Heterogeneous Treatment Effects of Medicaid and Efficient Policies: Evidence from the Oregon Health Insurance Experiment

Click here for latest version.

Abstract

The optional provision of Medicaid expansion, through the Affordable Care Act (ACA), has triggered a national debate among diverse stakeholders regarding the impacts of Medicaid coverage on various dimensions of public health, costs, and benefits. Randomized experiments like the Rand Health Insurance Experiment and the Oregon Health Insurance Experiment have generated some credible estimates of the average treatment effects of access to insurance. However, identical policy interventions can have heterogeneous effects on different subpopulations. This paper uses data from the Oregon Health Insurance Experiment to estimate the heterogeneous treatment effects of access to Medicaid on health care utilization, preventive care utilization, financial strain, and self-reported physical and mental health. I detect heterogeneous treatment effects using a cluster-robust generalized random forest, a causal machine learning approach. I find that the impact of Medicaid is more pronounced among relatively older non-elderly and poorer households, which is consistent with standard adverse selection theory. Furthermore, I implement the "efficient policy learning," another machine learning strategy, to identify policy changes that prioritize providing Medicaid coverage to the subgroups that are likely to benefit the most. On average, the proposed reforms would improve the average probability of outpatient visits, preventive care use, overall health outcomes, having a personal doctor and clinic, and happiness by a range of 2% to 9% over a random assignment baseline. These findings are helpful for designing Medicaid Section 1115 waiver.

Keywords: Insurance, causal machine learning, heterogeneous treatment effect, efficient policy learning

JEL Classification: I13, D04, C93, C44, C54

1 Introduction

As of September 20, 2019, 37 states and the District of Columbia have expanded Medicaid coverage for low-income adults to 138%[1] of the Federal Poverty Level through the Affordable Care Act (ACA). This optional[2] provision to expand the Medicaid program through the ACA has triggered a substantial nationwide debate among policymakers and diverse stakeholders about what effects, if any, Medicaid has on the various dimensions of health (Baicker, 2019).

Finkelstein et al. (2012) use random assignment of Medicaid, employing the Oregon Health Insurance Experiment (OHIE) dataset, and found mixed-bag effects[3] of Medicaid which have presented policymakers with tough choices in balancing the costs and benefits of Medicaid (Baicker, 2019). Meanwhile, states like Florida, Minnesota, and North Carolina are analyzing their Medicaid programs to find potential savings, some of which could be redirected to improve access and the quality of care to patients served by the Medicaid program (Rueben, 2019). Furthermore, another significant reform in Medicaid is the "Medicaid work requirements[4]," which take away Medicaid coverage from people not engaging in work or work-related activities for a specified number of hours each month (Katch et al., 2018).

The mixed-bag effect of Medicaid and policymakers' quest for Medicaid reforms are the main motivations of my research. In this paper, I provide answers to the questions of why previous literature finds the mixed-bag effect of Medicaid?, and how to think about Medicaid reforms while improving the effectiveness of Medicaid? To answer these research questions, I use the Oregon Health Insurance Experiment (OHIE) public-use data. This data set comprises the lottery assignment of Medicaid in Oregon, thus creates a randomized controlled study setting and allows causal analysis by comparing various outcomes of the lucky Oregonians who received Medicaid to those who did not (Klein, 2013). The primary rationale to use OHIE data is that random assignment of Medicaid allows circumventing the challenges of endogeneity. Endogeneity arises because it is difficult to control[5] for observed and unobserved confounding[6] variables among the insured and uninsured population (Levy and Meltzer, 2008).

Answers of the above research questions can contribute to two primary domains that are relevant

[1]Medicaid income eligibility limits for adults as a percent of the Federal Poverty Level, indeed, are different from states to states. Kaiser Family Foundation provides a table for the state by state Medicaid income eligibility levels for adults.

[2]Following the June 2012 Supreme Court decision, states face a decision about whether to adopt the Medicaid expansion. But, as per the Centers for Medicare and Medicaid Services (CMS) guidance, there is no deadline for states to implement the Medicaid expansion (Kaiser Family Foundation, 2019).

[3]Finkelstein et al. (2012) use OHIE data set, and found that, in the year following the random assignment of lottery Medicaid, the treatment group had higher health care use, lower out-of-pocket medical expenditures and medical debt, and better self-reported physical and mental health than the control group, but did not reflect any detectable improvements in physical health conditions like high blood pressure.

[4]Centers for Medicare Medicaid Services (CMS) guidance for state Medicaid waiver proposals, issued on January 2018, allows states, for the first time, to impose work requirements in Medicaid as a condition of eligibility. As a result, several states have received approval for or are pursuing these waivers. See Garfield et al. (2019) for details.

[5]For example, a comparison of the health between those with and without health insurance, can reveal that insurance is detrimental for one's health because people with poor health are more likely to get insurance compared to healthy people (Baicker and Finkelstein, 2011).

[6]Confounding variables are common causes that explain both treatment and outcome variables.

for policy development. First, unlike the series of papers[7] that have evaluated the average treatment effects of Medicaid, I contribute by estimating the heterogeneous treatment effect of lottery Medicaid, employing the Athey et al. (2019) cluster-robust generalized random forest, on several outcome variables. These outcomes variables are health care utilization, preventive care utilization, financial strain, self-reported physical and mental health, and several variables of potential mechanism to improve health. The primary rationale to understand the heterogeneous treatment effects is that the identical policy intervention can often distinctly affect different individuals and subpopulations in different ways. Along with average treatment effects, policymakers are usually interested in how the effects of intervention vary across subpopulations. Identifying the heterogeneous treatment effects accommodate the discovery of underlying mechanisms that drive the results and allow for efficient design and reform of policy.

Second, I contribute insights regarding how to target health insurance interventions for effective policymaking using the Athey and Wager (2019a) strategies of "efficient policy learning." Understanding "who should be treated" with intervention is ubiquitous in policymaking. It can be unfair, unethical, and sometimes illegal to target policy at only a particular subpopulation. Moreover, intervening everybody in the population (a blanket policy) is welfare-maximizing but can be costly.[8] The main logic of efficient policy learning is to identify treatment assignment policies based on easily observable individual characteristics. The treatment assignment, in this paper, represents Medicaid assignment.

To investigate the heterogeneous treatment effects, one can stratify the data in mutually exclusive groups or include interactions within a regression (Athey and Imbens, 2017a). However, for large-scale investigations of effect heterogeneity, $p-$values of standard "single" hypothesis tests are no longer valid because of the multiple hypothesis testing[9] problems (Lan et al., 2016; List et al., 2019). Moreover, performing ad-hoc searches or $p-$hacking[10] to detect the responsive subgroups may lead to false discoveries or may mistake noise for an actual treatment effect (Davis and Heller, 2017). To avoid many of the issues associated with data mining or $p-$hacking, researchers can commit in advance to study only a subgroup by a preregistered analysis plan.[11] However, this may also prevent discovering unanticipated results and

[7] See Allen et al. (2010); Baicker et al. (2013, 2017, 2014); Baicker and Finkelstein (2011); Finkelstein et al. (2012); Grossman et al. (2016); Taubman et al. (2014); Zhou et al. (2017).

[8] For example, a provision of the Affordable Care Act (ACA) was that the federal government would pay the full cost of coverage expansion through 2016. Moreover, it would reimburse at least 90% of the cost of covering the newly-insured population (Norris, 2018). Oregon responded to this incentive by expanding Medicaid in January 2014 and ensured insurance to everyone with incomes up to 133% of the federal poverty line. When the federal government gradually reduced their payments, the state budget of Oregon (nearly \$74 billion for 2017-2019) suffered about \$1 billion budget hole due to the cost of health care (Foden-Vencil, 2018).

[9] The "multiple hypothesis testing problems" leads to the so-called "ex-post selection problem," which is widely recognized in the program evaluation literature. For example, for fifty single hypotheses tests, the probability that at least one test falsely rejects the null hypotheses at the 5% significance level (assuming independent test statistics as an extreme case) is $1 - 0.95^{50} = 0.92$ or 92%.

[10] The $p-$hacking is an exhaustive search for statistically significant relations from combinations of variables or combinations of interactions of variables or subgroups. The $p-$hacking could lead to discovering the statistically significant relationship, when, in fact, there could have no real underlying effect.

[11] A preregistered analysis plan is sets of analyses plans released in the public domain by the researchers in advance prior they collect the data and learn about outcomes. For example, The American Economic Association's registry for randomized controlled trials is a reputable platform for conducting a preregistered analysis plan.

developing new hypotheses (Athey and Imbens, 2016).

I implement the cluster-robust generalized random forest methods, developed by the Athey et al. (2019), on the OHIE dataset to explore the heterogeneous treatment effects of Medicaid. The Athey et al. (2019) method re-engineers the strengths and innovations of the Breiman (2001) random forest[12], a predictive machine learning method for causal inference. The Athey et al. (2019) modifications[13] allow for a systematic investigation of the heterogeneous treatment effects that are not prone to data mining and $p-$hacking, and useful when research includes high-dimensional covariates.[14] Furthermore, OHIE provides individual-level data set, but the Medicaid lottery intervention occurred at household-level; therefore, the outcome variable may be arbitrarily correlated within a household. The Athey et al. (2019) allows a conservative approach of cluster-robust analysis to account for potential correlations within each household cluster.

Along with the heterogeneous treatment effects, the question of: "Who should get treatment?" is also a widespread issue in policy design. For example, who should get in youth employment programs (Davis and Heller, 2017), who should get Medicare funding for hip or knee replacement surgery (Kleinberg et al., 2015), who should get a job training, job searching support, and other assistance (Kitagawa and Tetenov, 2018). My paper implements the "efficient policy learning" strategies of Athey and Wager (2019a) to answer how to set eligibility criteria to intervene with Medicaid coverage. The Athey and Wager (2019a) approach allows identifying policy changes/reforms that prioritize providing Medicaid coverage to the subgroups that are likely to benefit the most.

I show efficient policy rules considering two rationales $-$ first, I exclude observable covariates like race, gender, and residence. Excluding these covariates are essential to allow ethical, legislative, and political considerations of policy design. Second, I follow the Kitagawa and Tetenov (2018) approach to design policy from an "intent-to-treat" perspective. This approach is crucial because the policy maker's problem is only a choice of the eligibility criteria and not the take-up[15] rate.

I find the heterogeneous treatment effect of Medicaid on health care utilization, preventive care utilization, financial strain, self-reported physical and mental health, and several variables of potential

[12]The Breiman (2001) random forest ensembles or bootstrap and aggregate many classifications and regression tree (CART) of the Breiman et al. (1984), and report the average. The CART recursively filters and partitions the large dataset into binary sub-groups (nodes) such that the samples within each subset become more homogeneous in their fit of the response variable, thus resulting in a tree-like format.

[13]The modifications are based on the "causal tree" (Athey and Imbens, 2016), "causal forest" (Wager and Athey, 2018) and the "generalized random forest" (Athey et al., 2019) methods. The "causal tree" approach re-engineers the Breiman et al. (1984) classification and regression tree (CART), a machine learning algorithms for causal inference. The remaining methods extend the "causal tree" approach utilizing the Breiman (2001) random forest machine learning algorithm for causal inference.

[14]The nearest-neighbor matching, kernel methods, and series estimation are classical approaches for nonparametric estimation of heterogeneous treatment effects (Crump et al., 2008; Lee, 2009; Willke et al., 2012), and performs well with a smaller set of covariates. However, these classic approaches break down quickly when covariates are large in numbers (Athey and Wager, 2019a).

[15]The take-up rate in our study is the percentage of eligible people who accept Medicaid benefits. Individuals decide the take-up rate for various reasons unknown to the policymakers.

mechanism to improve health. For each of these outcome variables, I display the causal thresholds for distinct subpopulations where the impacts of Medicaid intensify and subdue. These realms have not been explored earlier, and my results are some unique contributions to the literature. My findings, therefore, provide a holistic perspective toward the large, and at times contradictory research exploring the effects of Medicaid on health. I find that the heterogeneous impacts of Medicaid are more pronounced among poorer and older non-elderly households. These impoverished families may need more medical services, and when Medicaid provides an opportunity, these households utilize more health care compared to those who are uninsured, just as standard adverse selection theory would predict.

Furthermore, I find efficient policies or reforms for several selected outcome variables. I quantify the cost of estimated policy rules in comparison to the random assignment of Medicaid. On average, the proposed reforms would improve the average probability of outpatient visits, preventive care use, overall health outcomes, having a personal doctor and clinic, and happiness by a range of 2% to 9% over a random assignment baseline, and these improvements are likely to support a causal interpretation.

In summary, I use the Oregon Health Insurance Experiment public-use data and contribute to examining the sources of treatment heterogeneity on Medicaid programs and offering efficient policy rules or reforms that prioritize Medicaid allotments to subgroups that are likely to benefit the most. The findings of this paper are useful for analysts, policymakers, and insurance designers to discover the underlying mechanisms that drive the health outcome results and to design or reform policy. For example, the proposed reforms can help Oregon to develop a priority list against current blanket Medicaid policy which can help to reduce the state budget-deficits[16] without hampering the current Medicaid welfare.

Section 2 summarizes the institutional background of the Oregon Health Insurance Experiment. Section 3 summarizes approaches to study health insurance and health outcomes and explains how causal machine learning can help to analyze different research questions. Section 4 lays out identification strategy and empirical methods for the cluster-robust random forest for heterogeneous treatment estimation along with efficient policy learning strategies. Section 5 displays the results. Section 6 provides discussions on findings and concludes the study.

2 Oregon Health Insurance Experiment

Oregon's Medicaid program, the Oregon Health Plan (OHP), created by one of the first federal waivers of traditional Medicaid rules, has two separate parts. First is the "OHP Plus." It serves low-income children, pregnant women, welfare recipients, and poor elderly and disabled populations groups who are

[16]The federal government started to defund Oregon's Medicaid Expansion from 2016 which has led to a budget deficit and Oregon Measure 101 − a two-year budget fix to close state budget deficit by taxing hospital and insurance agencies − is nearing to end in 2020.

categorically eligible Medicaid populations in Oregon (Office for Oregon Health Policy and Research, 2009). Second is the "OHP Standard." It servers poor adults who are financially but not categorically eligible for the Plus program. Eligibility for the Standard plan is limited to adults ages 19−64 who are Oregon residents and U.S. citizens or legal immigrants, and have incomes below the 100% federal poverty level and/or who have been without health insurance for at least six months, and/or have less than $2,000 in assets (Office for Oregon Health Policy and Research, 2009; Allen et al., 2010).

Except for vision and non-emergency dental services, the OHP Standard provides relatively comprehensive benefits with no consumer cost-sharing. The OHP Standard coverage includes physician services, prescription drugs, all significant hospital benefits, behavioral health, and chemical dependency services (including outpatient services), hospice care, and some durable medical equipment (Finkelstein et al., 2012; Baicker and Finkelstein, 2011). In 2001−2004, the average annual Medicaid expenditures for an individual on the OHP Standard were about $3,000, with monthly premiums that ranged below $20 depending upon income and was $0 for those below 10% of the federal poverty level (Wallace et al., 2008).

In early 2002, OHP Standard covered nearly 110,00 people, but in 2004, a budgetary shortfall halted new enrollment in the OHP Standard; and by early 2008, attrition had reduced enrollment to about 19,000. However, in early 2008, the state of Oregon had the budget to enroll an additional 10,000 adults. Despite this newfound budget, the demand for the program among eligible individuals would far exceed the 10,000 available slots. Therefore, Oregon's Department of Human Services applied for and received permission from the Centers for Medicare and Medicaid Services to add new members through random lottery draws from a new reservation list (Finkelstein et al., 2012).

In early 2008, the state of Oregon campaigned an extensive public awareness program about the lottery opportunity focusing on the group that was not categorically eligible for the Plus program. Any qualified person could sign up from January 28 to February 29, 2008, by telephone, fax, in-person sign-up, mail, or online while providing very little demographic information. The sign up form required minimal demographics information such as sex, date of birth, address, telephone number, P.O. box, and preferred language of communication (either English or Spanish) along with the list of names, sex, and date of birth of anyone age nineteen and older in the household whom they wished to add to their sign up form (Allen et al., 2010).

No attempts were made to verify the information or to adequately screen for program eligibility at sign up for the lottery in order to keep the entry barrier low. During the window from January 28 to February 29, 2008, a total of 89,824 individuals signed up. Ineligible individuals for the OHP Standard were excluded before the lottery. The exclusion comprises individuals residing outside of Oregon, individuals born before 1944 or after 1989, individuals with the OHP standard plan as of January 2008, individuals with an institutional address, and individuals who sign up by an unrelated third party (Allen et al., 2010).

This exclusion leads to a sample that comprises 74,922 individuals (representing 66,385 households). After the sign-up phase, the state of Oregon conducted eight lottery drawings (occurred from March through September 2008) and randomly selected 29,834 individuals, and the remaining 45,088 individuals were kept as a control group.

Lottery selectees were sent a two-page application form[17]. Up to eight supplemental forms could accompany it (Allen et al., 2010). The selected individual was eligible to apply for OHP Standard for themselves and their family member (whether listed or not) and was required to submit the paperwork within 45 days. If they met the eligibility requirements, they could enroll in the Oregon Health Plan (OHP) Standard indefinitely. However, they had to verify their status every six months.

About 60% of the people who were selected by the lottery sent back the application. Half of those applications failed to meet the requirements. The primary reason was the requirement of income in the last quarter, corresponding to annual income below the poverty level. The federal poverty line in 2008 was \$10,400 for a single person and \$21,200 for a family of four (Allen et al., 2010). Therefore, about 30% of the total selected individuals successfully enrolled in the OHP Standard. Shortly after the random assignment of lottery and OHP Standard application form, an "initial survey" was conducted and followed by the "main survey" a year later. These surveys consist of data for 58,405 individuals comprising 29,589 individuals in treatment, and 28,816 individuals in the control group.

3 Approaches to Health Insurance & Health Outcomes

How does health insurance affect health? The answer seems obvious, but Levy and Meltzer (2008) review the literature and draw three conclusions. First, the problem of endogeneity makes causal claims tenuous. Second, the papers that establish causal evidence are focused on small subgroup populations. For example, public health insurance reduces mortality among infants and children (Currie and Gruber, 1996a,b; Hanratty, 1996), while for the elderly, public health insurance improves different outcomes but not mortality (Card and Maestas, 2008; Finkelstein and McKnight, 2008; McWilliams et al., 2007b,a). Third, the nature of these studies is not representative of the broader population, which prohibits generalizing for policy purposes. In this paper, I provide causal claims of the effects of Medicaid that qualify for subgroups and also allow results to generalize in out-of-samples.

[17] "The main form asked for the names of all household members applying for coverage and inquired about their Oregon residence, U.S. citizenship, insurance coverage over the past six months, household income over the past two months, and assets. Documentation of identity and citizenship and proof of income had to be returned with the completed form" (Allen et al., 2010).

Allen et al. (2010) point out three practical designs for insurance and health outcomes research: observational studies, quasi-experimental studies, and randomized experiments. Observational studies comprise the most substantial part of the literature. Most of these studies typically utilize "multivariate regression" approaches. When implemented correctly, these approaches control the observable confounding variables between health insurance and health outcomes. However, these approaches are less likely to address the issues of unobservable confounders between health insurance and health outcomes. Failure to control unobservable differences between the insured and the uninsured may drive the observed differences in health outcomes (Levy and Meltzer, 2004, 2008), which could lead to biased estimations.

The second set of studies exploit natural experiments to evaluate the effect of health insurance on health outcomes. These studies implement techniques like differences-in-differences estimations, regression discontinuity designs, and instrumental variables. These techniques exploit an exogenous event that results in variation within health insurance coverage – changes that are plausibly unrelated to health and other underlying determinants of health insurance coverage (Levy and Meltzer, 2008). Exploiting an exogenous event makes the variation of the health insurance coverage take-up as good as random. In other words, health insurance coverage varies in a way that is unrelated to the unobservable factor. Thus a comparison of various outcomes between insured and uninsured are likely to support a causal interpretation.

However, the results of natural experiments are valid for only specific population groups; thus, they cannot be generalized to the broader population. As explained earlier, several studies show that public health insurance reduces mortality among infants and children (Currie and Gruber, 1996a,b; Hanratty, 1996), while for the elderly, public health insurance does not reduce mortality (Card and Maestas, 2008; Finkelstein and McKnight, 2008; McWilliams et al., 2007b,a). The "one size fits all" policy approaches are unlikely to be useful for the broader population. For example, the channels or mechanisms through which having insurance affects health outcomes may be different for infants and children than they are for elderly adults.

The third set of studies are social experiments, which are the "gold standard" for establishing causality. The RAND Health Insurance Experiment (RAND) and the Oregon Health Insurance Experiment (OHIE) are only two of such kind in the United States. Newhouse (1994) provides details on the RAND experiment while Finkelstein et al. (2012) offer details on the Oregon experiments. Using RAND experiment data, Newhouse (1994) and Brook et al. (1983) find no significant effect of insurance on the health status of an average adult. Levy and Meltzer (2008) point out a weakness of the RAND experiment that it did not randomize people to receive any health insurance. Instead, random individuals have treated with health insurance with varying degrees of generosity. Finkelstein et al. (2012) study the Oregon health insurance experiment data. They find statistically significant higher health care utilization, lower out-of-pocket medical expenditures and medical debt, and better self-reported physical and mental health among the treatment group.

The observational studies, quasi-experimental studies, and randomized experiments often focus on causal inference and have been dominant in empirical policy research in health economics as well as economics in general. Recently, due to the availability of big-data and computing powers, machine learning approaches are gaining momentum among researchers and policymakers. Several scholars like Varian (2014), Mullainathan and Spiess (2017), and Athey (2018) have promoted the value of the big-data and machine learning method in the field of economics. Within the domain of machine learning in economics, two strands of literature are gaining momentum: machine learning for policy prediction problems and machine learning for causal inference problems.

The machine learning algorithms behave well for out-of-sample prediction as it utilizes flexible model selection, model ensembles, high dimensional data environment, and cross-validations. Therefore these algorithms are useful in many policy applications[18] where the causal inference is not central or potentially unnecessary. However, machine learning algorithms are not well suited for causal inference. Rather than just correctly predicting out-of-sample, establishing causal effect relates to understanding the counterfactual − what would happen with and without a policy (Athey, 2018). However, some slight modifications of "off-the-shelf" or readily-available machine learning algorithms can utilize the strengths and innovations of machine learning algorithms for causal inference. The predictive machine learning algorithms are readily available with the open-source routines for statistical software like Python and R.

The approaches that use machine learning methods for causal inference focus on estimating the average treatment effect, heterogeneous treatment effects, and optimal policies (Athey, 2018). In Appendix A, I provide a summary of these approaches. This paper implements a causal machine learning approach, the "generalized random forest" of Athey et al. (2019), to explore the heterogeneous treatment effects of expanding access to public health insurance on various dimensions of healthcare utilization, personal finance, health, and wellbeing. Then, I utilize efficient policy learning strategies of Athey and Wager (2019a) to explore some strategies that can help to reform access to public health insurance programs.

[18] For example, Kleinberg et al. (2015) consider a resource allocation problem in health policy in which a policymaker needs to decide which otherwise-eligible patients should not be given hip replacement surgery through Medicare. They predict the probability that a candidate for a joint replacement would die within a year from other causes. They then identify patients who are at particularly high risk and should not receive joint replacement surgery. Similarly, Henderson et al. (2012) use satellite data on lights at night to predict economic growth, and Glaeser et al. (2018) use Google Street View images to predict income in New York City. Glaeser et al. (2016) develop a system for allocating health inspectors to restaurants in Boston, and Naik et al. (2016) quantify the "urban appearance" from street-level imagery for 19 American cities and establish an empirical connection between the physical appearance of a city and the behavior and health of its inhabitants.

4 Empirical Strategy

4.1 Identification

Finkelstein et al. (2012) provides the most detailed explanations and analyses of the Oregon Health Insurance Experiment. They give the Intent-to-Treat (ITT) and the Local Average Treatment Effect (LATE) estimates for various outcome variables using the data from the "main survey" along with several other data sources. Shortly after the lottery assignment, – that allowed lucky Oregonians to apply for the OHP Standard Medicaid, and an "initial survey" was conducted to collect information from those that participate in the application. A year later, a follow-up survey or the "main survey" was performed. Therefore, the "initial survey" is pre-treatment, and the "main survey" is a post-treatment survey. These surveys consist of data of 58,405 individuals comprising 29,589 individuals in treatment, and 28,816 individuals in the control group.

Analyses in this paper consider similar outcome variables as Finkelstein et al. (2012). However, the interpretations are very distinct compared to the approach of Finkelstein et al. (2012). This paper contemplates a situation where analysts know their outcome variable (Y) at the post-treatment and have data of observables (X) at the pre-treatment period. Therefore, the sample in this study may not be independent because the covariates are all drawn from the "initial sample" and merged to the outcome variables that are from the "main survey" sample. For this reason, this paper analyzes the data as an observational, rather than a genuinely randomized study. This paper assumes unconfoundedness to identify causal effects. Unconfoundedness means that treatment assignment is as good as random conditional on observable covariates (Rosenbaum and Rubin, 1983).

Consider $i \in \{1, \ldots, N\}$ observations where the potential outcomes for each unit is either $\{Y_i(0), Y_i(1)\}$. Following Rosenbaum and Rubin (1983), the unit level causal effect is the difference in potential outcomes $\tau_i = Y_i(1) - Y_i(0)$, where, $W_i \in \{0, 1\}$ is a binary indicator for the treatment with $W_i = 0$ indicating that unit i did not received the treatment and $W_i = 1$ indicating that unit i received the treatment. X_i is a k-component vector of features or covariates unaffected by the treatment. The data consist of triple $\left(Y_i^{obs}, W_i, X_i\right)$, $\forall\ i = 1, \ldots, N$. The realized outcome for unit i is the potential outcomes corresponding to the treatment i.e. Y_i^{obs} is

$$Y_i^{obs} = Y_i(W_i) = \begin{cases} Y_i(0)\ if\ W_i = 0, \\ Y_i(1)\ if\ W_i = 1. \end{cases}$$

then, unconfoundedness can be formalized as:

$$\{Y_i\left(0\right), Y_i\left(1\right)\} \perp W_i | X_i.$$

4.2 Mean comparison of demographics

In this study, the outcome variables are health care utilization, preventive care utilization, financial strain, and health after a year of the OHP Standard or Medicaid experience. The treatment variable is lottery selection, and observable covariates comprise pre-treatment demographics. This paper begins the analyses by comparing the mean of control and treatment group demographics.

$$\tilde{x}_{i,h} = \gamma_0 + \gamma_1 W_{i,h} + \eta_{ih} \tag{1}$$

where $\tilde{x}$ is an observable demographic variable in the pre-treatment period, γ_0 is the mean of the control group and, γ_1 is the mean difference between the control and treatment group. One should expect γ_1 to be statistically zero for comparable control and treatment groups. The selected individuals were eligible to apply for OHP Standard for themselves and their family member (whether listed or not); therefore, standard errors are household-level clustered and heteroscedasticity-consistent.

4.3 Intent to Treat Effect of Lottery

Secondly, this paper estimates the "intent-to-treat" (ITT) effect of winning the lottery (i.e., the difference between treatment and controls). The ITT provides a causal assessment of the net impact of expanding access to public health insurance.

This paper utilizes the double-selection post-LASSO approach introduced by (Belloni et al., 2014b). This method is based on the"LASSO."[19] Under the assumption of sparsity[20], the double-selection post-LASSO approach select the observable confounders and covariates properly. Confounders are common-cause variables that affect both outcomes and treatments. Covariates are variables that might affect results but are not associated with anything else.

The double-selection post-LASSO procedure is comprised of the following steps (Belloni et al., 2014a). First, run LASSO of dependent variables on a large inventory of potential covariates to select a set of predictors for the dependent variable. Second, run LASSO of treatment variable (lottery) on an extensive list of potential covariates to choose a set of predictors for treatment. If the treatment is genuinely exogenous, one should expect this second step should not select any variables. Third, perform OLS

[19] The Least Absolute Shrinkage and Selection Operator (LASSO) is an appealing method to estimate the sparse parameter from a high-dimensional linear model is introduce by Frank and Friedman (1993) and Tibshirani (1996). LASSO simultaneously performs model selection and coefficient estimation by minimizing the sum of squared residuals plus a penalty term. The penalty term penalizes the size of the model through the sum of absolute values of coefficients. Consider a following linear model $\tilde{y}_i = \Theta_i \beta_1 + \varepsilon_i$, where Θ is high-dimensional covariates, the LASSO estimator is defined as the solution to $\min_{\beta_1 \in \mathbb{R}^p} E_n \left[(\tilde{y}_i - {}_i\Theta\beta_1)^2 \right] + \frac{\lambda}{n} \|\beta_1\|_1$, the penalty level λ is a tuning parameter to regularize/controls the degree of penalization and to guard against overfitting. The cross-validation technique chooses the best λ in prediction models and $\|\beta\|_1 = \sum_{j=1}^{p} |\beta_j|$. The kinked nature of penalty function induces $\hat{\beta}$ to have many zeros; thus LASSO solution feasible for model selection.

[20] The "sparse" outcome model means a model with a few meaningful covariates affect the average outcome.

regression of dependent variable on treatment variable, and the union of the sets of regressors chosen in the two LASSO implementations to estimate the effect of treatment on the dependent variable then correct the inference with usual heteroscedasticity robust OLS standard error.

$$Y_{i,h} = \beta_0 + \beta_1 W_{i,h} + x_{ih}\beta_2 + \varepsilon_{it} \tag{2}$$

where, β_1 is the main coefficient of interest and gives the average difference in (adjusted) means between the treatment group (the lottery winners) and the control group (those not selected by the lottery). β_1 is the impact of being able to apply for OHP Standard through the Oregon lottery (Finkelstein et al., 2012). The x_{ih} are selected from X_{it}, implementing the double-selection post-LASSO. x_{ih} includes the set of confounding variables that correlate with treatment probability (and potentially with the outcome) along with covariates that explain treatment and outcome. Therefore controlling these covariates helps to estimate the "unbiased" relationship between winning the lottery and the outcome.

4.4 Local Average Treatment Effect of Lottery

The ITT estimates from equation (2) provide the causal effect of winning the lottery to apply for the OHP Standard. Another interesting causal parameter would be the impact of actual OHP Standard Medicaid insurance coverage rather than just the impact of winning the lottery to be eligible for the OHP Standard (ITT). In other words, policymakers are interested in the causal effect of compliance to the lottery and not just winning the lottery. The "complier" is the subset[21] of individuals who obtain insurance from winning the lottery and who would not obtain insurance through the lottery selection. One way to retrieve this parameter is to utilize lottery selection as an instrument and perform a two-stage least square (2SLS). equation (3) represents the first stage equation and the second stage equation respectively.

$$\begin{aligned} Z_{i,h} &= \delta_0 + \delta_1 W_{i,h} + x_{ih}\delta_2 + \mu_{it} \\ Y_{i,h} &= \phi_0 + \phi_1 \widehat{Z}_{i,h} + x_{ih}\phi_2 + \nu_{it} \end{aligned} \tag{3}$$

where, $W_{i,h}$ is an instrumental variable of lottery assignment; $Z_{i,h}$ is an endogenous binary variable that takes a value of 1 if an individual is "ever in Medicaid" during the study period (from initial

[21] Imbens and Angrist (1994) point out that there exist four possible groups of individuals based upon the compliance types: complier, always-taker, never-taker, and defier. The "complier" is the subset of individuals who obtain insurance by winning the lottery and who would not obtain insurance without winning the lottery. Never takers are a subset of individuals who never get insurance even after winning the lottery. Always takers will get insurance regardless of the lottery. The defier insured themselves when they are in the control group, and don't take insurance when they are in the treatment group. So, always taker and defier have insurance though they are in the control group. The never taker and defier won't take insurance though they win the lottery.

notification period until September 2009), or 0 otherwise. The first stage equation provides $\widehat{Z}_{i,h}$, which is the predicted value of "ever in Medicaid." The main coefficient of interest is ϕ_1 and is interpreted as a local average treatment effect (LATE) of Medicaid insurance (Imbens and Angrist, 1994) and identifies the causal impact of insurance among the "compliers." For just identified model, the LATE estimates, ϕ_1, is the ratio of ITT estimates from equation (2) and the first-stage coefficient on winning the lottery from equation (3) or $\phi_1 = \frac{\beta_1}{\delta_1}$ (Finkelstein et al., 2012). Relative to the study population, "compliers" are somewhat older, more likely white, in worse health, and in lower socioeconomic status (Finkelstein et al., 2012).

4.5 Heterogeneous Treatment Effects

Numerous studies examine the population average treatment effect of having an insurance. This effect can be formalize using a potential outcome framework as $\tau = E\left[Y_i(1) - Y_i(0)\right]$. However, this paper's main contribution is examining the heterogeneous treatment effect of Medicaid on several health and personal finance related outcomes. The treatment heterogeneity can be expressed as the conditional average treatment effect (CATE) and can be formalized as $\tau(x) \equiv E\left[Y_i(1) - Y_i(0)|X_i = x\right]$.

This paper employs the cluster-robust random forest approach of Athey and Wager (2019b) to access the treatment heterogeneity. This approach is based on the "causal tree" (Athey and Imbens, 2016), "causal forest" (Wager and Athey, 2018) and the "generalized random forest" (Athey et al., 2019) methods. The "causal tree" approach re-engineers the Breiman et al. (1984) classification and regression tree (CART)[22], a machine learning algorithms for causal inference. The remaining methods extend the "causal tree" approach utilizing the Breiman (2001) random forest[23] machine learning algorithm for causal inference.

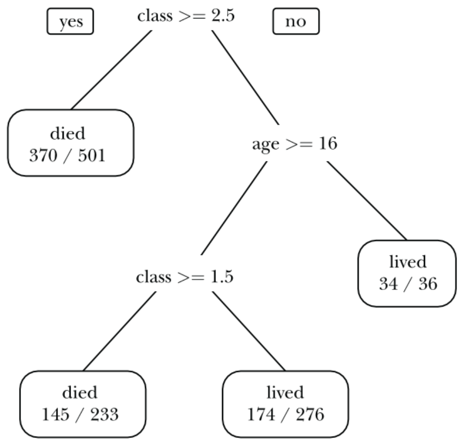
In essence, CART recursively filters and partitions the large data-set into binary sub-groups (nodes) such that the samples within each subset become more homogeneous in their fit of the response variable, thus resulting in a tree-like format. Figure 1 shows an example of features of the Titanic survivors using the CART method, as demonstrated by Varian (2014).

CART minimizes the mean-squared error of the prediction of outcomes to capture heterogeneity in outcomes. However, the "causal tree" minimizes the mean-squared error of treatment effects to capture treatment effect heterogeneity. The approach to estimate the "causal tree" is similar to the Imai and Ratkovic (2013) method. A sample is split into two halves. One half is used to determine the optimal partition of covariates space. The other half is used to estimate treatment effects based on the

[22] In simplest, the CART algorithm chooses a variable and split that variable above or below a certain level (which forms two mutually exclusive subgroups or leaves) such that the sum of squared residuals is minimized. This splitting process is repeated for each leave until the reduction in the sum of squared residuals is below a certain level (as defined by users), thus resulting in a tree format (Athey and Imbens, 2017b).

[23] The Breiman (2001) random forest ensembles or bootstrap and aggregate many CART and report the average.

Figure 1: A Classification Tree for Survivors of the Titanic



Source: Varian (2014).
Interpretation: The leftmost terminal node can be interpreted as, if the class of travel is more than 2.5 (a third-class accommodation), 370 out of 501 died. The rightmost terminal node can be interpreted as, out of 36 people of the age-cohort 16 or below who were in the first and second-class accommodation, 34 survived. Those who were age-cohort more than 16, if they were in second-class accommodation, 145 died out of 233 (second from the leftmost terminal node), while 174 out of 276 died if they were in the first-class accommodation (second from the rightmost terminal node). These rules fit the data reasonably well, misclassifying about 30 percent of the observations in the testing set.

optimal partition of covariates selected from the first partition (Athey and Imbens, 2016). This sample-splitting approach is known as an "honest" estimation because model training and model estimation are independent. This approach leads to loss of precision, as only half of the data is used to estimate the effect. However, this approach generates a treatment effect and a confidence interval for each subgroup that is valid no matter how many covariates are used in estimation. This paper employs the Chernozhukov et al. (2018a) cross-fitting approach, which will be covered later in this section, to prevent the loss of precision.

One caveat of the causal tree is that it does not provide personalized estimates. Wager and Athey (2018) utilize the "random forest" machine learning approach and propose a "causal forest" method, where many different causal trees are generated and averaged, which can provide personalized estimates. This method offers causal effects that change more smoothly with covariates and provides distinct individualized estimates and confidence intervals. Wager and Athey (2018) also provide an essential finding that the predictions from causal forests are asymptotically normal and centered on the true conditional average treatment effect for each individual. Athey et al. (2016) extend the approach to other models for causal effects, such as instrumental variables, or other models that can be estimated using the generalized method of moments (GMM). In each case, the goal is to evaluate how a causal parameter of interest varies with covariates.

4.6 Cluster-robust Random Forest

4.6.1 Random Forest

Essentially, the Breiman (2001) random forest approach makes prediction from an average of b CARTs or trees, as follow: (1) for each tree $b = 1, \ldots, B$, draw a subsample $S_b \subseteq \{1, \ldots, n\}$; (2) grow a tree via recursive partitioning on each such subsample of the data; and (3) make a prediction by averaging the prediction made by individual tree as:

$$\hat{\mu}(x) = \frac{1}{B} \sum_{b=1}^{B} \sum_{n=1}^{n} \frac{Y_i \mathbf{1}(\{X_i \in L_b(x), i \in S_b\})}{|\{i : X_i \in L_b(x), i \in S_b\}|} \tag{4}$$

where, $L_b(x)$ denotes the leaf of the b^{th} tree containing the training sample x. For out-of-bag prediction, one can estimate the average as $\hat{\mu}^{(-i)}(x)$ by only considering those trees b for which $i \notin S_b$. $(-i)$ superscript denote "out-of-bag" or "out-of-fold" prediction

4.6.2 *R*−Learner Objective Function

Nie and Wager (2017) showed that "$R-$learner" objective function for heterogeneous treatment effect estimation as

$$\hat{\tau}(\cdot) = \arg\min_{\tau} \left\{ \sum_{i=1}^{n} \left(\left(Y_i - \hat{m}^{(-i)}(X_i) \right) - \tau(X_i) \left(W_i - \hat{e}^{(-i)}(X_i) \right) \right)^2 + \lambda_n(\tau(\cdot)) \right\} \tag{5}$$

where, $\lambda_n(\tau(\cdot))$ is a "regularizer" that controls the complexity of the learned conditional average treatment effect $\hat{\tau}(\cdot)$ function. $e(x) = P[W_i | X_i = x]$ is the propensity score or probability of being treated; $m(x) = E[Y_i | X_i = x]$ is expected outcomes marginalizing over treatment; $(-i)$ superscript denote "out-of-bag" or "out-of-fold" prediction.

4.6.3 Causal Random Forest

As explained earlier, random forest ensembles of many trees and provides prediction as an average prediction made by many individual trees. Athey et al. (2019) show that a random forest can be equivalent as an adaptive kernel method and re-express the random forest from equation (4) as

$$\hat{\mu}(x) = \sum_{i=1}^{n} a_i(x) Y_i; \quad a_i(x) = \frac{1}{B} \sum_{b=1}^{B} \frac{Y_i \mathbf{1}(\{X_i \in L_b(x), i \in S_b\})}{|\{i : X_i \in L_b(x), i \in S_b\}|} \tag{6}$$

where, $a_i(x)$ is a data-adaptive kernel or simply weights that measure how often the i^{th} training example appears in the same leaf as the test point x. Causal forests can be seen as a forest-based method motivated by "$R-$learner". Causal forest has several tuning parameters[24] and the cross-validation on the "$R-$learner" objective function helps to select these tuning parameters. The kernel-based perspective on forests suggests a natural way to use them for treatment estimation by first growing a forest to get weights $a_i(x)$, and then set

$$\hat{\tau} = \frac{\sum_{i=1}^{n} a_i(x_i)\left(Y_i - \hat{m}^{(-i)}(X_i)\right)\left(W_i - \hat{e}^{(-i)}(X_i)\right)}{\sum_{i=1}^{n} a_i(x_i)\left(W_i - \hat{e}^{(-i)}(X_i)\right)} \tag{7}$$

Athey et al. (2019) discuss this approach in more detail, including how to design a splitting rule of a forest that will be used to estimate prediction via equation (7). At the implementation level, the causal forest starts by fitting two separate regression forests to estimate $\hat{m}(\cdot)$ and $\hat{e}(\cdot)$ and making out-of-bag predictions using these two first-stage forests. Then the model uses these out-of-bag predictions as inputs to the causal forest where cross-validation on the "$R-$learner" objective function, as given in equation (5), chooses the tuning parameters for the causal forest.

The random forests in this paper employs the Wager and Athey (2018) "honest" estimation, as explained earlier. Furthermore, the lottery assignment was to the household rather than to an individual. Therefore, this paper grows random forests by drawing a subsample at household level rather than individual-level. Similarly, the out-of-bag predictions are made using the household that was not in the training sample. equation (8) exhibits effectiveness of intervention, or Medicaid in individual, household, and global levels.

$$\hat{\tau}_h = \frac{1}{n_h}\sum_{\{i:H_i=h\}} \hat{\Gamma}_i, \quad \hat{\tau} = \frac{1}{H}\sum_{h=1}^{H}\hat{\tau}_h, \quad \hat{\sigma}^2 \frac{1}{H(H-1)}\sum_{h=1}^{H}\left(\hat{\tau}_h - \hat{\tau}\right)^2,$$
$$\widehat{\Gamma}_i = \hat{\tau}^{(-i)}(X_i) + \frac{W_i - \hat{e}^{(-i)}(X_i)}{\hat{e}^{(-i)}(X_i)\left(1 - \hat{e}^{(-i)}(X_i)\right)}\left(Y_i - \hat{m}^{(-i)}(X_i) - \left(W_i - \hat{e}^{(-i)}(X_i)\right)\hat{\tau}^{(-i)}(X_i)\right) \tag{8}$$

where, for the individual with household index $A_i \in \{1, \ldots, H\}$, the individual level effectiveness of lottery intervention is $\hat{\Gamma}_i$ and estimated based on the "doubly-robust" estimator with cross-fitting (Chernozhukov et al., 2018a). The household-level effectiveness of lottery intervention or the doubly-robust Average Treatment Effect (ATE) is $\hat{\tau}_h$. The global effectiveness of lottery intervention is $\hat{\tau}$ with standard error of $\hat{\sigma}^2$. The "doubly-robust" estimator is a variant of the augmented inverse-probability weighting. The name "doubly-robust" means in the sense that estimates are consistent whenever either the propensity fit, $\hat{e}(\cdot)$, or the outcome fit, $\hat{m}(\cdot)$, is consistent, and are asymptotically efficient in a semiparametric specifications. The cross-fitting, as suggested by Chernozhukov et al. (2018a), is similar to the Athey

[24] These tuning parameters include the number of variables to try for each split, number of trees grown in the forest, a target for the minimum number of observations in each tree leaf, number of minimum node size for tree.

and Imbens (2016) "honest" estimation. A sample is split into two halves. The first half (main sample) is used to determine the optimal partition of covariates space. The second half (auxiliary sample) is used to estimate treatment effects within the leave based on the optimal partition of covariates selected from the first partition. Then flip the role of the main and auxiliary samples. Each of the estimates is "honest" or the two estimators will be approximately independent, so simply averaging them offers an efficient procedure (Chernozhukov et al., 2018a).

4.6.4 Assessing Treatment Heterogeneity

A heuristic approach to gain qualitative insights about the strength of heterogeneity is to see how different are the doubly-robust average treatment effects for the subgroup whose out-of-bag CATE estimates are below or above median CATE (Athey and Wager, 2019b). Davis and Heller (2017) and Athey and Wager (2019b) have used this approach to test for heterogeneity.

However, another test is based on "best linear predictor" or BLP method of Chernozhukov et al. (2018b). First test if the model is calibrated or not, and second, test for the existence of treatment heterogeneity. For this Chernozhukov et al. (2018b) suggest to create three variables: $B_i = Y_i - \hat{y_i}^{(-i)}$; $C_i = \bar{\tau} W_i - \bar{\tau} \hat{e_i}^{(-i)}$; and $D_i = (\hat{\tau}^{(-i)}(X_i) - \bar{\tau})(W_i - \hat{e_i}^{(-i)})$. $\bar{\tau}$ is out-of-bag ATE, and $\hat{e_i}^{(-i)}$ is out-of-bag propensity score.

The mean forest prediction or regressing B_i and C_i, should yield $\frac{dB_i}{dC_i} = 1$. A coefficient of one for mean forest prediction (MFP) suggests that the mean forest prediction is correct. Next, the differential forest prediction (DFP), or regressing B_i and D_i, if $\frac{dB_i}{dC_i} = 1$, it suggests that the forest has captured heterogeneity in the underlying signal. The $p-$value of the DFP coefficient also acts as an omnibus test for the presence of heterogeneity: if the coefficient is significantly greater than 0, then one can reject the null of no heterogeneity. However, asymptotic results justifying such inference are not presently available.

4.7 Estimation of Treatment Policies

The optimal policy estimation has received greater attention in the machine learning literature[25] (Athey, 2018). The optimal policy function maps the observable characteristics of an individual to a policy or treatment assignment. In simplest, the main goal of optimal policy estimation is to answer "who should be treated?" or the optimal treatment allocation. The understanding of optimal policy is essential in policymaking because an ad-hoc targeting of a specific subpopulation with positive interventions can be unfair, unethical, illegal, and politically incorrect to some other subpopulations while intervening

[25] See Strehl et al. (2010); Dudík et al. (2011); Li et al. (2012); Dudk et al. (2014); Swaminathan and Joachims (2015); Jiang and Li (2015); Thomas and Brunskill (2016) and Kallus (2018).

everyone in the population (a blanket policy) is welfare-maximizing but can be extremely costly.

The optimal policy estimation, or optimal treatment allocation, has been recently studied in using causal machine learning in economics, mainly by Kitagawa and Tetenov (2018) and Athey and Wager (2019a). The main idea is to select a policy function that minimizes the loss from failing to use the ideal policy, referred to as the regret of the policy. Note that estimating conditional average treatment effect or heterogeneous treatment effect focus on the squared-error loss while the optimal policy estimation focuses on utilitarian regret (Athey and Wager, 2019a).

Once a policymaker understands the heterogeneity effect, they would like to assign the correct treatment to each individual or subpopulation. For that, I implement the Athey and Wager (2019a) strategy to find the policy function π that can map the observable characteristic of individuals, X_i, to an available set of treatment, W_i.

$$\pi : X_i \rightarrow W_i \in \{+1, -1\}$$

Note, $W_i \in \{1, 0\}$ is reindexed as $W_i \in \{+1, -1\}$ which will help to formulate an optimal policy assignment strategy later. Then an optimal treatment assignment policy can be given as π^* that maximizes expected utility, in our case, the health outcomes.

$$\pi^* \in \arg \max_{\pi \in \prod} E\left[Y_i\left(\pi\left(X_i\right)\right)\right]$$

Alternatively, any other non-optimal policy experiences the regret of $R(\pi)$, and we would like to minimize the regret function:

$$R(\pi) = E\left[Y_i\left(\pi^*\left(X_i\right)\right)\right] - E\left[Y_i\left(\pi\left(X_i\right)\right)\right] \tag{9}$$

Under unconfoundedness, the overlapping assumptions and binary treatment assignment Athey and Wager (2019a) propose a technique to estimate the regret, regret convergence, and bound of the regret. They first determine the treatment effect, $\hat{\Gamma}_i$, for each i using the double-robust estimation technique called double machine learning of Chernozhukov et al. (2018a) and given in equation (8).

Equation (8) is a doubly-robust estimator because only one of $\hat{\mu}$ or $\hat{e}$ needs to be correctly specified, and the term double machine learning is used because $\hat{\mu}$ and $\hat{e}$ can be semi- or non-parametric estimators. If the estimate is a positive treatment effect $\hat{\Gamma}_i$, I assign individual to treatment $(\pi(X_i) = 1)$ and if not then I assign individual to control $(\pi(X_i) = 0)$ and penalize for mismatch and maximize the following Q function to assess the effective policy:

$$\hat{Q}(\pi) = n^{-1} \sum_i \pi\left(X_i\right) |\hat{\Gamma}_i| sign\left(\hat{\Gamma}_i\right)$$

Further, Athey and Wager (2019a) show that the regret has $\sqrt{n}\left(\hat{R}_{DML}(\pi) - R(\pi)\right) \xrightarrow{d} N\left(0, \sigma^2(\pi)\right)$ convergence and is bounded with the order of $\sqrt{VC(\Pi)/n}$ where $\hat{R}_{DML}(\pi)$ is the double machine learning estimates of regret. The bound provides a robust theoretical prediction that the test-error on any out-of-sample data is upper bounded with the sum of training error and $\sqrt{VC(\Pi)/n}$.

5 Results

The analysis presented in this paper utilizes data from the "initial survey" and the "main survey." The "initial survey" (administered shortly after random assignment of lottery and mailing of the OHP Standard application form to the lottery selectee) and the "main survey" (conducted a year after the random assignment of the lottery) collect data from very similar questionnaire from 58,405 individual comprising 29,589 individuals in treatment and 28,816 individuals in the control group. Each of these individuals is adults of ages $19-64$ who are Oregon residents, the U.S. citizens, or legal immigrants without health insurance for at least six months, and/or are below the federal poverty level and/or have assets below \$2,000.

5.0.1 Pre-treatment Comparison of Demographic Characteristics

Employing equation 1, Table 1 begins the analysis by presenting how different are treatment and control groups in their demographics in the pre-treatment period. These demographics are retrieved from the lottery list data and the initial survey data. Table 1 illustrates the mean of the control group and the difference of means between the treatment group and the control group. Given the random assignment of insurance, one should expect that the mean of the treatment and control group should be statistically similar. Except for a few variables, the differences in the means between treatment and control group are statistically zero. There exist some anomalies where the mean difference of few demographics are statistically nonzero, but close to zero, which could be due to the large sample size. This evidence suggests that treatment or lottery was assigned randomly.

5.1 ITT, LATE and Heterogeneous Treatment Effects

The treatment effect often varies with individuals' observable characteristics. For example, if the treatment is costly and less accessible, then only those who are likely to benefit most will take the treatment. In this case, the availability of the treatment may reduce the average effect among the treatment recipients. While, on the other hand, if the treatment provided to the individuals who are less likely to benefit, then the availability of the treatment may increase the average effect among the treatment recipients. Therefore understanding the heterogeneity in treatment effects has important implications

Table 1: Pre-treatment Comparison of Demographic Characteristics

Variable	Control mean	Mean diff	Variable	Control mean	Mean diff
% Female §	0.600	-0.015*** (0.006)	% dont currently work	0.527	-0.007 (0.008)
% English preferred §	0.921	-0.009** (0.004)	% work below 20 hours/week	0.096	-0.002 (0.005)
% Self signup §	0.880	-0.045*** (0.004)	% work 20–29 hours/week	0.111	-0.003 (0.005)
% Signed up on first day §	0.102	0.004 (0.004)	% work 30+ hrs/week	0.266	0.012* (0.007)
% PO Box address §	0.127	0.000 (0.005)	% income the FPL below 50%	0.436	-0.029*** (0.009)
% MSA §	0.750	-0.004 (0.006)	% income the FPL 50–75%	0.125	0.005 (0.006)
Age (as of 2008) §	42.33	-0.108 (0.169)	% income the FPL 75–100%	0.154	0.000 (0.006)
% Race as White	0.838	-0.009 (0.006)	% income the FPL 100–150%	0.171	0.012* (0.007)
% Race as Black	0.031	-0.001 (0.003)	% income the FPL above 150%	0.114	0.011* (0.006)
% Race as Spanish/Hispanic/Latino	0.100	0.009* (0.005)	% Insurance	0.293	0.145*** (0.008)
% 4-year college degree or more	0.113	0.000 (0.005)	% OHP	0.067	0.158*** (0.006)
% High school diploma or GED	0.506	-0.007 (0.008)	% Private insurance	0.028	-0.002 (0.003)
% Less than high school	0.168	0.002 (0.006)	% Other insurance	0.055	0.00 (0.004)
% Vocational training or 2-year degree	0.212	0.004 (0.007)	Household size	2.884	0.094*** (0.029)

Notes: The initial survey consists of data of 58,405 individual comprising 29,589 individuals in the treatment group and 28,816 individuals in the control group. The variables collected from the lottery list for the population that appeared in the “initial survey” are marked with §. Enclosed in the parenthesis are household-level clustered heteroscedasticity-consistent standard errors. The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. the FPL represents the FPL; in 2008, it was $10,400 for a single person and $21,200 for a family of four Allen et al. (2010). The variables presented in this table are similar to Finkelstein et al. (2012) paper. However, these estimates are different from theirs. They compare the means of treatment and control group using lottery list data (marked as §) for the observation of $n = 74922$ and the “main survey” data while this table utilizes “initial survey” data.

for policymakers, mainly to yield valuable insights about how to distribute scarce social resources in an unequal society (Xie et al., 2012) by balancing the competing policy objectives, such as reducing cost, maximizing average outcomes, and reducing variance in outcomes within a given population (Manski, 2009).

As noted earlier, this paper contemplates a situation where analysts know their outcome variable, (Y), at post-treatment and have data of observables, (X), at the pre-treatment period. For this reason, this paper analyzes the data as an observational rather than a genuinely randomized study. Therefore, treatment heterogeneity is likely because such a situation could arise if there are unobserved household-level features that are an essential treatment effect modifiers. For example, some households may have better access to care and probably implement the intervention better than others or may have the knowledge to utilize resources to benefit from the treatment.

To generalize the results outside the sample size, one needs to robustly account for the sampling variability of potentially unexplained household-level effects. This study takes a conservative approach and assumes that the outcome variables of an individual within the same household may be arbitrarily correlated within a household (or "cluster"); therefore, it utilizes the cluster-robust analysis. Furthermore, to generalize beyond the household given in the data, each household is equally weighted such that the model allows the prediction of the effect on a new individual from a new household.

Tables 2, 3, 4 and 5 comprise various estimates for health care/preventive utilization, financial strain, self-reported health and potential mechanisms, respectively. These outcome variables are taken from the "main survey" and proxy the causal effects after one year of Medicaid experiences. Each of these tables has several estimates. The estimates in Column (1) represent "intent-to-treat" effects implementing double-selection post-LASSO method. Column (2) shows local average treatment effects, which can be interpreted as the impact of Medicaid among compliers. Column (3) presents the doubly-robust average treatment effect, which presents the average effectiveness of the lottery intervention on the outcomes.

For each Table, Columns (4), (5), and (6) explore the treatment heterogeneity. Column (4) provides "heuristic", or qualitative, insights about the strength of heterogeneity, and it groups the out-of-bag CATE estimates to above or below the median CATE estimate then estimates average treatment effects in these two subgroups separately using the doubly-robust approach to test if those average treatment effects are statistically similar or not. Columns (5) and (6) provide a test calibration for causal forest or the omnibus evaluation of the quality of the random forest-based on the "best linear predictor" method of Chernozhukov et al. (2018b). It computes the best linear fit of the target "estimand" using the forest prediction (on held-out data) as well as the mean forest prediction as to the sole two regressors. A coefficient of one for mean forest prediction (MFP) suggests that the mean forest prediction is correct, whereas a coefficient of one for differential forest prediction (DFP) additionally suggests that the forest has captured heterogeneity in the underlying signal. The $p-$value of the DFP coefficient also acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly higher than 0, then we can reject the null hypothesis of no heterogeneity. Though the treatment heterogeneity is not detected, this does not exclusively indicate the non-existence of treatment heterogeneity. Therefore, a heatmap plot is provided for a closer look at the location of heterogeneity.

The heatmap helps to characterize which subpopulations are more or less inclined to Medicaid. However, a heatmap is a partial representation of overall treatment heterogeneity. It requires caution while interpreting because it only presents two-dimensions: age in the x-axis and household income as a percentage of the FPL. Indeed, there may exist several variables that should be taken into consideration for proper interpretation of heterogeneous treatment effects. Appendix B provides a list of relevant variables to explain each of the heatmaps in this section.

5.1.1 Health Care Utilization

Table 2 Panel A describes health care utilization on extensive and intensive margins. The health care utilization extensive margin relates to if an individual is currently taking any medication, has any outpatient visits, has any emergency visits, or has any inpatient hospital admission in the last six months. While the health care utilization intensive margins quantify how many times an individual is currently taking medication, has outpatient visits, has emergency visits, has inpatient hospital admission in the last six months.

Table 2: Health Care Utilization

Outcome variables	ITT (1)	LATE (2)	ATE (3)	Heuristic (4)	MFP (5)	DFP (6)
Panel A: Health care utilization						
Extensive margins						
Currently taking any prescription medications	0.021**	0.067**	0.007	-0.018	0.801	-0.494
	(0.009)	(0.03)	(0.009)	(0.018)	(1.015)	(0.734)
Outpatient visits last six months	0.07***	0.224***	0.062***	0.055***	1.028***	1.316***
	(0.009)	(0.027)	(0.009)	(0.017)	(0.145)	(0.312)
ER visits last six months	0.009	0.029	0.005	-0.014	0.696	-3.331
	(0.008)	(0.024)	(0.008)	(0.015)	(1.172)	(1.816)
Inpatient hospital admissions last six months	0.002	0.005	0.001	-0.006	0.272	-0.626
	(0.004)	(0.014)	(0.005)	(0.009)	(2.322)	(1.4)
Intensive margins						
Number of prescription medications currently taking	0.104*	0.342*	0.042	-0.119	0.899	-0.383
	(0.055)	(0.177)	(0.055)	(0.109)	(1.219)	(1.005)
Number of Outpatient visits last six months	0.335***	1.087***	0.304***	0.426***	1.037***	1.502***
	(0.052)	(0.166)	(0.055)	(0.11)	(0.188)	(0.373)
Number of ER visits last six months	0.006	0.018	-0.003	-0.115***	1.97	-10.89
	(0.016)	(0.053)	(0.017)	(0.035)	(14.846)	(2.98)
Number Inpatient hospital admissions last six months	0.007	0.024	0.007	0.008	0.713	-2.071
	(0.007)	(0.021)	(0.007)	(0.014)	(0.661)	(1.974)
Panel B: Preventive care utilization						
Blood cholesterol checked (ever)	0.036***	0.116***	0.035***	0.00	1.043***	1.022*
	(0.008)	(0.026)	(0.008)	(0.016)	(0.236)	(0.73)
Blood tested for high blood sugar/diabetes (ever)	0.038***	0.121***	0.035***	0.003	0.982***	-1.588
	(0.008)	(0.025)	(0.008)	(0.017)	(0.235)	(1.618)
Mammogram within last 12 months (women 40 + age)	0.078***	0.249***	0.063***	0.048*	0.992***	2.036***
	(0.013)	(0.039)	(0.014)	(0.027)	(0.213)	(0.697)
Pap test within last 12 months (women)	0.053***	0.18***	0.047***	0.037*	1.003***	2.159***
	(0.01)	(0.034)	(0.011)	(0.022)	(0.23)	(0.671)

Notes: The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. Enclosed in the parenthesis are household-level clustered heteroscedasticity-consistent standard errors. The regressions in Columns (1) and (2) include household size dummies, survey wave dummies, and survey wave interacted with household size dummies. For the LATE estimates in Column (2), the instrumental variable is lottery assignment, and the endogenous variable is "Ever in Medicaid". The ITT and LATE estimates are base on the double-selection post-LASSO.

The ITT and LATE estimate in Table 2 Panel A shows that on both margins of the health care utilization, there are substantial and (mostly) statistically significant increases in prescription drugs and outpatient use. However, the doubly-robust ATE estimates illustrate a significant effect for the outpatient usages only. The average treatment effect of winning the lottery is associated with about a 0.30 (std. err. = 0.06) increase in the number of outpatient usages. Table 2 Panel B depicts the preventive care utilization. The ITT and ATE estimates are similar and statistically significant, suggesting that winning the lottery increases the likelihood for preventive cares like a blood test for cholesterol and diabetes, mammograms (for women of age 40+), or Pap tests (for women). However, these estimates are small in size and also do not shed light on the treatment heterogeneity. There is likely no effect among a particular subgroup, while another subgroup may be uniquely affected.

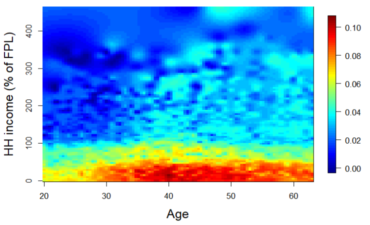
Table 2 Column (4) renders the heuristic approach to test the treatment heterogeneity. Evidence of treatment heterogeneity for outpatient usages and preventive care utilization is found. Table 2 Column (5) shows the MFP and Column (6) represents the DFP. The MFP and DFP are close to unit and statistically nonzero, suggesting treatment heterogeneity among these variables.

Note that 2000 causal trees were assembled to develop a cluster-robust random forest. Among these 2000 causal trees, the algorithm always selects the age and the household income below the federal poverty level along with the household size and other variables like education and employment. Appendix A provides the variable importance table for all of the outcomes analyzed. It lists the variables which were split (more than average) by the random forest. Only for illustration purpose of treatment heterogeneity, I develop a heatmap by grouping age and percentage of household income below the FPL and average the out-of-bag conditional average treatment. The heatmap has age on the x−axis and Household Income below the FPL (in percentage) on the y−axis.

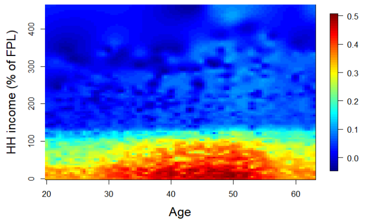
Figure 2 Panel (a) to (e) renders graphical depictions that compare the treatment and control group to exhibit the treatment heterogeneity for the outpatient usages and preventive care. Figure 2 Panel (a) and (b) portray an insight into outpatient utilization, CATE, over Age and household income. It appears that outpatient usage CATE (in extensive margin) for lottery winners is high and positive for those who belong to a household whose income lies below 100% of the FPL, regardless of age cohorts. The findings are similar for the intensive margin of outpatient usage CATE; however, there exist some additional heterogeneity for different age-cohorts.

Figure 2 Panel (c) exhibits treatment heterogeneity if the blood test for cholesterol level were ever done within the study period. Mostly younger age cohorts, between 20 to 40, who belong to a more impoverished household, have a higher likelihood of this preventive test. Figure 2 Panel (e) shows the treatment subgroup who are in a household below 80% of the FPL are more likely to the blood test for diabetes. Figure 2 Panel (e) and (f) illuminates CATE for the Mammogram test (for women whose age is above 40) and the Pap test (for women). It appears that women aged 40 years and above who belong

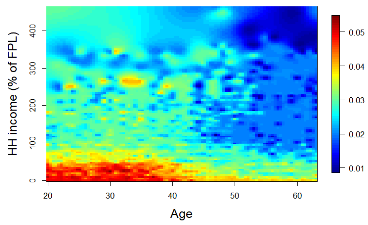
Figure 2: Health Care and Preventive Care Utilization



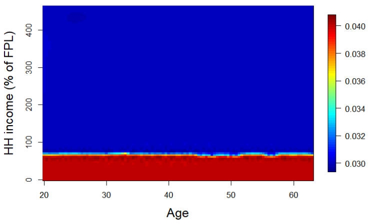
(a) Outpatient visits last six months



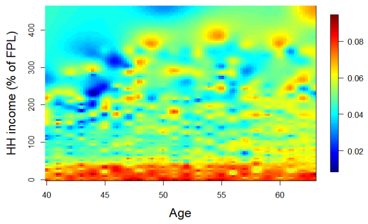
(b) Number of outpatient visits last six months



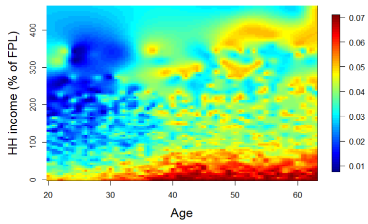
(c) Blood cholesterol checked (ever)



(d) Blood tested for high blood sugar/diabetes (ever)



(e) Mammogram within last 12 months (women 40+)



(f) Pap test within last 12 months (women)

Notes: The heatmap helps to exhibit which subpopulations are more or less susceptible to Medicaid. For each heatmap, age is in the x-axis and household income as a percentage of the FPL is in the y-axis. For each grid of x-axis and y-axis, the color maps the intensity of individualized treatment effect. However, a heatmap is a partial representation of overall treatment heterogeneity and requires caution to interpret. Indeed there may exist several variables which should be taken into consideration for proper interpretation of heterogeneous treatment effect. Appendix B provides relevant variables list to explain each of the heatmaps in this section.

below 50% of the FPL households are highly likely to elect to have a Mammogram test performed. Post 50 years, women are likely to have a Mammogram test regardless of the household income is below the FPL. The heatmap of the Pap test shows, women from households close to the FPL or below 100% the FPL are likely to participate for the Pap test.

5.1.2 Financial Strain

Table 3 displays extensive margins and intensive margins of financial strains. Winning the insurance lottery is associated with lower financial strains both in extensive and intensive margins. The ITT and ATE estimates for financial strains in intensive margins quantify the results in dollar terms as the net effect of winning the lottery. The ITT and ATE ranges describe that winning the lottery relates to reductions of various types of out-of-pocket costs for the past six months. The ITT and ATE estimate ranges depicts on average $20 reductions on out-of-pocket costs for doctors visits, clinics or health centers; nearly $40 to $49 reduction on out-of-pocket costs for emergency room or overnight hospital care; about $13 to $15 reduction on out-of-pocket costs for medical care and nearly about $50 reduction on the total out-of-pocket cost for medical care. Other than these financial strains, the group that received insurance through the lottery also has nearly \$450 to \$500 on average reduction of their medical debts.

Table 3: Financial Strain

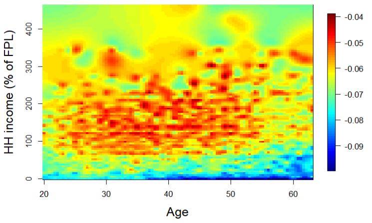
Outcome variables	ITT (1)	LATE (2)	ATE (3)	Heuristic (4)	MFP (5)	DFP (6)
Extensive margins						
Any out-of-pocket medical expenses, last six months	-0.073***	-0.238***	-0.073***	0.028	1.021***	1.449***
	(0.009)	(0.029)	(0.009)	(0.018)	(0.125)	(0.562)
Owe money for medical expenses currently	-0.053***	-0.17***	-0.058***	0.038**	1.076***	0.87
	(0.009)	(0.027)	(0.009)	(0.018)	(0.169)	(1.253)
Borrowed money or skipped other bills to pay medical bills, last six months	-0.057***	-0.184***	-0.064***	0.008	1.061***	0.473
	(0.009)	(0.028)	(0.009)	(0.017)	(0.145)	(1.323)
Refused treatment because of medical debt, last six months	-0.012**	-0.037**	-0.013***	0.006	1.054***	-3.706
	(0.005)	(0.015)	(0.005)	(0.009)	(0.387)	(2.121)
Intensive margins						
out-of-pocket costs for doctors visits, clinics or health centers, past 6 months	-19.308***	-61.429***	-20.175***	-8.47	0.999***	0.371
	(3.46)	(10.919)	(3.594)	(7.192)	(0.179)	(0.664)
out-of-pocket costs for emergency room or overnight hospital care, past 6 months	-49.519**	-157.71**	-40.73**	14.213	1.035**	0.211
	(21.611)	(67.674)	(18.46)	(36.89)	(0.468)	(0.689)
out-of-pocket costs for prescription medicine, past 6 months	-15.042**	-45.756**	-12.747**	2.234	0.889**	-1.116
	(6.941)	(22.054)	(6.012)	(12.067)	(0.403)	(1.405)
out-of-pocket costs for other medical care, past 6 months	-3.431	-10.577	-3.052	-7.223*	0.894*	-3.693
	(2.088)	(6.55)	(2.083)	(4.188)	(0.617)	(1.492)
Total out-of-pocket costs for medical care, last 6 months	-48.203***	-152.815***	-53.793***	13.3	1.034***	0.489
	(9.552)	(30.393)	(9.751)	(19.707)	(0.188)	(0.732)
Total amount currently owed for medical expenses	-442.39***	-1447.906***	-496.084***	167.277	1.038***	-0.298
	(96.744)	(318.1)	(105.023)	(208.674)	(0.223)	(1.125)

Notes: The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. Enclosed in the parenthesis are household-level clustered heteroscedasticity-consistent standard errors. The regressions in Columns (1) and (2) include household size dummies, survey wave dummies, and survey wave interacted with household size dummies. For the LATE estimates in Column (2), the instrumental variable is lottery assignment, and the endogenous variable is "Ever in Medicaid". The ITT and LATE estimates are base on the double-selection post-LASSO.

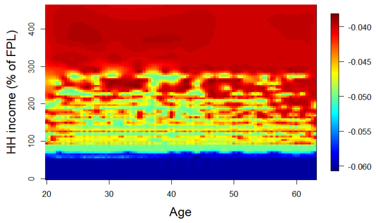
The "best linear prediction" (BPL) model narrates the treatment heterogeneity in the out-of-pocket expenses (last six months) only. Again, this does not necessarily mean that there is no heterogeneity because the BPL acts as an omnibus test for the presence of heterogeneity. A closer look at the heatmap in Figure 3 illuminates some sources of treatment heterogeneity.

The heatmap of Figure 3 Panel (a) shows a reduction for the extensive margin on the out-of-pocket medical expenses (last 6 months) suggesting lower financial strain for lottery winners of all age groups and all households but the effects are more pronounced for lottery winning households with income that ranges below 80% the FPL and belong to the age group of 40 years and above. Figure 3 Panel (b) exhibits a sharp discontinuity of owing money for medical expenses for lottery winning households with income below 100% the FPL. These differences suggest that at least within a low-income and relatively older population, individuals who select health insurance coverage are in poorer health (and therefore

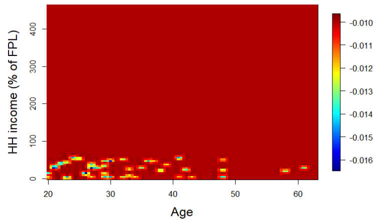
Figure 3: Financial Strain



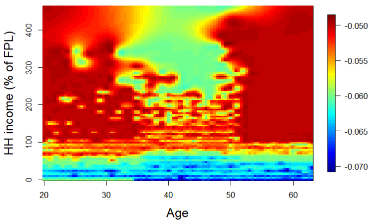
(a) Any out-of-pocket medical expenses, last 6 months



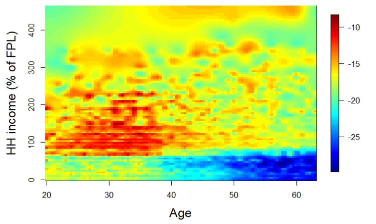
(b) Owe money for medical expenses currently



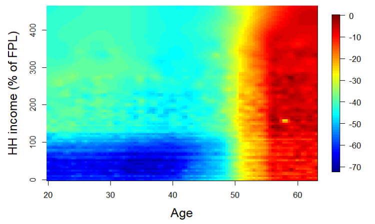
(c) Refused treatment because of medical debt, last six months



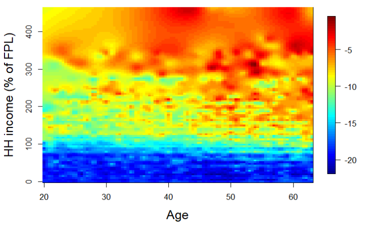
(d) Borrowed money or skipped other bills to pay medical bills, last 6 months



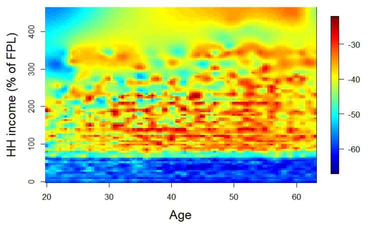
(e) out-of-pocket costs for doctors visits, clinics or health centers, past 6 months



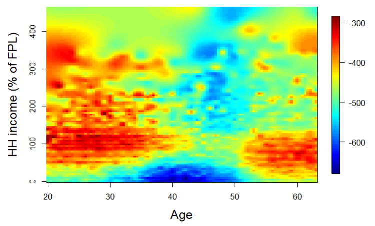
(f) out-of-pocket costs for emergency room or overnight hospital care, past 6 months



(g) out-of-pocket costs for prescription medicine, past 6 months



(h) Total out-of-pocket costs for medical care, last 6 months



(i) Total amount currently owed for medical expenses

Notes: The heatmap helps to exhibit which subpopulations are more or less susceptible to Medicaid. For each heatmap, age is on the x-axis, and household income as a percentage of the FPL is on the y-axis. For each grid of the x-axis and y-axis, the color maps the intensity of the individualized treatment effect. However, a heatmap is a partial representation of overall treatment heterogeneity and requires caution to interpret. Indeed there may exist several variables that should be taken into consideration for proper interpretation of heterogeneous treatment effect. Appendix B provides a relevant variables list to explain each of the heatmaps in this section.

demand more medical care) than those who are uninsured, just as standard adverse selection theory would predict Finkelstein et al. (2012).

Figure 3 Panel (c) shows no heterogeneity of being refused for treatment because of medical debt. Privately-owned hospitals may refuse patients in a non-emergency, but public hospitals cannot turn away patients. The Emergency Medical and Treatment Labor Act (EMTLA) enacted by Congress in 1986, explicitly prohibits public hospitals from denying care to indigent or uninsured patients even if they cannot pay.

Figure 3 Panel (d) shows that lottery winners have an overall reduction of borrowing money or skipping other bills to pay medical costs compared to the control group. However, the effect is more pronounced for lottery winning households with income below 100% of the FPL compared to a similar control group. These estimates are for the extensive margin only. The next figure exhibits some of the intensive margins of financial strains.

Figure 3 Panel (e) shows that more than \$25 to \$30 reductions of out-of-pocket costs (for doctors visits, clinics or health centers in past six months) for age group 50 plus who belongs to the lottery winning household with an income below 80% of the FPL compared to the control group. The below 40 age group from the lottery winning households within the range of 80% to 200% of the FPL have less than about \$15 reductions of such cost compared to the similar control group. The rest of the lottery winning subgroup has roughly an average of \$20 cuts of such cost, compared to the control group.

Figure 3 Panel (f) shows about \$60 to \$70 or more reduction in the out-of-pocket costs for emergency room or overnight hospital care in the past six months for the age group of 40 below of the lottery winning household with income below 100% of the FPL. The reduction of such costs is less than \$20 for the 50 and above age group, regardless of their household-level income status. The remaining subgroup of these aged below 50 who belong to a household with income more than 100% of the FPL has about \$30 to \$50 reductions in the costs of the out-of-pocket payments for emergency room visits or overnight hospital stays.

Figure 3 Panel (g) exhibits that the lottery winners who belong to the household with income below 100% the FPL (regardless of their age) report more than \$15 of reductions in the out-of-pocket costs for prescription medicine in past six months.

Figure 3 Panel (h) illuminates that the lottery winners who belong to a household with income below 100% the FPL (regardless of their age) have more than \$50 of reductions in the total out-of-pocket cost for medical care in last six months.

Figure 3 Panel (i) exhibits the decline of the total amount currently owed for medical expenses. Compared to the control group, the treatment group with the age of 35 to 50, have medical debt reductions. Such medical debt reductions are more pronounced (more than \$600) if the person belongs to a household with an income of 50% below the FPL.

As pointed out by Finkelstein et al. (2012), these results suggest that some of the financial benefits from Medicaid coverage can spillover beyond the insured. For example, the declines in out-of-pocket expenses and a reduction in the difficulty of paying non-medical bills means a reduction in the costs of unpaid care for medical providers. Furthermore, insurance can reduce extreme adverse shocks to consumption and can lead to consumption-smoothing.

5.2 Self-reported Health

Table 4 describes the effectiveness of the Oregon Health Insurance Experiment in the various dimensions of the perceived physical and mental health outcomes after a year. The ITT and ATE are similar and positive, suggesting lottery winners, on average, self-reported higher health in comparison with the control group. The LATE relates to the effect that is even higher for the compliance subgroup. There exists detectable treatment heterogeneity.

Table 4: Self-reported Health

Variables	ITT (1)	LATE (2)	ATE (3)	Heuristic (4)	MFP (5)	DFP (6)
Self-reported health good/very good/excellent (not fair or poor)	0.046***	0.15***	0.046***	0.032*	0.984***	1.485***
	(0.009)	(0.028)	(0.009)	(0.017)	(0.19)	(0.431)
Self-reported health not poor (fair, good, very good, or excellent)	0.033***	0.107***	0.033***	0.044***	1.036***	1.085***
	(0.006)	(0.019)	(0.006)	(0.012)	(0.188)	(0.316)
Health about the same or gotten better over last six months	0.035***	0.115***	0.039***	0.078***	1.086***	1.748***
	(0.008)	(0.026)	(0.008)	(0.016)	(0.223)	(0.437)
Number of days physical health good, past 30 days	0.557***	1.796***	0.602***	0.431	1.037***	1.011***
	(0.182)	(0.587)	(0.183)	(0.364)	(0.312)	(0.4)
Number days poor physical or mental health did not impair usual activity, past 30 days	0.432**	1.397**	0.454**	1.333***	1.157**	1.286***
	(0.198)	(0.641)	(0.197)	(0.392)	(0.511)	(0.421)
Number of days mental health good, past 30 days	0.741***	2.479***	0.806***	0.807**	1.041***	0.815***
	(0.209)	(0.675)	(0.207)	(0.411)	(0.27)	(0.311)
Did not screen positive for depression, last two weeks	0.024***	0.079***	0.027***	0.023	1.055***	0.657
	(0.008)	(0.027)	(0.008)	(0.017)	(0.338)	(0.81)

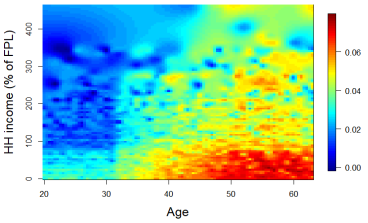
Notes: The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. Enclosed in the parenthesis are household-level clustered heteroscedasticity-consistent standard errors. The regressions in Columns (1) and (2) include household size dummies, survey wave dummies, and survey wave interacted with household size dummies. For the LATE estimates in Column (2), the instrumental variable is lottery assignment, and the endogenous variable is "Ever in Medicaid". The ITT and LATE estimates are base on the double-selection post-LASSO.

The survey has a self-reported health section. The responders had five options to choose (excellent, very good, good, fair, and poor) to report their health for different time frames. These are ordinal questions in nature, and there is no doubt that responders may have different perceptions of what good health represents for each individual. These options are recoded as binary for the self-reported health: good/very good/excellent to 1 and not fair or poor to 0.

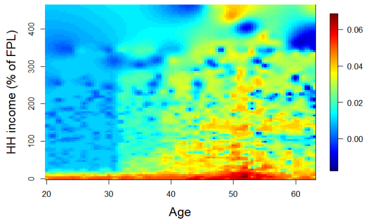
Figure 4 Panel (a) shows, compared to the control group, the lottery winning subgroup, those aged 40 and above, from a household whose income is below 100% of the FPL are more likely to report better health. Only the small subgroup of those aged 50 reported at least not poor health, as exhibited in Figure 4 Panel (b).

Figure 4 Panel (c) depicts heterogeneity for another question regarding the responder's perceptions of better or worse health outcomes throughout the last six months. The lottery winners from households whose income is below 70% of the FPL report better health to compare to the control group. When asked

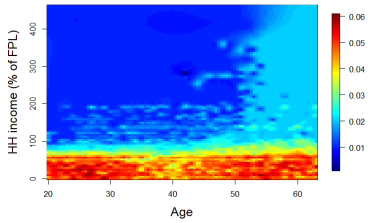
Figure 4: Self-reported Health



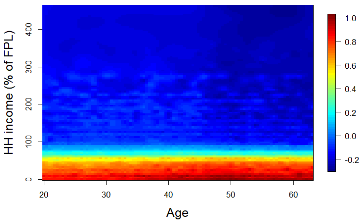
(a) Self-reported health good/very good/excellent (not fair or poor)



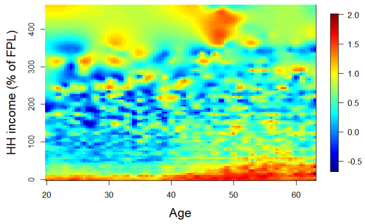
(b) Self-reported health not poor (fair, good, very good, or excellent)



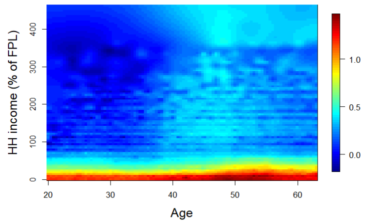
(c) Health about the same or gotten better over last six months



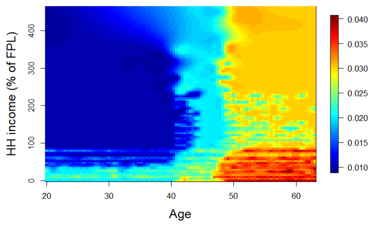
(d) Number of days physical health good, past 30 days



(e) Number of days mental health good, past 30 days



(f) Number days poor physical or mental health did not impair usual activity, past 30 days



(g) Did not screen positive for depression, last two weeks

Notes: The heatmap helps to exhibit which subpopulations are more or less susceptible to Medicaid. For each heatmap, age is on the x-axis, and household income as a percentage of the FPL is on the y-axis. For each grid of the x-axis and y-axis, the color maps the intensity of the individualized treatment effect. However, a heatmap is a partial representation of overall treatment heterogeneity and requires caution to interpret. Indeed there may exist several variables that should be taken into consideration for proper interpretation of heterogeneous treatment effect. Appendix B provides a relevant variables list to explain each of the heatmaps in this section.

to quantify the number of good physical health days in the past 30 days, lottery winning households closer to the FPL report higher numbers, as presented in Figure 4 Panel (d).

However, in Figure 4 Panel (e), the number of good mental health days in the past 30 days is reported to be higher for the age group above 40 from the lottery winning households closer to the FPL. The severity of mental and physical health is captured from the question to quantify the number of poor physical or mental health days did not impair the usual activity, past 30 days. Again, households closer to the FPL report higher numbers of days that were not impaired by poor physical and mental health, as plotted in Figure 4 Panel (f).

Figure 4 Panel (g) shows a group of those aged 50 and above who are from a household below 100% of the FPL are more likely to not be among those detected with depression (in the last two weeks). In all these Panels, it is repeatedly observed that lottery winning, poorer households report slightly better health outcomes compared to those who were not selected in the lottery. These results could arise due to adverse selection. As the theory suggests, those who are typically viewed as poorer/older require more health services than their counterparts. When they can receive that care, they will report better health outcomes compared to the groups that are unable to acquire that care.

5.3 Potential Mechanism for Improved Health

Table 5 depicts some potential mechanisms by which health insurance could have improved objective physical health along the heterogeneities in these mechanisms. Table 5 Columns (1), (2), and (3) present statistically significant increases of self-reported access to care (Panel A), quality of care (Panel B), and happiness (Panel C). Overall, the evidence suggests that people feel better off due to insurance, but Finkelstein et al. (2012) point-out that with the current data, it is difficult to determine the fundamental drivers of this improvement. One way to look at the drivers of this improvement is to capture the treatment heterogeneities. Except for the use of ER for a non-emergence (last six months), there are treatment heterogeneities in the access to care, quality of care, and happiness detailed in Table 5 Columns (4), (5), and (6).

Figure 4 illustrates the heatmap with age in the x−axis and percentage of household income below the FPL in the y−axis. The treatment effects are plotted for every possible grid of age and percentage of household income below the FPL. Figure 4 Panel (a) exhibits a particular threshold that households with income below the FPL 90% are more likely to have the usual place of clinic-based care than the control subgroup of similar attributes.

Figure 4 Panel (b) depicts households with income above 100% of the FPL with those aged 40 years and above are less likely to have a personal doctor compared to the households with income below 100% of the FPL with those aged under 40. Most of the poorer households are likely to get all their needed medical care (Figure 4 Panel (c)) and medications (Figure 4 Panel (d)) while households with income below 50% of the FPL and aged 40 age and above are less likely to utilize the ER in instance of non-emergencies (Figure 4 Panel (e)). Perceived quality of care is very uniformly distributed among the households and all age groups (Figure 4 Panel (f)). However, aged 40 and above in households with income below 180% of the FPL are more likely to have perceived happiness (Figure 4 Panel (d)).

Table 5: Potential Mechanism for Improved Health

Variables	ITT (1)	LATE (2)	ATE (3)	Heuristic (4)	MFP (5)	DFP (6)
Panel A: Access to care						
Have usual place of clinic-based care	0.087***	0.274***	0.086***	0.041**	1.012***	2.185***
	(0.009)	(0.029)	(0.009)	(0.018)	(0.109)	(0.736)
Have personal doctor	0.073***	0.235***	0.072***	0.101***	1.031***	1.329***
	(0.009)	(0.029)	(0.009)	(0.018)	(0.127)	(0.202)
Got all needed medical	0.085***	0.274***	0.085***	0.095***	1.019***	1.985***
	(0.009)	(0.028)	(0.009)	(0.017)	(0.106)	(0.332)
Got all needed drugs, last six months	0.07***	0.227***	0.073***	0.058***	1.016***	1.733***
	(0.008)	(0.026)	(0.008)	(0.016)	(0.112)	(0.416)
Didn't use ER for non emergency, last six months	0.00	0.00	0.003	-0.04***	1.163	-4.168
	(0.005)	(0.015)	(0.005)	(0.01)	(1.469)	(2.29)
Panel B: Quality of care						
Quality of care received last six months good/very good/excellent (conditional on any)	0.049***	0.15***	0.053***	-0.312***	1.028***	-402.796
	(0.01)	(0.03)	(0.01)	(0.019)	(0.179)	(19.252)
Panel C: Happiness						
Happiness, very happy or pretty happy (vs. not too happy)	0.062***	0.202***	0.069***	0.057***	1.049***	1.551***
	(0.009)	(0.029)	(0.009)	(0.017)	(0.134)	(0.379)

Notes: The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. Enclosed in the parenthesis are household-level clustered heteroscedasticity-consistent standard errors. The regressions in Columns (1) and (2) include household size dummies, survey wave dummies, and survey wave interacted with household size dummies. For the LATE estimates in Column (2), the instrumental variable is lottery assignment, and the endogenous variable is "Ever in Medicaid". The ITT and LATE estimates are base on the double-selection post-LASSO.

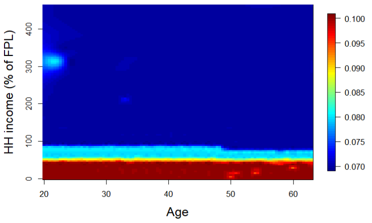
5.4 Efficient Policies

The previous section describes the ITT, LATE, and ATE along with the test of treatment heterogeneity. A more interesting question is whether we can find ways to prioritize treatment to some subgroups of Medicaid eligible registrants who are more likely to benefit from it. Following the out-of-bag prediction using generalized random forests of the Athey and Wager (2019b), I compute doubly-robust scores for the treatment effect as in equation **??**, and learning policies empirical maximization as in equation 9.

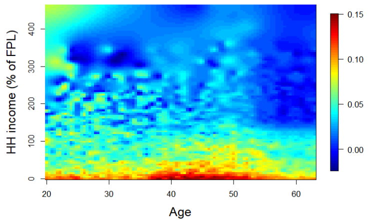
Table 6 Column (1) details the average outcome for each policy variable of interest under the random assignment of treatment. Table 6 Columns (2) to (5) present the estimates of the average outcome improvement (in percentage) of various policies over a random assignment baseline for the selected variable of interest. Efficient policy for each of the variables of interest uses a particular set of covariates, as given in appendix B. However, I did not use covariates like gender and race for the ethical and political rationale because these covariates cannot legally be used for treatment allocation.

In Table 6 Column (2) the assignment policy is based on a probability rule. The probability rule allocates Medicaid for those whose probability is less than the average probability of each outcome of interest. The generalized random forest provides the probability for each outcome of interest. In Table 6 Column (3), the assignment policy is the CATE rule, i.e., assign Medicaid if CATE is positive. In Table

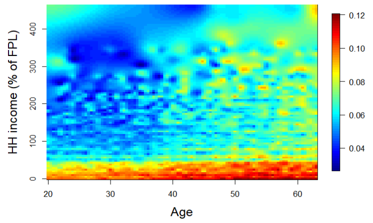
Figure 5: Potential Mechanism for Improved Health



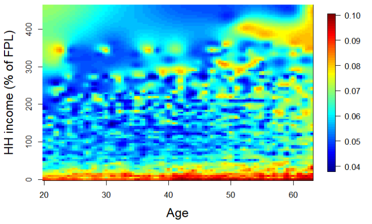
(a) Have usual place of clinic-based care



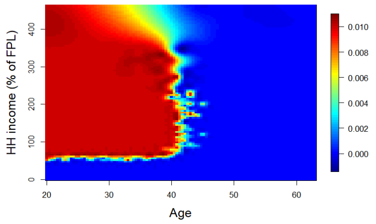
(b) Have personal doctor



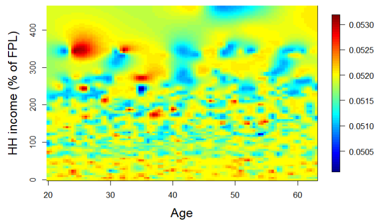
(c) Got all needed medical care, last six months



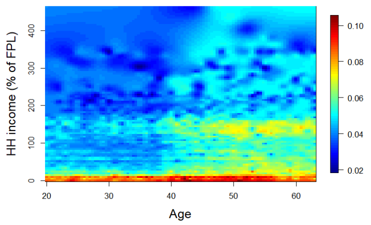
(d) Got all needed drugs, last six months



(e) Didn't use ER for non-emergency, last six months



(f) Quality of care received last six months good/very good/excellent (conditional on any)



(g) Happiness, very happy or pretty happy (vs. not too happy)

Notes: The heatmap helps to exhibit which subpopulations are more or less susceptible to Medicaid. For each heatmap, age is on the x-axis, and household income as a percentage of the FPL is on the y-axis. For each grid of the x-axis and y-axis, the color maps the intensity of the individualized treatment effect. However, a heatmap is a partial representation of overall treatment heterogeneity and requires caution to interpret. Indeed there may exist several variables that should be taken into consideration for proper interpretation of heterogeneous treatment effect. Appendix B provides a relevant variables list to explain each of the heatmaps in this section.

6 Columns (4) and (5), the shallow and deeper causal tree provides the Medicaid assignment policies. The shallow causal tree allows a max-depth of 3 policy trees while the deeper causal tree allows the max-depth of policy tree to be obtained by optimal pruning of the causal tree using cross-validation. Caution is warranted as asymptotic results hold only for trees with little complexity.

Table 6 Panel A, Column (1) describes the percentage of the households with outpatient visits over the last six months using the full sample data. About 60% of the whole sample has an outpatient visit in the previous six months. Note, this estimate is based on the lottery assignment of the OHP Standard or Medicaid. Panel A, Column (2) presents that if the Medicaid or OHP Standard is assigned among the

Table 6: Estimate of the utility improvement of various policies over a random assignment baseline.

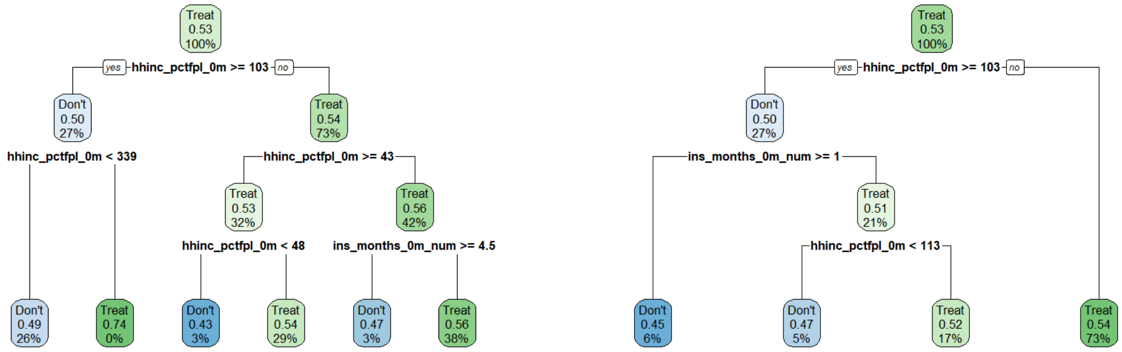
Variable	Baseline (1)	Probability rule (2)	CATE rule (3)	Shallow tree (4)	Deeper tree (5)
Panel A: Health care utilization					
Outpatient visits last six months	0.604***	4.74***	5.119***	4.228***	2.898***
	(0.002)	(0.182)	(0.17)	(0.197)	(0.177)
Panel B: Preventive care utilization					
Blood cholesterol checked (ever)	0.659***	0.575***	3.023***	1.934***	1.59***
	(0.005)	(0.176)	(0.146)	(0.166)	(0.154)
Blood tested for high blood sugar/diabetes (ever)	0.625***	1.066***	3.059***	2.665***	2.068***
	(0.003)	(0.157)	(0.124)	(0.137)	(0.178)
Mammogram within last 12 months (women + 40)	0.331***	7.008***	10.228***	9.26***	5.75***
	(0.002)	(0.482)	(0.398)	(0.552)	(0.42)
Pap test within last 12 months (women)	0.411***	3.489***	5.682***	4.955***	4.058***
	(0.003)	(0.286)	(0.24)	(0.315)	(0.316)
Panel C: Self-reported health					
Self-reported health good/very good/excellent (not fair or poor)	0.579***	1.952***	4.186***	4.225***	2.588***
	(0.003)	(0.174)	(0.145)	(0.201)	(0.195)
Panel D: Potential mechanism					
Have usual place of clinic-based care	0.558***	5.462***	7.44***	7.305***	4.718***
	(0.002)	(0.227)	(0.203)	(0.237)	(0.202)
Have personal doctor	0.544***	6.114***	6.432***	6.144***	4.576***
	(0.003)	(0.192)	(0.207)	(0.244)	(0.181)
Happiness, very happy or pretty happy (vs. not too happy)	0.629***	2.137***	4.883***	5.042***	3.306***
	(0.002)	(0.196)	(0.174)	(0.218)	(0.166)

Notes: The ***, **, and * represent 1%, 5%, and 10% level of significance, respectively. Enclosed in the parenthesis are standard errors. The estimates in Columns (1) represents the averages of each variable based on the random assignment baseline and considered as a parameter measuring the cost of treatment. The estimates in Column (2) to (5) presents the estimates of the average outcome improvement (in percentage) of various policies over a random assignment baseline for selected variable of interest. Policies learned on different subsets of the data will in general be different from the policies learned on the full data. Therefore, to examine the stability of the learned rule, 100 different policy are learned from randomly sample subdata and estimates are based on the out-of-bag sample.

eligible registrants using the probability rule, then it would improve outpatient visits by an additional 4.74%. Panel A, Column (3), exhibits, if the Medicaid assignment is based on the CATE rule, then it would improve outpatient visits by 5.12%. The optimal depth-3 policy tree or shallow tree would improve outpatient visits by an additional 4.23%. The optimal depth for policy trees based on the cross-validation for pruning would improve outpatient visits by an extra 2.9%. All of these improvements are statistically significant.

Figure 6 is a graphical depiction of the proposed efficient policy with the shallow tree in Panel (a) and deep tree in Panel (b). Note that the policies learned on different subsets of the data will, in general, be different from the policies acquired on the full data. It can be interesting to examine them to gain an intuition for the stability of the learned rule. Table 6 exhibits the stability of learned rule. However, Figure 6 is a graphical depiction of a learned policy and can vary for different subsets of the data. To save space, learned efficient policies for the rest of the variables that are presented in Table 6 are compiled in

Figure 6: Efficient policy to improve outpatient visits



(a) Shallow tree **(b)** Deep tree

Notes: The hhinc_pctthe FPL_0m shows household income as percentage of the federal poverty line in the baseline and the ins_months_0m_num shows numbers of months that a responder has insurance in last six months. Policies learned on different subsets of the data will in general be different from the policies learned on the full data, and it can be interesting to examine them to gain intuition for the stability of the learned rule. Table 6 exhibits the stability of learned rule, however, Figure 6 is a graphical depiction of a learned policy and can be different to different subsets of the data.

Appendix C.

6 Discussion and Conclusion

In 2008, 10,000 low-income Oregonian adults (19 to 64 years of age) were randomly chosen to qualify for Medicaid, which provides a unique opportunity to study the causal effect of Medicaid coverage. Finkelstein et al. (2012) found in the year following the random assignment of Medicaid, the treatment group had higher health care use, lower out-of-pocket medical expenditures and medical debt, and better self-reported physical and mental health than the control group, but it did not have detectable improvements in physical health conditions like high blood pressure. However, these mixed-bag effects of Medicaid puzzle researchers to determine what drives the relationship between Medicaid and other outcomes of interest. My paper puts forward an argument of heterogeneous treatment effect where Medicaid distinctly affects different individuals and subpopulations differently. Furthermore, I use these heterogeneous treatment effects to reveal policy reforms. These reforms prioritize Medicaid allotments to the subgroups that are likely to benefit the most. I also quantify how much these reforms improve from the baseline Medicaid impacts on health care use, personal finance, health, and well-being.

In this section, I present discussions on some of the obvious questions that the reader may have. This paper contemplates a situation where analysts know their outcome variable, (Y), at the post-treatment and have data of observables, (X), at the pre-treatment period. This situation may be a standard for many researchers. For this reason, this paper analyzes the data as an observational rather than a genuinely randomized study. Therefore, the unconfoundedness assumption to identify causal effects is crucial for this paper.

This paper focuses on "intent-to-treat" rather than "local average treatment effects." A local average treatment effect can be interpreted as the impact of Medicaid among compliers while an intent-to-treat estimates the net impact of expanding access to Medicaid. The results present both facts, but I mainly focus on the intent-to-treat because the problem policymakers face only a choice of the eligibility criteria and not the take-up. There can be many reasons for eligible people (lottery winner) not to accept Medicaid and people who do not win the lottery to get other insurance from other sources. This is the consumer's sovereignty, and policymakers cannot micromanage.

The heterogeneous effects of Medicaid are pronounced among households below 100% of the federal poverty line. A possible answer would be that more impoverished families may need more medical care. Medicaid provides an opportunity for these households to gain access to health care, and they, therefore, may utilize health care more than those who are uninsured that can be an exemplification of a standard adverse selection theory prediction. Also, I did not use the covariates like gender and race for the ethical and political rationale because these covariates cannot legally be used for treatment allocation. However, these are essential covariates, and not including these covariates can lead to higher standard errors in the estimates.

The proposed policy can be thought of as small reforms in Medicaid. Rather than a blanket policy that can be welfare-maximizing yet highly costly, these reforms target the subpopulation who are more likely to derive benefit and because these reforms are aimed, therefore, can be less expensive. For example, the federal government started to defund Oregons Medicaid Expansion from 2016 which has led to a budget deficit and Oregon Measure101 a two-year budget fix to close the state budget deficit by taxing hospital and insurance agencies, is nearing to end in 2020, these proposed reforms can help Oregon to reduce the state budget deficit.

To generalize the results outside the sample size, one needs to robustly account for the sampling variability of potentially unexplained household-level effects. This study takes a conservative approach and assumes that the outcome variables of an individual within the same household may be arbitrarily correlated within a household (or "cluster"), and therefore, utilizes the cluster-robust analysis. Each household is equally weighted, such that the model allows the prediction of the effect on a new individual from a new household, to generalize beyond the households given in the data. However, caution must be taken. First, these estimates are the one-year impact of expanding Medicaid access, and effects can change over longer time horizons than we can analyze. Second, these findings are the partial equilibrium effects of covering a small number of people, holding constant the rest of the health care system; the results of much more extensive health insurance expansions might differ because of supply-side responses by the health care sector. Third, the population is not representative of the low-income uninsured adults in the rest of the United States on several observable (and presumably unobservable) dimensions.

To conclude, I provide some evidence of heterogeneous treatment effects of Medicaid that can reconcile the mixed-bag results of Medicaid, as reported by previous literature. I also proposed some reforms that can improve program effectiveness. The Medicaid expansion, through the Affordable Care Act (ACA) and the contemporary fiscal pressure, has triggered a national debate amongst diverse stakeholders regarding the impacts of Medicaid coverage on various dimensions of public health, costs, and benefits. Some have argued that Medicaid decreases total health care spending by improving health and reducing inefficient hospital and emergency room utilization. Others have disputed that Medicaid reneges the promised benefits because Medicaid reimburses providers insufficiently, and therefore, recipients struggle to obtain access to care, and the low income uninsured already have reasonable access to care through clinics, uncompensated care, emergency departments, and out-of-pocket spending. Both of these arguments eventually motivate a need for substantial discussion and rigorous empirical assessment of what effects, if any, Medicaid coverage has on health care, health, and well-being and how to strike a balance between cost and benefits.

References

Allen, H., Baicker, K., Finkelstein, A., Taubman, S., and Wright, B. J. (2010). What the Oregon Health Study can Tell us about Expanding Medicaid. *Health Affairs*, 29(8):1498–1506.

Athey, S. (2018). The Impact of Machine Learning on Economics. In *The Economics of Artificial Intelligence: An Agenda*, number Ml, pages 1–27. University of Chicago Press.

Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360.

Athey, S. and Imbens, G. (2017a). Chapter 3 - the econometrics of randomized experiments. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 73 – 140. North-Holland.

Athey, S. and Imbens, G. W. (2017b). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.

Athey, S., Imbens, G. W., and Wager, S. (2018). Approximate residual balancing: debiased inference of average treatment effects in high dimensions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4):597–623.

Athey, S., Tibshirani, J., and Wager, S. (2016). Solving Heterogeneous Estimating Equations with Gradient Forests. Research Papers 3475, Stanford University, Graduate School of Business.

Athey, S., Tibshirani, J., Wager, S., et al. (2019). Generalized random forests. *The Annals of Statistics*, 47(2):1148–1178.

Athey, S. and Wager, S. (2019a). Efficient Policy Learning. pages 1–37.

Athey, S. and Wager, S. (2019b). Estimating treatment effects with causal forests: An application. *arXiv preprint arXiv:1902.07409*.

Baicker, K. (2019). The effect of health insurance on spending, health, and well-being − evidence and implications for reform.

Baicker, K., Allen, H. L., Wright, B. J., and Finkelstein, A. N. (2017). The Effect Of Medicaid On Medication Use Among Poor Adults: Evidence from Oregon. *Health Affairs*, 36(12):2110–2114.

Baicker, K. and Finkelstein, A. (2011). The Effects of Medicaid Coverage Learning from the Oregon Experiment. *New England Journal of Medicine*, 365(8):683–685.

Baicker, K., Finkelstein, A., Song, J., and Taubman, S. (2014). The Impact of Medicaid on Labor Market Activity and Program Participation: Evidence from the Oregon Health Insurance Experiment. *American Economic Review*, 104(5):322–328.

Baicker, K., Taubman, S. L., Allen, H. L., Bernstein, M., Gruber, J. H., Newhouse, J. P., Schneider, E. C., Wright, B. J., Zaslavsky, A. M., and Finkelstein, A. N. (2013). The Oregon Experiment Effects of Medicaid on Clinical Outcomes. *New England Journal of Medicine*, 368(18):1713–1722.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014a). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2):29–50.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014b). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2):608–650.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. (1984). Classification and regression trees Regression trees. *Wadsworth: Belmont, CA*, (June):358.

Brook, R. H., Ware, J. E., Rogers, W. H., Keeler, E. B., Davies, A. R., Donald, C. A., Goldberg, G. A., Lohr, K. N., Masthay, P. C., and Newhouse, J. P. (1983). Does Free Care Improve Adults' Health? *New England Journal of Medicine*.

Card, D. and Maestas, N. (2008). Care Utilization : Evidence from Medicare. *American Economic Review*, 98(5):2242–2258.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018a). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1):C1–C68.

Chernozhukov, V., Demirer, M., Duflo, E., and Fernandez-Val, I. (2018b). Generic machine learning inference on heterogenous treatment effects in randomized experiments. Technical report, National Bureau of Economic Research.

Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2008). Nonparametric Tests for Treatment Effect Heterogeneity. *The Review of Economics and Statistics*, 90(3):389–405.

Currie, J. and Gruber, J. (1996a). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *The Quarterly Journal of Economics*, 111(2):431–466.

Currie, J. and Gruber, J. (1996b). Saving Babies : The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy*, 104(6):1263–1296.

Davis, J. M. and Heller, S. B. (2017). Using Causal Forests to Predict Treatment Heterogeneity: An application to Summer Jobs. *American Economic Review*, 107(5):546–550.

Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. *CoRR*, abs/1103.4601.

Dudk, M., Erhan, D., Langford, J., and Li, L. (2014). Doubly robust policy evaluation and optimization. *Statist. Sci.*, 29(4):485–511.

Finkelstein, A. and McKnight, R. (2008). What Did Medicare Do? The Initial Impact of Medicare on Mortality and Out of Pocket Medical Spending. *Journal of Public Economics*.

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

Foden-Vencil, K. (2018). Oregon Measure 101: What You Need To Know . News — OPB.

Frank, L. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*.

Garfield, R., Rudowitz, R., Orgera, K., and Damico, A. (2019). Understanding the intersection of medicaid and work: What does the data say?

Glaeser, E. L., Hillis, A., Kominers, S. D., and Luca, M. (2016). Crowdsourcing City Government: Using Tournaments to Improve Inspection Accuracy. *American Economic Review*, 106(5):114–118.

Glaeser, E. L., Kominers, S. D., Luca, M., and Naik, N. (2018). Big Data and Big Cities: the Promises and Limitations of Improved Measures of Urban Life. *Economic Inquiry*, 56(1):114–137.

Grossman, R. L., Heath, A. P., Ferretti, V., Varmus, H. E., Lowy, D. R., Kibbe, W. A., and Staudt, L. M. (2016). Effect of Medicaid Coverage on ED Use Further Evidence from Oregon's Experiment. *New England Journal of Medicine*, 363(1):1–3.

Hanratty, M. J. (1996). American Economic Association Canadian National Health Insurance and Infant Health. *The American Economic Review*, 86(1):276–284.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *Source: The American Economic Review American Economic Review*, 102(1022):994–1028.

Imai, K. and Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jiang, N. and Li, L. (2015). Doubly robust off-policy evaluation for reinforcement learning. *CoRR*, abs/1511.03722.

Kaiser Family Foundation.

Kaiser Family Foundation (2019). Status of state action on the medicaid expansion decision.

Kallus, N. (2018). Balanced policy evaluation and learning. In *Advances in Neural Information Processing Systems*, pages 8895–8906.

Katch, H., Wagner, J., and Aron-Dine, A. (2018). Taking medicaid coverage away from people not meeting work requirements will reduce low-income families access to care and worsen health outcomes.

Kitagawa, T. and Tetenov, A. (2018). Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice. *Econometrica*.

Klein, E. (2013). Here's what the Oregon Medicaid study really said - The Washington Post. *The Washington Post*, (August 2012).

Kleinberg, J., Ludwig, J., Mullainathan, S., and Obermeyer, Z. (2015). Prediction Policy Problems. *American Economic Review*.

Knaus, M. C., Lechner, M., and Strittmatter, A. (2017). Heterogeneous Employment Effects of Job Search Programmes: A Machine Learning Approach. Economics Working Paper Series 1711, University of St. Gallen, School of Economics and Political Science.

Lan, W., Zhong, P.-S., Li, R., Wang, H., and Tsai, C.-L. (2016). Testing a single regression coefficient in high dimensional linear models. *Journal of Econometrics*, 195(1):154 – 168.

Lee, M. J. (2009). Non-parametric tests for distributional treatment effect for randomly censored responses. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71(1):243–264.

Levy, H. and Meltzer, D. (2004). What Do We Really Know About Whether Health Insurance Affects Health? *Health policy and the uninsured*, pages 179–204.

Levy, H. and Meltzer, D. (2008). The Impact of Health Insurance on Health. *Annual Review of Public Health*, 29(1):399–409.

Li, L., Chu, W., Langford, J., Moon, T., and Wang, X. (2012). An unbiased offline evaluation of contextual bandit algorithms with generalized linear models. In Glowacka, D., Dorard, L., and Shawe-Taylor, J., editors, *Proceedings of the Workshop on On-line Trading of Exploration and Exploitation 2*, volume 26 of *Proceedings of Machine Learning Research*, pages 19–36, Bellevue, Washington, USA. PMLR.

List, J. A., Shaikh, A. M., and Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*.

Manski, C. F. (2009). *Identification for prediction and decision*. Harvard University Press.

McWilliams, J. M., Meara, E., Zaslavsky, A. M., and Ayanian, J. Z. (2007a). Health of Previously Uninsured Adults after Acquiring Medicare Coverage. *JAMA - Journal of the American Medical Association*.

McWilliams, J. M., Meara, E., Zaslavsky, A. M., and Ayanian, J. Z. (2007b). Use of Health Services by Previously Uninsured Medicare Beneficiaries. *New England Journal of Medicine*.

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

Naik, N., Raskar, R., and Hidalgo, C. A. (2016). Cities Are Physical Too: Using Computer Vision to Measure the Quality and Impact of Urban Appearance. *American Economic Review*, 106(5):128–132.

Newhouse, J. P. (1994). Free for All: Lessons from the RAND Health Insurance Experiment. *BMJ*.

Nie, X. and Wager, S. (2017). Quasi-oracle estimation of heterogeneous treatment effects. *arXiv preprint arXiv:1712.04912*.

Norris, L. (2018). Oregon and the ACA's Medicaid expansion: eligibility, enrollment and benefits — healthinsurance.org.

Office for Oregon Health Policy and Research (2009). Trends in oregons healthcare market and the oregon health plan: A report to the 75th legislative assembly.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 56(4):931–954.

Rosenbaum, P. R. and Rubin, D. B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1):41–55.

Rueben, B. (2019). Let data guide medicaid reforms — opinion.

Strehl, A. L., Langford, J., and Kakade, S. M. (2010). Learning from logged implicit exploration data. *CoRR*, abs/1003.0120.

Swaminathan, A. and Joachims, T. (2015). Batch learning from logged bandit feedback through counterfactual risk minimization. *Journal of Machine Learning Research*, 16(52):1731–1755.

Taubman, S. L., Allen, H. L., Wright, B. J., and Baicker, K. (2014). Oregon's Health Insurance Experiment. *Science*, 343(6168):263–268.

Thomas, P. S. and Brunskill, E. (2016). Data-efficient off-policy policy evaluation for reinforcement learning. *CoRR*, abs/1604.00923.

Tibshirani, R. (1996). Regression Selection and Shrinkage via the Lasso.

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2):3–28.

Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242.

Wallace, N. T., McConnell, K. J., Gallia, C. A., and Smith, J. A. (2008). How effective are copayments in reducing expenditures for low-income adult medicaid beneficiaries? Experience from the Oregon Health Plan. *Health Services Research*, 43.

Willke, R. J., Zheng, Z., Subedi, P., Althin, R., and Mullins, C. D. (2012). From concepts, theory, and evidence of heterogeneity of treatment effects to methodological approaches: A primer.

Xie, Y., Brand, J. E., and Jann, B. (2012). Estimating Heterogeneous Treatment Effects with Observational Data. *Sociological methodology*, 42(1):314–347.

Zhou, R. A., Baicker, K., Taubman, S., and Finkelstein, A. N. (2017). The uninsured do not use the emergency department more- they use other care less. *Health Affairs*, 36(12):2115–2122.

A Causal machine learning approaches

A.1 Average treatment effect

In this paragraph, I show a few examples of a causal machine learning approach to estimate the average treatment effect. For example, Belloni et al. (2014b) and Belloni et al. (2014a) utilize "off-the-shelf" or readily available predictive machine learning algorithm called the "LASSO"[26] method and purpose a correction[27] called the "double-selection post-LASSO"[28] method. This method is useful for estimating the average treatment effect when the analyst is required to select a "sparse" outcome model[29] from high-dimensional observables when some covariates correlate with treatment and outcome, and the analyst does not know which ones are important. Similarly, Athey et al. (2018) utilize "doubly-robust"[30] method

[26]The Least Absolute Shrinkage and Selection Operator (LASSO) is an appealing method to estimate the sparse parameter from a high-dimensional linear model is introduce by Frank and Friedman (1993) and Tibshirani (1996). The LASSO simultaneously performs model selection and coefficient estimation by minimizing the sum of squared residuals plus a penalty term. The penalty term penalizes the size of the model through the sum of absolute values of coefficients. Consider a following linear model $\tilde{y}_i = \Theta_i\beta_1 + \varepsilon_i$, where Θ is high-dimensional covariates, the LASSO estimator is defined as the solution to $\min_{\beta_1 \in \mathbb{R}^p} E_n\left[(\tilde{y}_i - {}_i\Theta\beta_1)^2\right] + \frac{\lambda}{n}\|\beta_1\|_1$, the penalty level λ is a tuning parameter to regularize/controls the degree of penalization and to guard against over-fitting. The cross-validation technique chooses the best λ in prediction models and $\|\beta\|_1 = \sum_{j=1}^{p}|\beta_j|$. The kinked nature of penalty function induces $\hat{\beta}$ to have many zeros; thus LASSO solution feasible for model selection.

[27]When LASSO of outcome variable is implemented to select the covariates while always restricting the treatment indicator, the estimated treatment effect is biased because LASSO's sole objective is to select variables that predict outcome thus LASSO fails to select confounders that are also strong predictor of treatment assignment.

[28]Belloni et al. (2014a) simplify the double-selection post-LASSO procedure as following. First, run LASSO of outcome variables on a large list of potential covariates to select a set of predictors for the outcome variable. Second, run LASSO of treatment variable on a large list of potential covariates to select a set of predictors for treatment. If the treatment is truly exogenous, we should expect this second step should not select any variables. Third, run OLS regression of outcome variable on treatment variable, and the union of the sets of regressors selected in the two LASSO runs to estimate the effect of treatment on the outcome variable then correct the inference with usual heteroscedasticity robust OLS standard error.

[29]The "sparse" outcome model means a model with a few meaningful covariates that affect the average outcome. These few meaningful covariates are selected from a given list of many observables covariates, and potentially a situate when numbers of observables k are greater than numbers of observations n, i.e., $k > n$. When $k > n$, an estimation based on the least-squares estimation is infeasible. However, traditionally, the principal component analysis (PCA) is commonly used to reduce dimension when the likelihood function is normal. The PCA creates principal components using linear combinations of a much larger set of variables from a multivariate data-set. Interpreting the coefficients on the principal components requires the researcher first to interpret the principal components, which can prove a challenge as all variables have non-zero loadings.

[30]The "doubly-robust" estimator proceeds by taking the average of the efficient score, which involves the estimation of conditional mean of outcomes given covariates as well as the inverse propensity score Athey (2018).

and LASSO method and purpose "residual balancing"[31] approach for estimating average treatment effect under the assumption of unconfoundedness[32] and the assumption of the outcome model is linear and sparse. Similarly, Chernozhukov et al. (2018a) purpose "double machine learning" for estimating the average treatment effect under unconfoundedness. The idea is to first run any feasible machine learning methods of outcomes on covariates, and then second run another feasible machine learning methods of the treatment indicator on covariates; then, the residuals from the first machine learning are regressed on the residuals from the second machine learning to estimate the average treatment effect. This idea is similar to Frish-Waugh-Lovell theorem[33] and close to the concept of Robinson (1988) residual-on-residual regression approaches where the estimator was a kernel regression.

A.1.1 Heterogeneous treatment effects

Along with the average treatment effect, heterogeneous treatment effects estimation interests policymakers because it helps to quantify the sizes of effects on different subpopulations, which is valuable to improve program targeting and to understand the underlying mechanisms driving the results. Usually, data are stratified in mutually exclusive groups or include interactions in a regression to explore heterogeneous treatment effects. However, ad-hoc searches for the responsive subgroups may lead to false discoveries or may mistake noise for a true treatment effect (Davis and Heller, 2017). Knaus et al. (2017) points out that for large-scale investigations of effect heterogeneity, standard $p-$values of standard (single) hypothesis tests are no longer valid because of the multiple hypothesis testing problems (Lan et al., 2016; List et al., 2019) and leads to so-called "ex-post selection" problem which is widely recognized in the program evaluation literature. For example, for fifty single hypotheses tests, the probability that at least one test falsely rejects the null hypotheses at the 5% significance level (assuming independent test statistics as an extreme case) is $1 - 0.95^{50} = 0.92$ or 92%.

The new avenue of causal machine learning provides a better systematic approach to search the groups with heterogeneous treatment effects. One intuitive approach proposed by Imai and Ratkovic (2013) is to sample-split and use the first sample to run the LASSO regression model with the treatment indicator interacted with covariates and perform variables selections then use the selected model with the second sample to perform an ordinary least squares regression to guard against over-fitting. While Athey and

[31] The "residual balancing" replaces inverse propensity score weights with weights obtained using quadratic programming, where the weights are designed to achieve balance between the treatment and control group. The conditional mean of outcomes is estimated using LASSO Athey (2018).

[32] The unconfoundedness assumption implies treatment is randomly assigned and knowing observable characteristics of an individual, and their treatment status gives no additional information on the potential outcomes. This means the treatment assignment is independent of the outcome variable.

[33] The Frisch-Waugh-Lovell theorem is that estimating a parameter in a multiple regression is equivalent to estimating the same parameter in a simple regression of the residual of the regress and regressed on all other predictors on the residual of the regressor regressed on all other predictors.

Imbens (2016) utilizes the Breiman et al. (1984) classification and regression tree (CART)[34] machine learning algorithms and purpose "causal tree" method. The CART recursively filters and partitions the large data-set into binary sub-groups (nodes) such that the samples within each subset become more homogeneous that fit the response variable. Unlike the CART that minimizes the mean-squared error of the prediction of outcomes to capture heterogeneity in outcomes, the "causal" tree minimizes the mean-squared error of treatment effects to capture treatment effect heterogeneity. The approach to estimate the "causal" tree is similar to Imai and Ratkovic (2013) approach, in which half of the sample is used to determine the optimal partition of covariates space, while the other half is used to estimate treatment effects within the leave based on the optimal partition of covariates selected from the first partition (Athey and Imbens, 2016). The sample-splitting approach also known as "honest" estimation lead to loss of precision as only half of the data is used to estimate the effect, but generates a treatment effect and a confidence interval for each subgroup that is valid no matter how many covariates are used in estimation. Athey and Imbens (2017b) points out that the researcher is free to estimate a more complex model in the second part of the data, for example, if the researcher wishes to include fixed effects in the model, or model different types of correlation in the error structure.

The causal tree doesn't provide personalized estimates, Wager and Athey (2018) utilize the "random forest" machine learning approach and propose a "causal forest" method, where many different causal trees are generated and averaged. This method provides causal effects that change more smoothly with covariates and provides distinct individualized estimates and confidence intervals. Wager and Athey (2018) also shows that the predictions from causal forests are asymptotically normal and centered on the true conditional average treatment effect for each individual. Athey et al. (2016) extend the approach to other models for causal effects, such as instrumental variables, or other models that can be estimated using the generalized method of moments (GMM). In each case, the goal is to evaluate how a causal parameter of interest varies with covariates.

A.1.2 Optimal policy estimation

The optimal policy estimation have received greater attention in the machine learning literature[35] (Athey, 2018). The optimal policy function map the observable characteristics of an individual to a policy or treatment assignment. In simplest, the main goal of optimal policy estimation is to answer—- "who should be treated?" or optimal treatment allocation. The understanding of optimal policy is essential

[34] In simplest, the CART algorithm chooses a variable and split that variable above or below a certain level (which forms two mutually exclusive subgroups or leaves) such that the sum of squared residuals is minimized. This splitting process is repeated for each leave until the reduction in the sum of squared residuals is below a certain level as defined by users, thus resulting in a tree format (Athey and Imbens, 2017b).

[35] See Strehl et al. (2010); Dudík et al. (2011); Li et al. (2012); Dudk et al. (2014); Swaminathan and Joachims (2015); Jiang and Li (2015); Thomas and Brunskill (2016) and Kallus (2018).

in policymaking because an ad-hoc targeting a specific subpopulation with positive interventions can be unfair, unethical, illegal, and unpolitical to some other subpopulations while intervening everyone in the population (a blanket policy) is welfare-maximizing but can be extremely costly.

The optimal policy estimation or optimal treatment allocation has been recently studied in using causal machine learning in economics, mainly by Kitagawa and Tetenov (2018) and Athey and Wager (2019a). The main idea is to select a policy function that minimizes the loss from failing to use the ideal policy, referred to as the regret of the policy. Note that estimating conditional average treatment effect or heterogeneous treatment effect focus on the squared-error loss while the optimal policy estimation focuses on utilitarian regret Athey and Wager (2019a).

B Variable importance

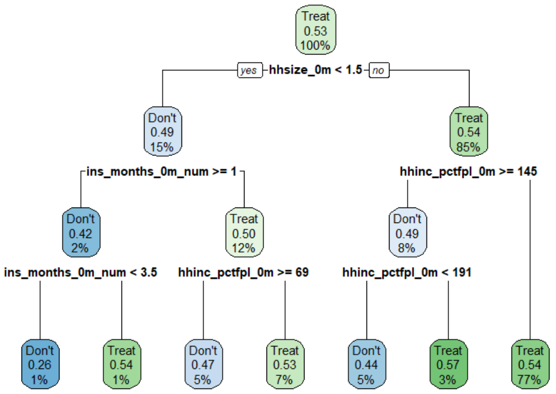
Table 7: Variable importance

Variables	FPL	Age	HHS	INS	Other variables
Currently taking any prescription medications	✓	✓	✓		% MSA
Outpatient visits last six months	✓	✓		✓	
ER visits last six months	✓	✓	✓		
Inpatient hospital admissions last six months	✓	✓	✓	✓	
Number of prescription medications currently taking	✓	✓	✓	✓	
Number of Outpatient visits last six months	✓	✓		✓	
Number of ER visits last six months	✓	✓	✓	✓	% High school disploma or GED
Number Inpatient hospital admissions last six months	✓	✓	✓		% Self signup
Any out of pocket medical expenses, last six months	✓	✓	✓	✓	% MSA
Owe money for medical expenses currently	✓	✓	✓		
Borrowed money or skipped other bills to pay medical bills, last six months	✓	✓	✓		
Refused treatment because of medical debt, last six months	✓	✓	✓		
Out of pocket costs for doctors visits, clinics or health centers, past 6 months	✓	✓			% work 30+ hrs/week
Out of pocket costs for emergency room or overnight hospital care, past 6 months	✓	✓	✓		
Out of pocket costs for prescription medicine, past 6 months	✓	✓	✓	✓	
Out of pocket costs for other medical care, past 6 months	✓	✓		✓	
Total out of pocket costs for medical care, last 6 months	✓	✓	✓	✓	% work 30+ hrs/week
Total amount currently owed for medical expenses	✓	✓	✓	✓	
Have usual place of clinic-based care	✓	✓			
Have personal doctor	✓	✓		✓	% work 30+ hrs/week
Got all needed medical care, last six months	✓	✓	✓	✓	% work 30+ hrs/week
Got all needed drugs, last six months	✓	✓	✓		% dont currently work
Didn't use ER for non emergency, last six months	✓	✓	✓		% work 30+ hrs/week
Quality of care received last six months good/very good/excellent (conditional on any)	✓	✓	✓		% MSA
Happiness, very happy or pretty happy (vs. not too happy)	✓	✓	✓	✓	
Blood cholesterol checked (ever)	✓	✓	✓	✓	
Blood tested for high blood sugar/diabetes (ever)	✓	✓	✓		
Mammogram within last 12 months (women 40)	✓	✓	✓	✓	% work 30+ hrs/week
Pap test within last 12 months (women)	✓	✓	✓		% work 30+ hrs/week
Self-reported health good/very good/excellent (not fair or poor)	✓	✓	✓	✓	
Self-reported health not poor (fair, good, very good, or excellent)	✓	✓		✓	
Health about the same or gotten better over last six months	✓	✓	✓		% High school diploma or GED
Number of days physical health good, past 30 days	✓	✓	✓		
Number days poor physical or mental health did not impair usual activity, past 30 days	✓	✓	✓	✓	
Number of days mental health good, past 30 days	✓	✓	✓	✓	% Female
Did not screen positive for depression, last two weeks	✓	✓	✓		

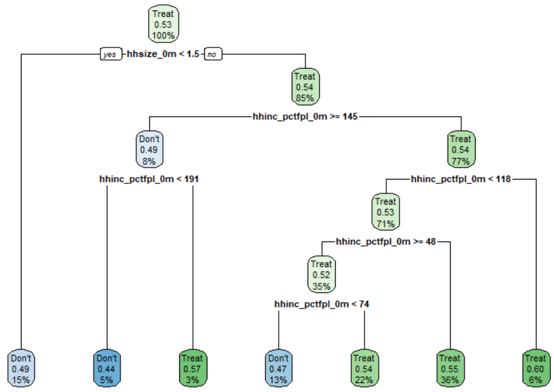
Notes: FPL represents household below the federal poverty line (in %), HHS represents household size, INS represents the nummber of non insurance months in last six months. The random forest model always splits on FPL and Age along with HHS and INS. Along with these variables the random forest also splits on different variables included in the last column. For example, consider the model called "Currently taking any prescription medications", the random forest splits (more than average) the data on FPL, Age, HHS, and % MSA. Therefore, the treatment heterogeneity is likely within these variables.

C Efficient Policies

Figure 7: Efficient policy to improve the blood cholesterol check participation.



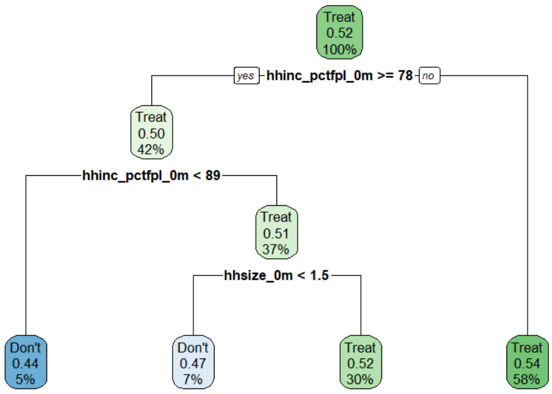
(a) Shallow tree



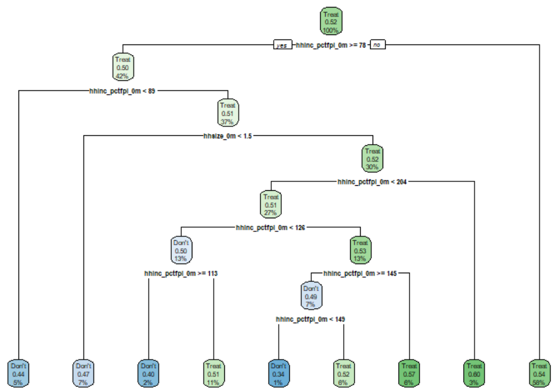
(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size.

Figure 8: Efficient policy to improve blood tests participation for high blood sugar/diabetes.



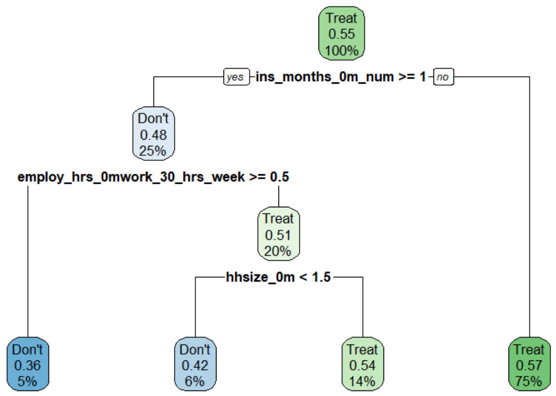
(a) Shallow tree



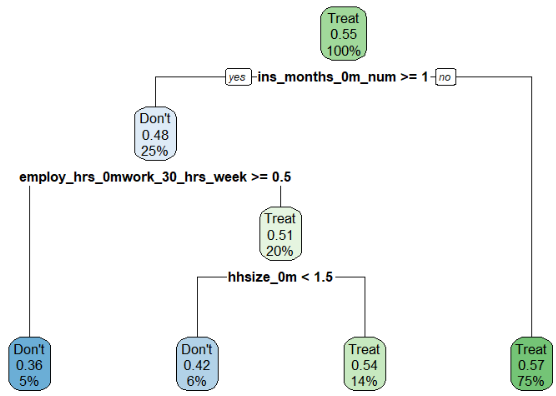
(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size.

Figure 9: Efficient policy to improve Mammogram test participation for women.



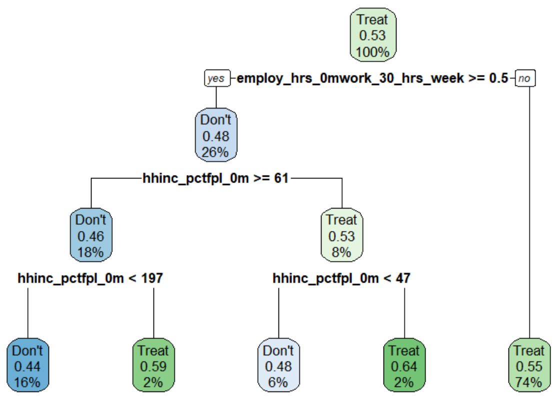
(a) Shallow tree



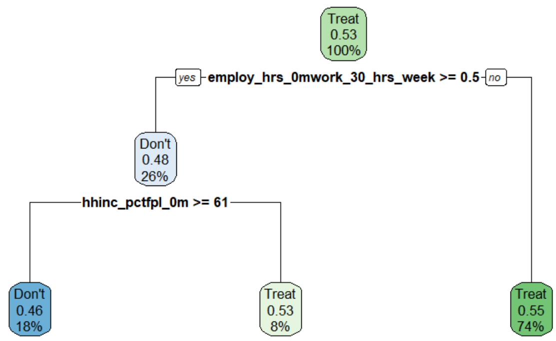
(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size. Valid only for women.

Figure 10: Efficient policy to improve Pap test participation for women.



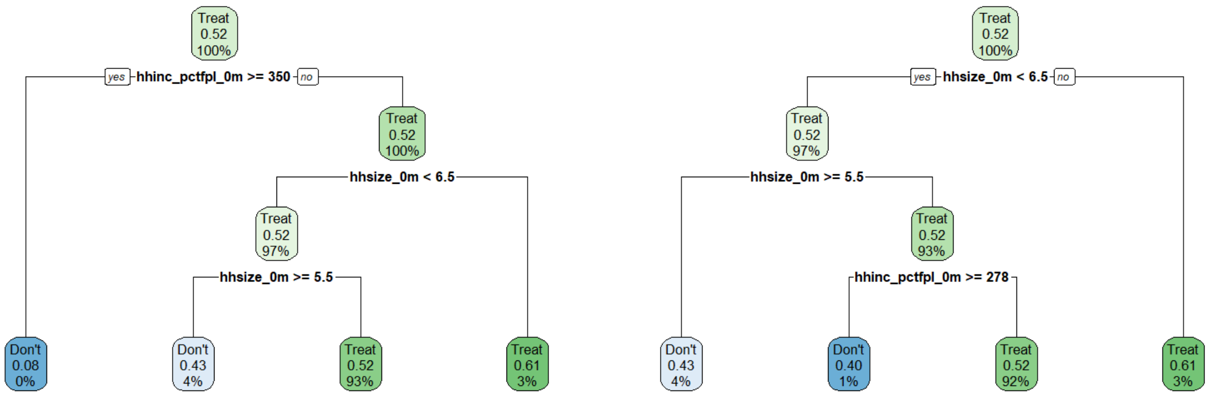
(a) Shallow tree



(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size. Valid only for women.

Figure 11: Efficient policy to improve Self-reported health.

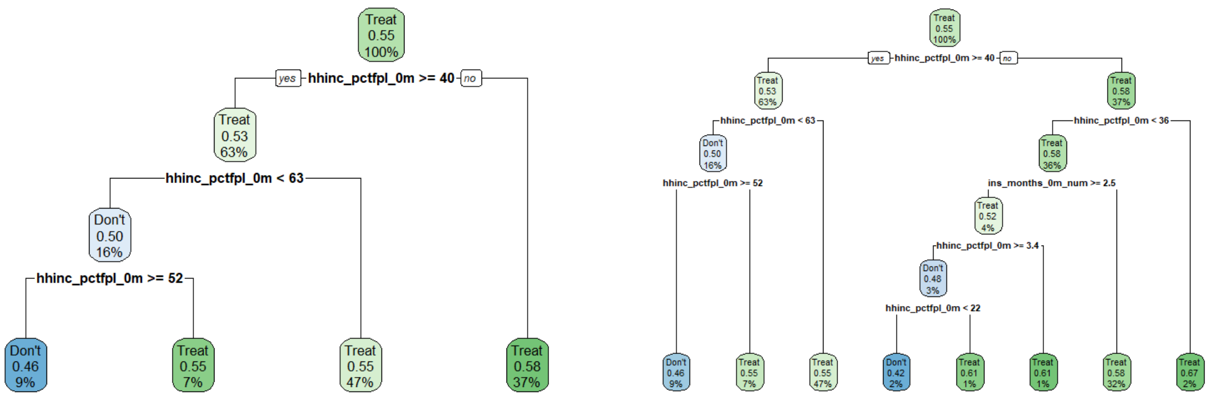


(a) Shallow tree

(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size. Valid only for women.

Figure 12: Efficient policy to improve to have usual place of clinic-based care.

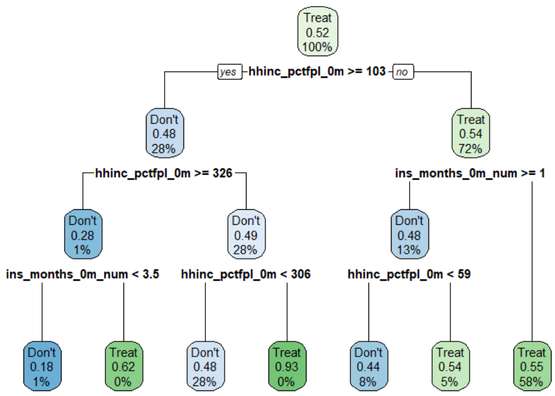


(a) Shallow tree

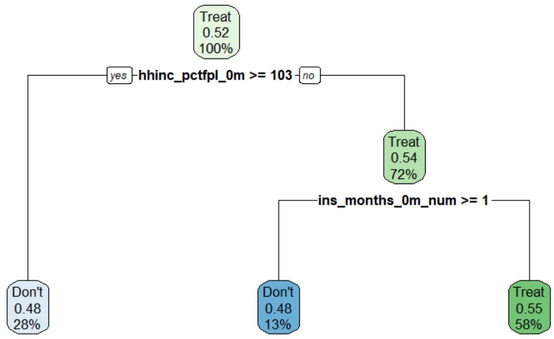
(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size. Valid only for women.

Figure 13: Efficient policy to improve to have have a personal doctor.



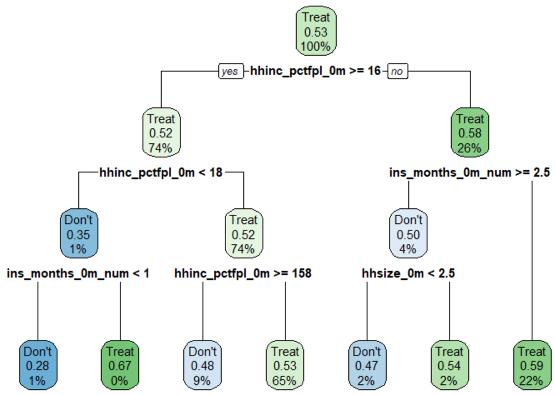
(a) Shallow tree



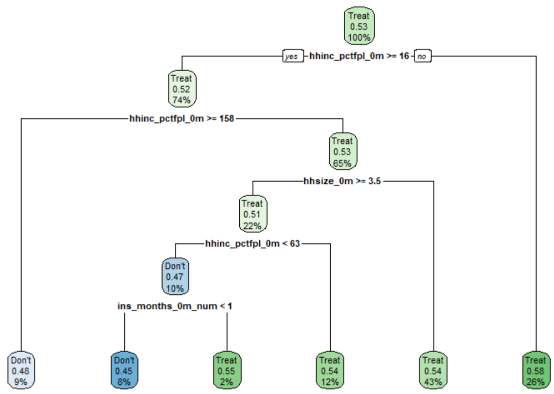
(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size. Valid only for women.

Figure 14: Efficient policy to improve to post health-care service happiness.



(a) Shallow tree



(b) Deep tree

Notes: The hhinc_pctfpl_0m shows household income as percentage of the federal poverty line. The ins_months_0m_num shows numbers of months that a responder has insurance in last six months. The employ_hrsmwork_30_hrs_week shows > 0.5 shows the responder work more than 30 hours per/week. The hhsize_0m is household size. Valid only for women.