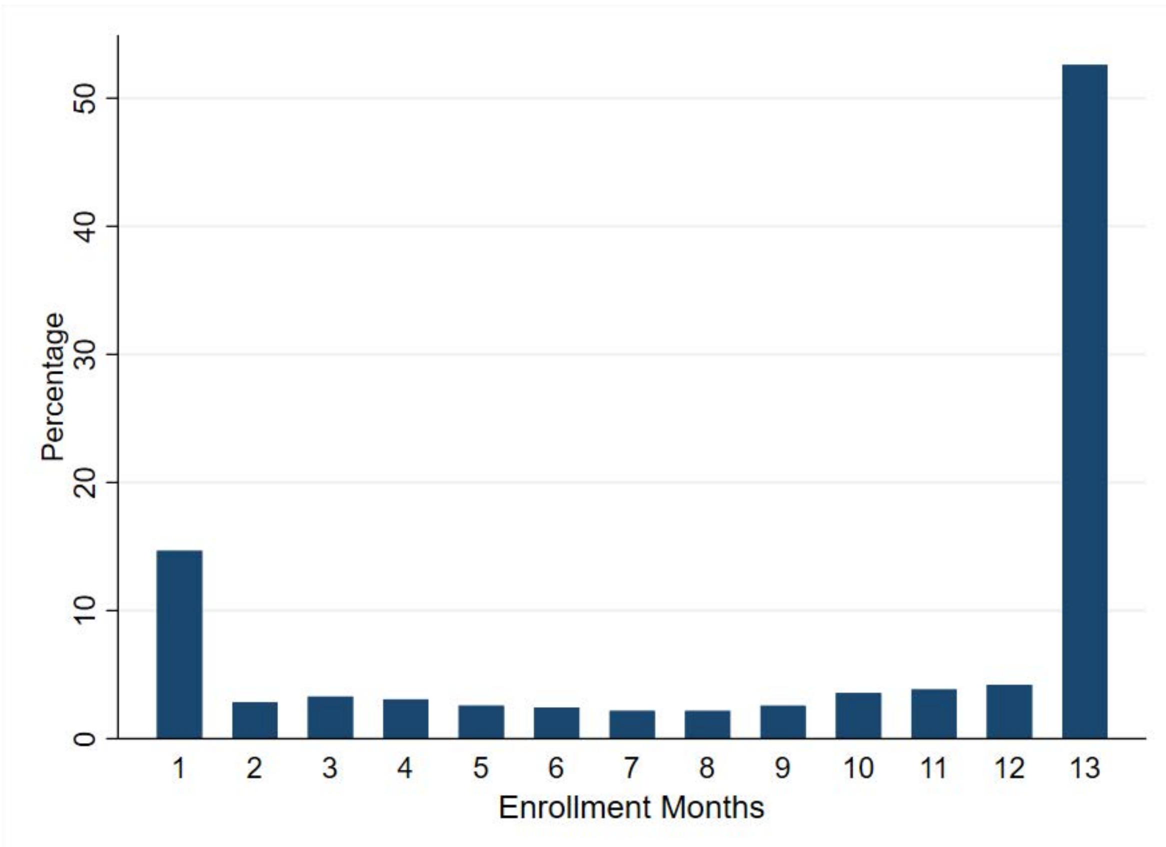

**Notes:** The figure displays the median household income by enrollment group from 2013. The annual 2013 household income of those who never purchase individual health insurance is \$52,000. Households who purchased individual health insurance before the ACA have a 2013 median income of \$70,000.

Figure 8: Drop-out Rates by Income Groups



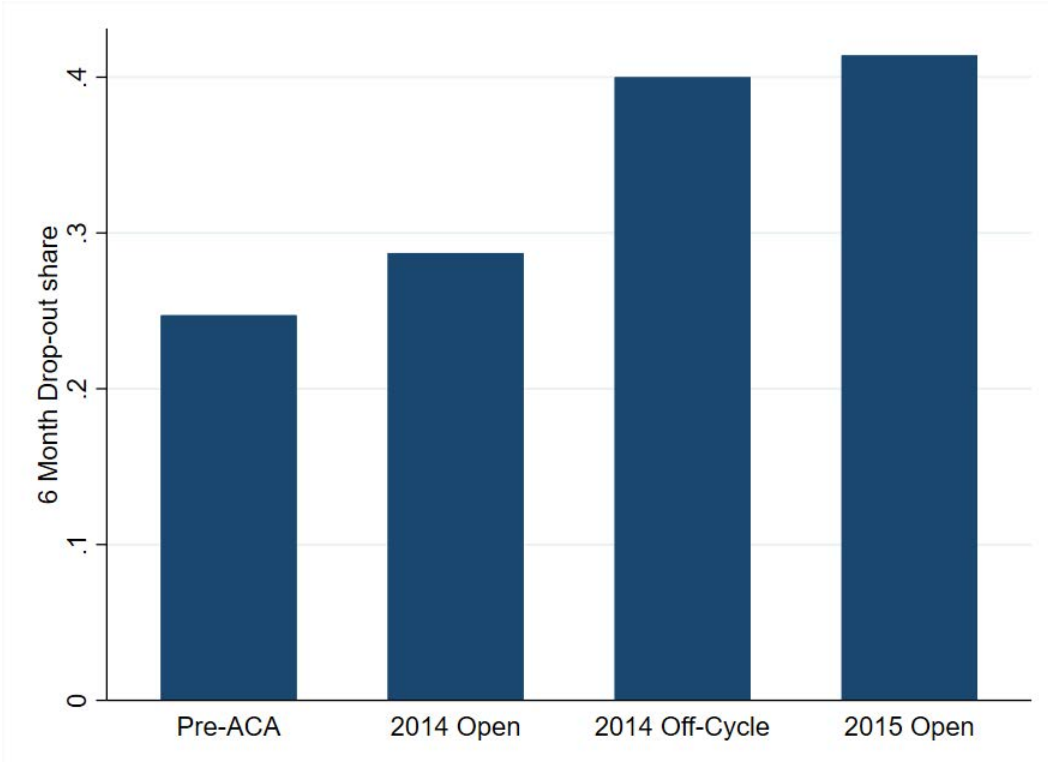

**Notes:** This figure shows the net share of households who continue to pay insurance premiums in each month, conditional on participating in the 2014 open enrollment period (N = 15,207). We break out this share by observed income. We observe that drop-out is highly prevalent among ACA enrollees across all income levels, with the strongest effects coming from the lowest income households. 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 9: Distribution of Drop-outs Across Months, for 2014 Open Enrollment Participants



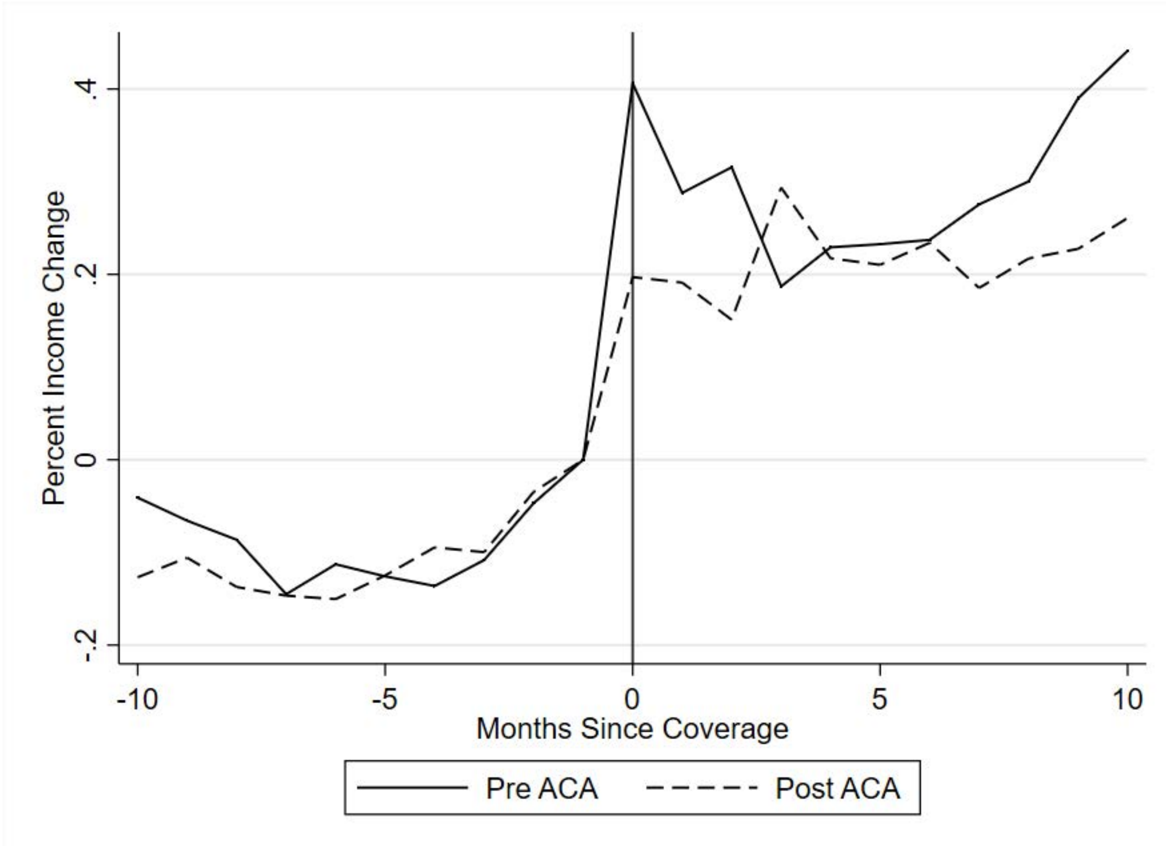

**Notes:** This figure illustrates the distribution of drop-outs across total months of coverage, conditional on participation in the 2014 open enrollment period. Households coded as 13 months remained enrolled for the full year or more.

Figure 10: Six Month Drop-out Rates by Enrollment Group




**Notes:** This figure shows the percent of households who drop their insurance coverage within six months of sign-up. This definition excludes drop-out in December, which is not subject to mandate penalties. The sample also does not include households who drop insurance within two months of exit from the transactions dataset, to ensure accurate measurement of drop-out.

Figure 11: Monthly Income Around Drop-out

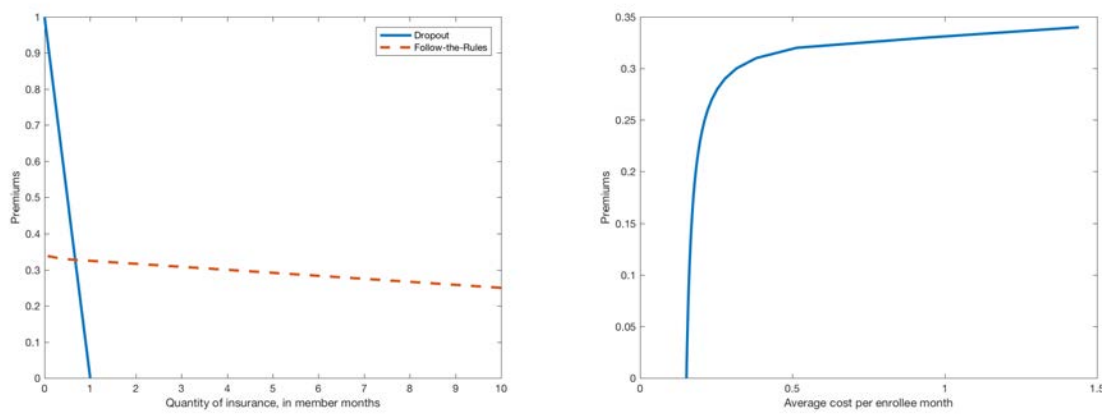



**Notes:** This figure compares the income dynamics around sign-up and drop-out, separately for the pre-ACA and post-ACA period. The x-axis measures months since coverage, where negative values indicate months since enrollment and strictly positive values indicate months since dropout. All months of 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 12: Demand for Insurance and Average Costs

(a) Demand for insurance, in member months

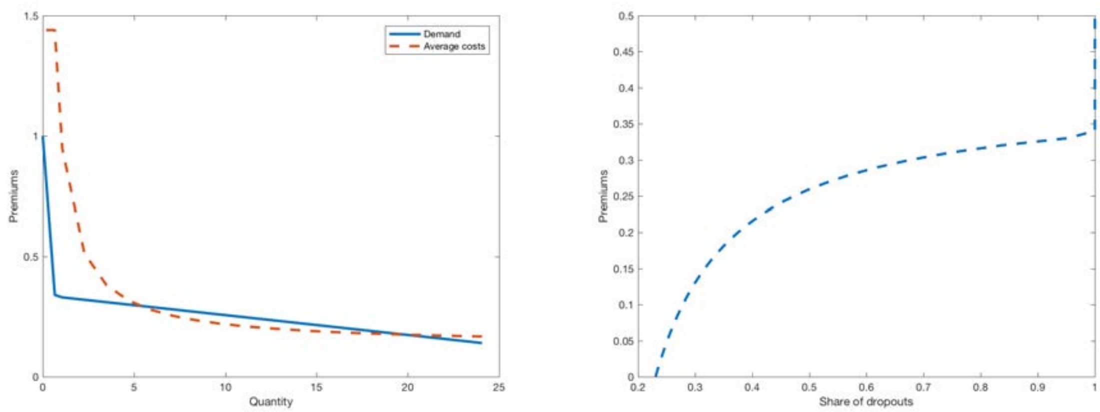
(b) Average costs per enrollee




**Notes:** This figure plots demand and average cost functions for the drop-out types and follow-the-rules types, respectively, for our example market. Figure 12a illustrates that the drop-out type (blue solid line) changes his quantity of coverage demanded, in member months, very little in response to changes in premiums. The follow-the-rules type (red dashed line) is more price sensitive. Figure 12b illustrates how the cost per enrollee to the insurer changes as the premium changes. With higher premiums, the quantity demanded by drop-out types falls less than the quantity demanded by follow-the-rules types, leading to higher average costs.

Figure 13: Adverse Selection from Drop-out Behavior

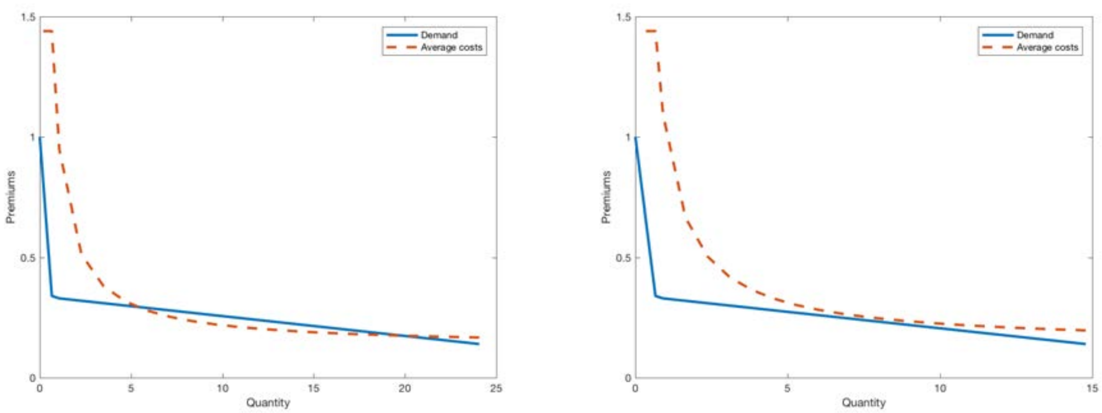
(a) Demand for insurance and average costs (b) Share of drop-outs vs. premiums




**Notes:** This figure plots average costs and market demand against premiums in the context of our example with two types of consumers, follow-the-rules types and drop-out types. In our example, both consumer types have identical underlying health risks. Figure 13a illustrates that as premiums rise, the average cost in the enrolled population increases. Thus, adverse selection arises despite the fact that the health risks are equal across the two types. The figure further illustrates a range of premiums for which demand exceeds average costs, and thus insurers will find it profitable to enter the market. At those premium levels, both follow-the-rules types and drop-out types pool in the market. Figure 13b illustrates how the share of enrolled drop-out types varies with changes in the premium level.

Figure 14: Adverse Selection from Drop-out Behavior and Market Unraveling

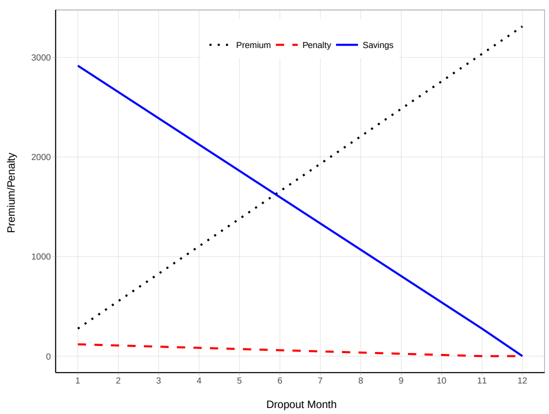
(a) Share of drop-out types is 10% (b) Share of drop-out types is 20%



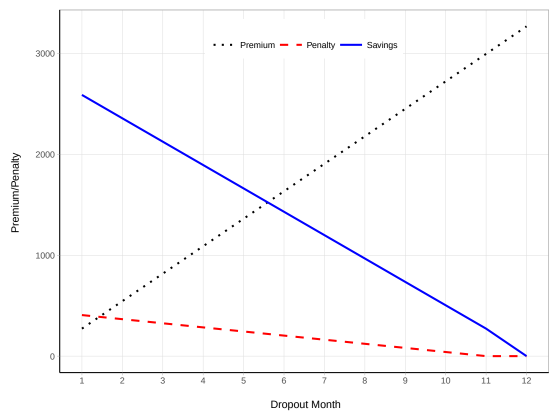

**Notes:** This figure illustrates the likelihood of unraveling as a function of the share of drop-out types in the potential enrollee pool in our example. Figures 14a and 14b plot the demand and average cost curves assuming 10% and 20% drop-out types in the potential enrollee pool, respectively. Figure 14a shows that insurers would find it profitable to enter the market in the presence of 10% drop-out types. In Figure 14b, the demand curve always lies below the average cost curve, indicating full unraveling when the share of drop-out types is 20%. That is, no insurer enters the market at any premium level.

Figure 15: **Back of the Envelope Penalty Estimates**

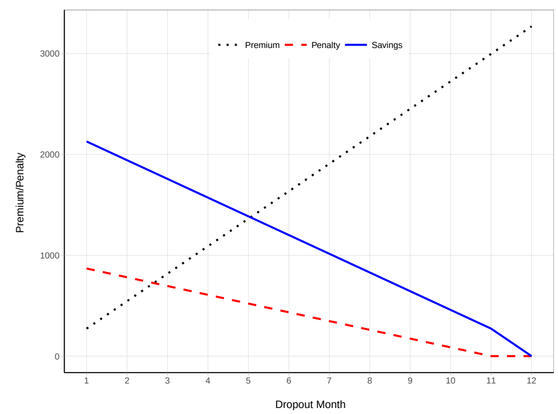
(a) 1 adult/1 child household, 2014



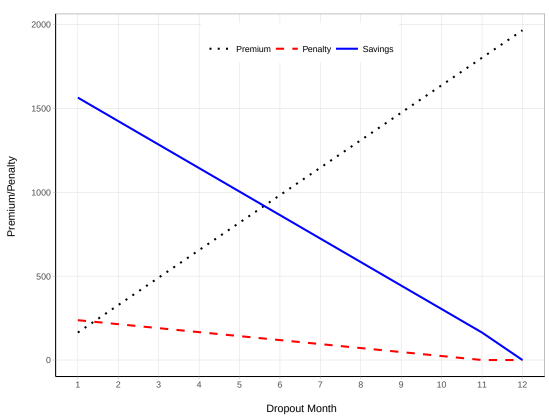

(b) 1 adult/1 child household, 2015



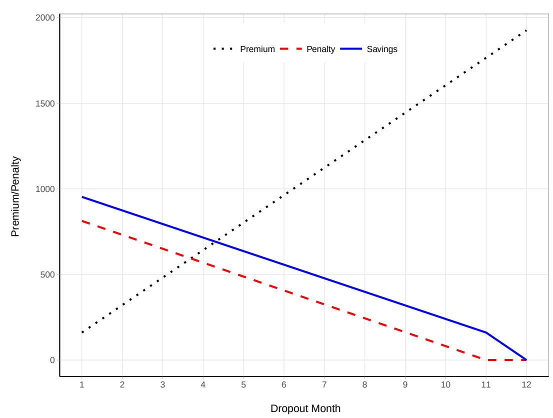

(c) 1 adult/1 child household, 2016



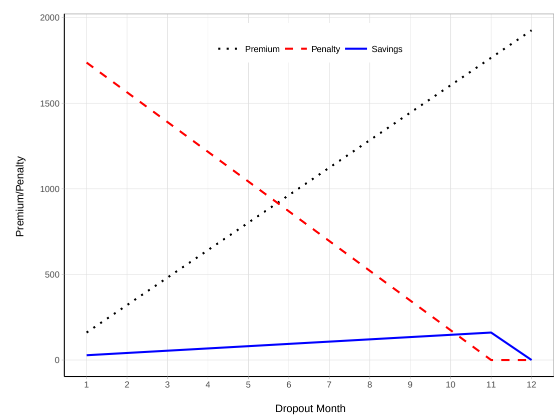

(d) 2 adults/2 children household, 2014




(e) 2 adults/2 children household, 2015




(f) 2 adults/2 children household, 2016




**Notes:** The figure illustrates a household's cumulative premium payment, penalty, and net savings from dropping coverage depending upon (a) the duration of coverage and (b) household size, while fixing household annual income at $40,000. Drop-out month $n$ is defined as the month during which an individual drops ACA insurance coverage. Penalties equal zero in months 11 and 12 due to the "short gap" exemption of two months. Monthly penalties and premiums are cumulative at each drop-out month. Savings is defined as the cumulative difference between monthly premium and penalty for each drop-out month, compared to the amount an individual who never drops coverage would pay. Monthly premium is a function of household size and annual income, accounting for federal premium subsidies. Monthly penalties are computed on the per-person basis, based on IRS rules.

Table 1: 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 2: 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 3: **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 3.5). 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 4: **Healthcare Consumption for 2014 Open Enrollees**

| | Y < 20K | 20K - 40K | 40K - 60K | 60K - 100K | 100K - 200K |
|---|---|---|---|---|---|
| | **Panel A: Change in health transactions** | | | | |
| Diff | 0.708*** | 0.365*** | 0.309*** | 0.181*** | 0.062 |
| | (0.077) | (0.049) | (0.055) | (0.04) | (0.045) |
| Diff-Diff | 0.514*** | 0.27*** | 0.193*** | 0.134*** | 0 |
| | (0.096) | (0.061) | (0.065) | (0.049) | (0.054) |
| | **Panel B: Change in health spending** | | | | |
| Diff | 0.471*** | 0.232** | 0.247*** | 0.038 | -0.002 |
| | (0.115) | (0.098) | (0.087) | (0.068) | (0.072) |
| Diff-Diff | 0.45*** | 0.248** | 0.232** | 0.055 | -0.003 |
| | (0.142) | (0.122) | (0.103) | (0.077) | (0.086) |
| | **Panel C: Change in drug transactions** | | | | |
| Diff | 0.289*** | 0.127*** | 0.05** | 0.021 | 0.032 |
| | (0.049) | (0.022) | (0.024) | (0.021) | (0.025) |
| Diff-Diff | 0.114* | 0.057** | 0.041 | 0.01 | 0.006 |
| | (0.06) | (0.028) | (0.028) | (0.025) | (0.027) |
| | **Panel D: Change in drug spending** | | | | |
| Diff | 0.382*** | 0.148*** | 0.002 | 0.045 | 0.066** |
| | (0.054) | (0.029) | (0.03) | (0.031) | (0.032) |
| Diff-Diff | 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:** Robust standard errors in parentheses. * $p<0.1$, ** $p<0.05$, *** $p<0.01$
The treatment group includes individuals who enrolled during the 2014 open enrollment period, from January through March. 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.

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** | | | | |
| Diff | 0.788*** | 0.312*** | 0.457** | 0.098 | -0.006 |
| | (0.219) | (0.096) | (0.215) | (0.08) | (0.141) |
| Diff-Diff | 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** | | | | |
| Diff | 0.924*** | 0.444** | 0.191 | -0.149 | -0.268 |
| | (0.281) | (0.174) | (0.189) | (0.139) | (0.165) |
| Diff-Diff | 0.968*** | 0.378* | 0.203 | -0.104 | -0.208 |
| | (0.356) | (0.223) | (0.227) | (0.165) | (0.187) |
| | **Panel C: Change in drug transactions** | | | | |
| Diff | 0.163 | 0.095 | -0.011 | 0.109* | -0.108* |
| | (0.113) | (0.062) | (0.07) | (0.059) | (0.064) |
| Diff-Diff | 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** | | | | |
| Diff | 0.154 | 0.131 | -0.127 | 0.21** | -0.085 |
| | (0.137) | (0.085) | (0.102) | (0.105) | (0.083) |
| Diff-Diff | 0.031 | 0.021 | -0.163 | 0.043 | -0.02 |
| | (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:** Robust standard errors in parentheses. * $p$<0.1, ** $p$<0.05, *** $p$<0.01
The treatment group includes individuals who enrolled during the 2014 open enrollment period, from January through March. 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 and Drop-Out

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | <20K | 20K-40K | 40K-60K | 60K-100K | 100K-200K |
| **Panel A: Change in health spending: Pre ACA Sign-up/Drop-out** | | | | | |
| signup | 1.170*** | 0.133 | 0.176 | 0.0990 | 0.339 |
| | (0.384) | (0.158) | (0.139) | (0.108) | (0.218) |
| dropout | -0.947** | -0.0312 | -0.0720 | -0.126 | -0.0764 |
| | (0.368) | (0.167) | (0.141) | (0.0957) | (0.223) |
| Pre-Signup Monthly Health Spend | 20.45 | 20.03 | 29.16 | 30.92 | 36.20 |
| Number of observations | 25641 | 17879 | 15311 | 19255 | 15296 |
| **Panel B: Change in health spending: Post ACA Sign-up/Drop-out** | | | | | |
| signup | 0.275*** | 0.542*** | 0.110 | 0.185* | 0.0934 |
| | (0.0968) | (0.131) | (0.100) | (0.103) | (0.129) |
| dropout | -0.403*** | -0.505*** | -0.0529 | -0.275*** | -0.178 |
| | (0.0786) | (0.146) | (0.0988) | (0.104) | (0.124) |
| Pre-Signup Monthly Health Spend | 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** | | | | | |
| signup | 0.318*** | 0.180* | 0.0394 | 0.0504 | 0.279*** |
| | (0.0938) | (0.103) | (0.0818) | (0.0828) | (0.0879) |
| dropout | -0.0846 | -0.0569 | -0.0702 | -0.120* | -0.118 |
| | (0.0932) | (0.0964) | (0.0824) | (0.0647) | (0.0945) |
| Pre-Signup Monthly Transactions | 0.21 | 0.25 | 0.33 | 0.40 | 0.35 |
| Number of observations | 25641 | 17879 | 15311 | 19255 | 15296 |
| **Panel D: Change in health transactions: Post ACA Sign-up/Drop-out** | | | | | |
| signup | 0.426*** | 0.359*** | 0.175*** | 0.143** | 0.0726 |
| | (0.0580) | (0.0600) | (0.0529) | (0.0560) | (0.0485) |
| dropout | -0.280*** | -0.236*** | -0.187*** | -0.142** | -0.111** |
| | (0.0567) | (0.0611) | (0.0527) | (0.0560) | (0.0519) |
| Pre-Signup Income | 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 7: Drug Spending Around Sign-up and Drop-Out

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | <20K | 20K-40K | 40K-60K | 60K-100K | 100K-200K |
| **Panel A: Change in drug spending: Pre ACA Sign-up/Drop-out** | | | | | |
| signup | 0.773*** | 0.291*** | 0.283*** | 0.533** | 0.560*** |
| | (0.0907) | (0.0720) | (0.0727) | (0.210) | (0.117) |
| dropout | -0.195** | 0.0945 | -0.0469 | -0.117 | -0.00115 |
| | (0.0962) | (0.128) | (0.0775) | (0.115) | (0.113) |
| Pre-Signup Monthly Drug Spend | 15.09 | 18.46 | 23.79 | 24.76 | 27.01 |
| Number of observations | 25641 | 17879 | 15311 | 19255 | 15296 |
| **Panel B: Change in drug spending: Post ACA Sign-up/Drop-out** | | | | | |
| signup | 0.0726** | 0.0630 | 0.0707* | 0.0733 | -0.0353 |
| | (0.0330) | (0.0495) | (0.0424) | (0.0536) | (0.0473) |
| dropout | -0.0748* | -0.105** | -0.162*** | -0.186*** | -0.0922** |
| | (0.0384) | (0.0495) | (0.0446) | (0.0562) | (0.0441) |
| Pre-Signup Monthly Drug Spend | 22.41 | 21.91 | 27.91 | 32.43 | 43.66 |
| Number of observations | 72677 | 55702 | 36951 | 38203 | 28006 |
| **Panel C: Change in drug transactions: Pre ACA Sign-up/Drop-out** | | | | | |
| signup | 0.718*** | 0.374*** | 0.346*** | 0.457*** | 0.481*** |
| | (0.0705) | (0.0570) | (0.0629) | (0.0602) | (0.0788) |
| dropout | -0.194*** | -0.156*** | -0.0714 | -0.0633 | -0.0396 |
| | (0.0675) | (0.0577) | (0.0617) | (0.0603) | (0.0668) |
| Pre-Signup Monthly 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** | | | | | |
| signup | 0.0932*** | 0.0636** | 0.0699** | -0.00562 | 0.0175 |
| | (0.0241) | (0.0259) | (0.0304) | (0.0247) | (0.0255) |
| dropout | -0.0720*** | -0.0944*** | -0.121*** | -0.102*** | -0.0836*** |
| | (0.0237) | (0.0249) | (0.0306) | (0.0279) | (0.0276) |
| Pre-Signup Month Transactions | 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. 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 8: Urgent and Essential Health Spending, for Continuing Enrollees vs. Dropouts

| VARIABLES | % 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 3.5). 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 9: **Pricing Regressions**

| | (1)<br>**Price Change (%)** | (2)<br>**Price Change (%)** | (3)<br>**Price Change (%)** | (4)<br>**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 A3 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.

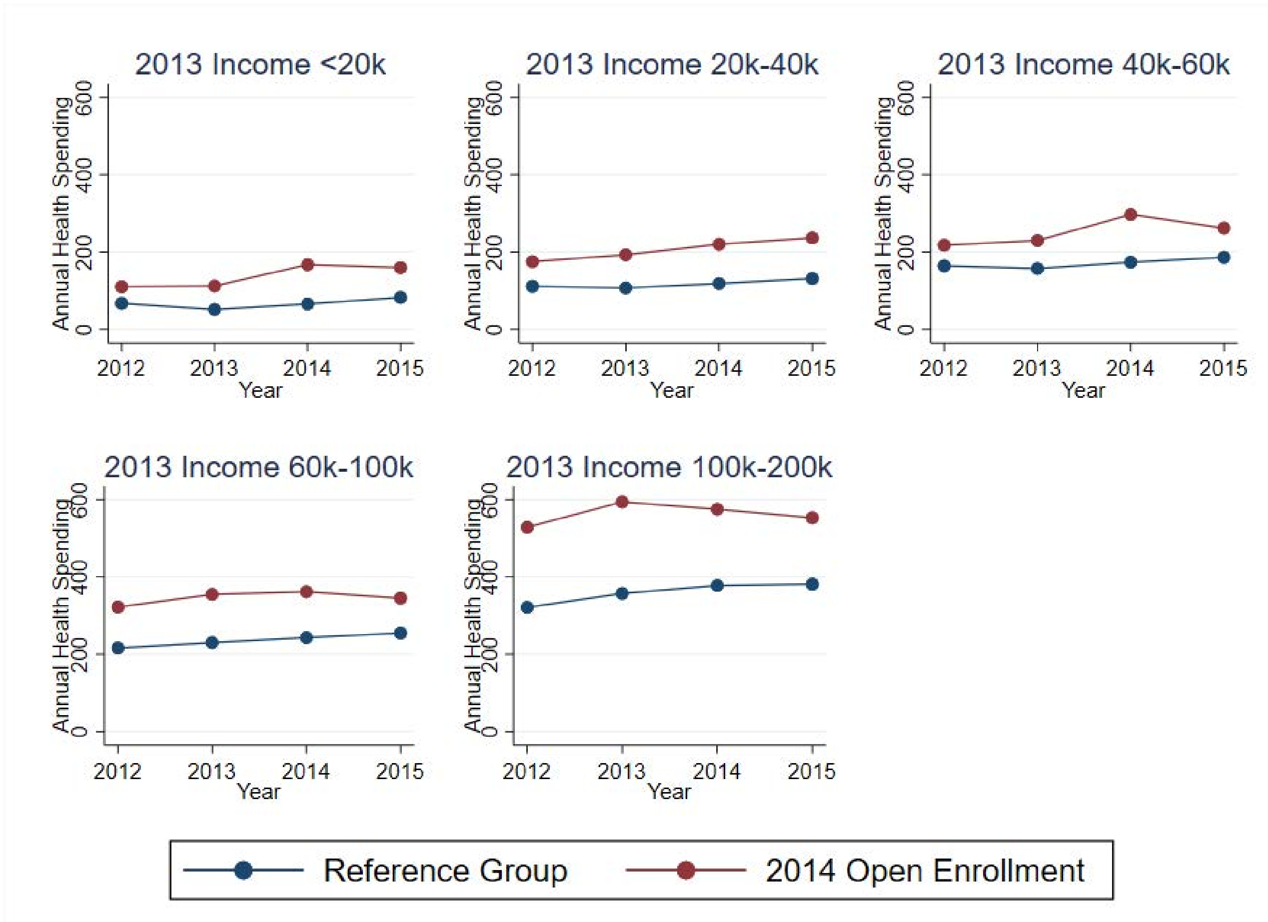
## A.3 Classifying Transactions

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.[41] 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.

[41] 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.

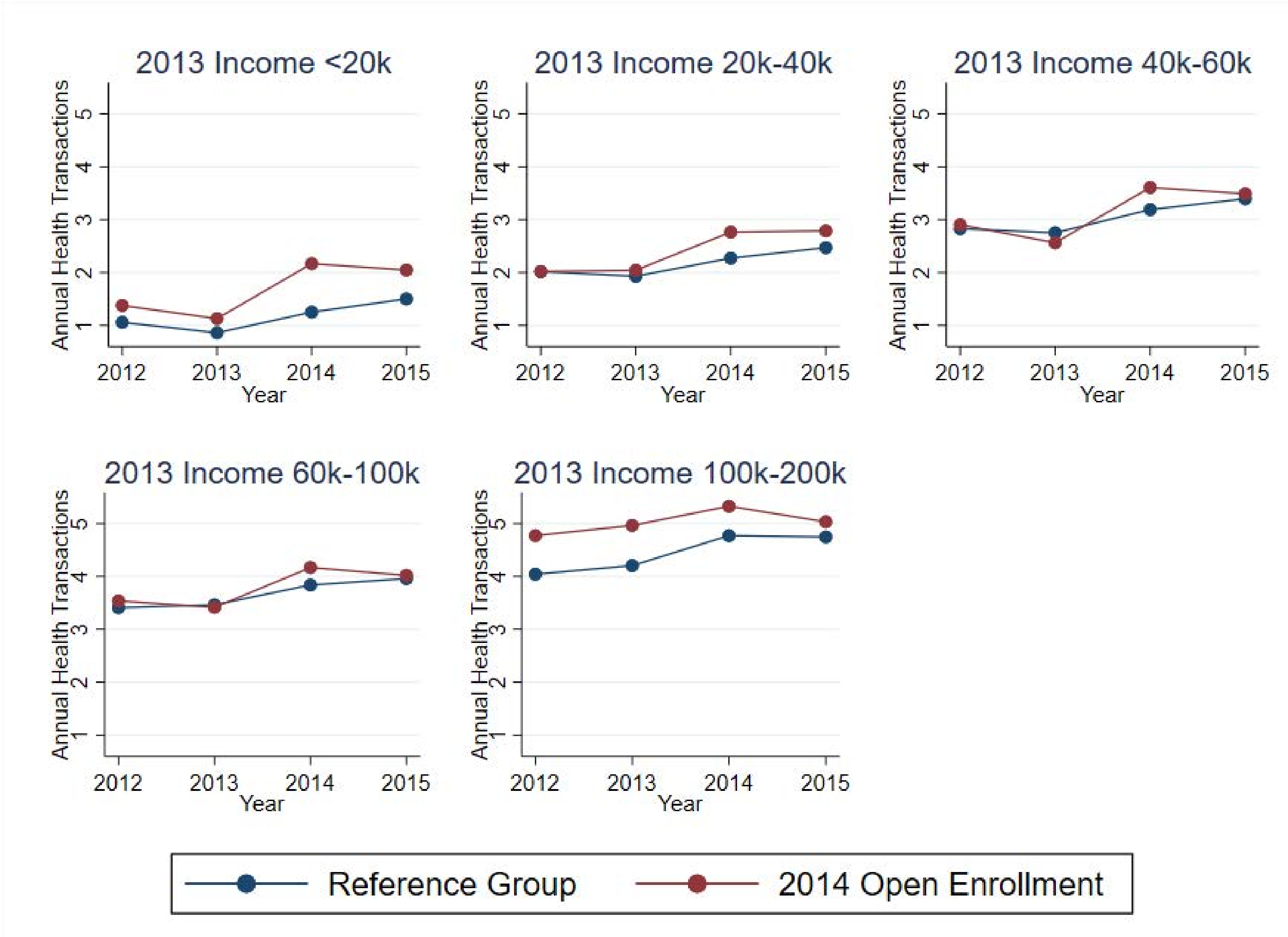
## A.4 Appendix Figures and Tables

Figure A1: Annual Health Spending: 2014 Open Enrollees vs Reference Group



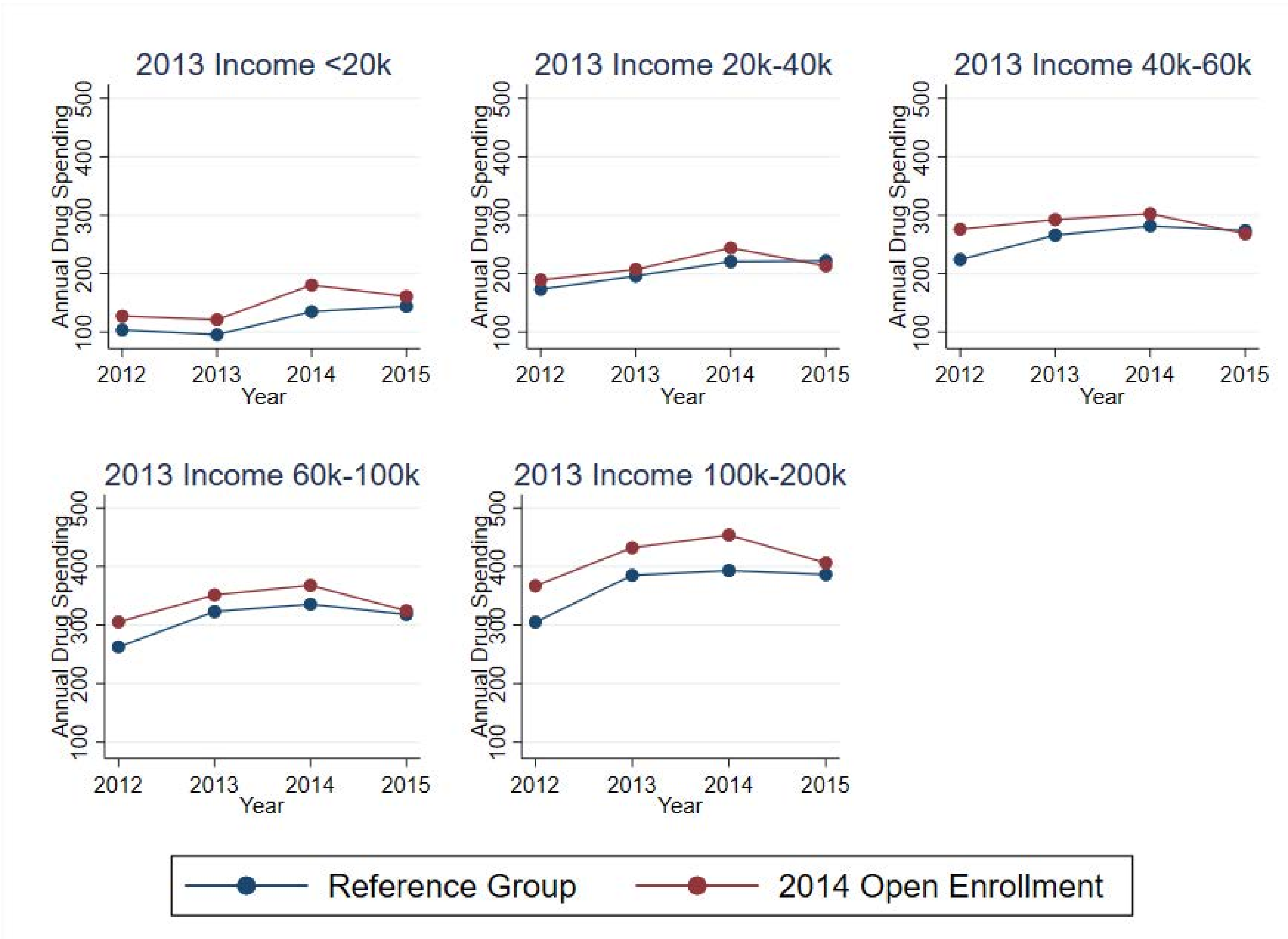

**Notes:** Figures show average annual health care spending by income group for 2014 open enrollees and a "reference group" of those who never purchase individual health insurance.

Figure A2: Annual Health Transactions: 2014 Open Enrollees vs Reference Group



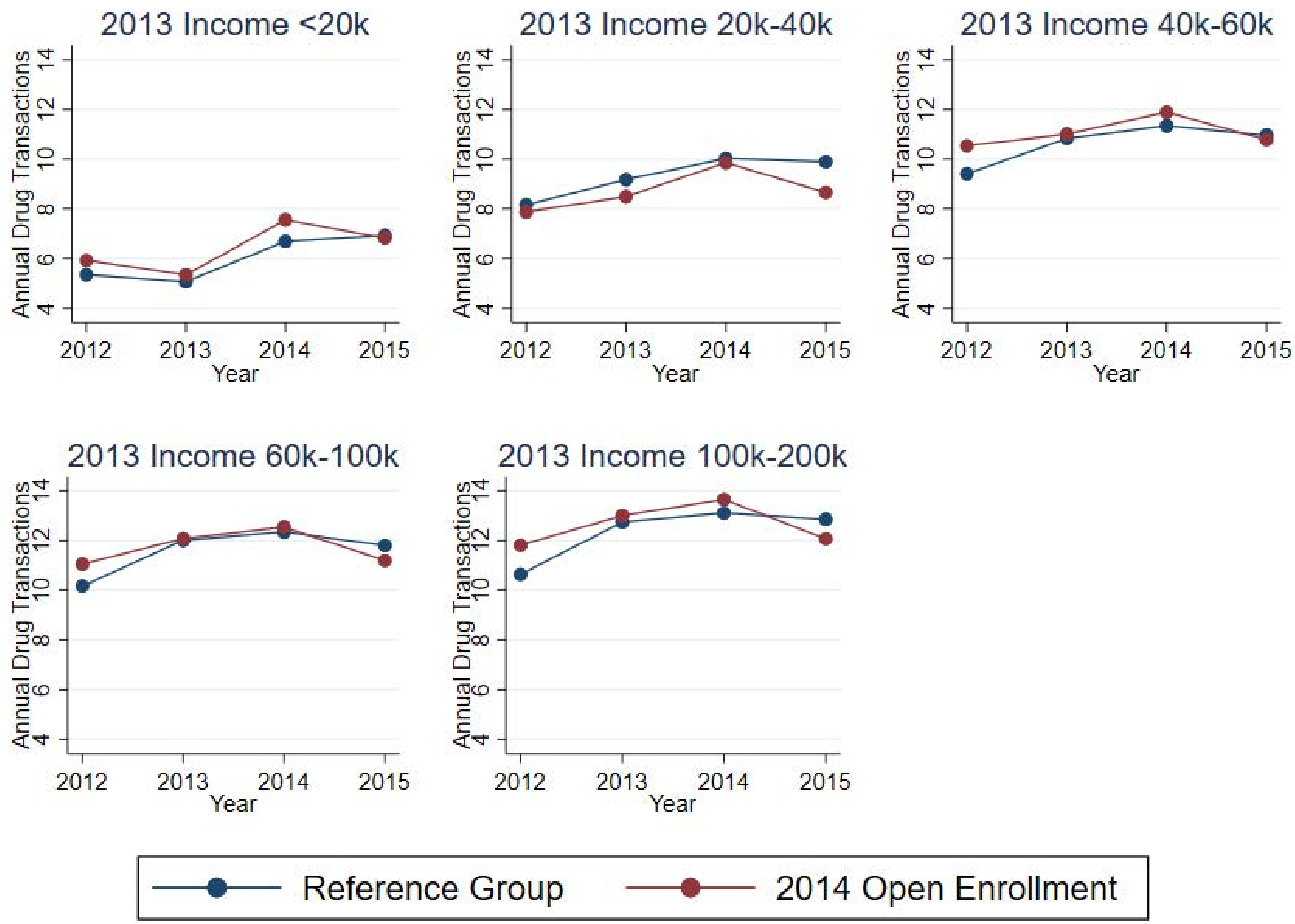

**Notes:** Figures show average annual health care transactions by income group for 2014 open enrollees and a “reference group” of those who never purchase individual health insurance.

Figure A3: 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 A4: 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<br>$250 brand drugs | $0 | $0 |
| Annual Out-of-Pocket,<br>Maximum Individual and Family | $6,250 individual<br>$12,500 family | $6,250 individual<br>$12,500 family | $6,250 individual<br>$12,500 family | $4,000 individual<br>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: 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.

Table A4: 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.

## A.5 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)