Effects of Health Insurance Coverage on Health & Financial Outcomes*

Timothee Maret, Yasmine Ouattara, Bhavya Pandey, Donna Zheng

March 8, 2024

1 Abstract

The Oregon Health Insurance Experiment (OHIE), initiated in 2008, provides a rare opportunity to study the impact of health insurance on health and financial outcomes through a randomized controlled design. Using data from follow-up surveys after the rollout of the program, we sought to find the impact of health insurance coverage on health and financial outcomes. This project uses a two-pronged strategy to first use Causal Forests to identify relevant heterogeneous treatment affects, and then deploy a Double Machine Learning strategy for taking into account missing data from non-responses in the survey to compute the average treatment effects. We find strong positive effects of health insurance coverage on financial outcomes while some estimates pertaining to health outcomes are diminished than those noted in previous research, through our DML analysis. Further, the causal forest method reveals interesting heterogenous treatment effects on health outcomes, particularly by subgroups such as age (birth year), race, and gender.

*This paper is a final project for ECMA 31350 Machine Learning for Economists, taught by Professor Kirill Ponomarev in Winter 2024, at the University of Chicago. The authors would like to thank Professor Kirill Ponomarev and TA Chanwool Kim for their guidance. Code for replication: https://github.com/Yasmineouattara/ML_for_Econ_Final_Project/tree/main

2 Introduction & Motivation

Expanding health insurance coverage presents a complex scenario with varied predictions from economic theory regarding its impact on healthcare consumption, financial health, and long-term well-being. The Oregon Health Insurance Experiment (OHIE), initiated in 2008, provides a rare opportunity to study the impact of health insurance on health and financial outcomes through a randomized controlled design. This was one of the earliest works that provided experimental evidence to study the effects of health insurance coverage. In our project, we aim to analyze the effects of Medicaid coverage by building on previous research methodologies around causal inference, to estimate the relevant Average Treatment Effects (ATEs) and the Heterogeneous ATEs in this setting, using machine learning methods and addressing data gaps through imputation techniques.

Although the quasi-experimental approaches used in many studies can be an improvement over observational designs and often involve larger samples than are feasible with a randomized design, in such settings, we cannot eliminate confounding factors as effectively as random assignment. The Oregon lottery thus presented an unprecedented opportunity to isolate the causal effect of insurance on the health care use, health outcomes, financial strain, and well-being of low-income adults.

In the first one to two years of coverage, Medicaid improved self-reported health and reduced depression, but had no statistically significant effect on several measures of physical health (Finkelstein et al., 2012). Using commercial credit reports and surveys, it was also found that Medicaid reduced the financial hardship of medical care for recipients (Finkelstein et al., 2016).

However, some of these studies acknowledged their limitations – this quote from Finkelstein et al., 2012, for instance:
"*The survey data [...] with a 50% effective response rate, are subject to potential non-response bias [...] These outcomes are only available for individuals who responded to the mail survey and may therefore not be representative of the full sample.*"
Such limitations within the data are concerning because we do not have non-response at random.

Further, existing research that study the impact of health insurance coverage on financial or health outcomes are also heterogeneous to the extent that they study two different effects – that of the lottery versus that of the enrollment in the health insurance itself, are generally computed in limited settings thus introducing a potential confounding factor, or measure narrow impacts on specific subgroups. Further, modern machine learning methods in the context of health insurance literature are largely limited to prediction-type analyses. Given the scope this study provides for the deployment of machine learning methods for causal investigation, to address this gap in the literature, and to leverage the existence of multiple potential covariates to turn to methods beyond the classic non-parametric methods, we advance the work of the OHIE in this paper.

3 Literature Review

In the context of studying the impact of the Medicaid lottery, various research have contributed significantly to understanding its implications. Finkelstein et al., 2016 focused on the effect of Medicaid on emergency department use over approximately 18 months and found a 40% increase in emergency department (ED) visits among Medicaid recipients, with increases across a range of visit types, challenging assumptions that Medicaid would reduce unnecessary ED use by improving access to primary care. More recently, this has also been studied by Denteh and Liebert, 2023. Baicker, Allen, Wright, et al., 2017c examined Medicaid's effect on managing depression and found that Medicaid coverage nearly halved the prevalence of undiagnosed depression and reduced untreated depression by more than 60%, highlighting Medicaid's potential to significantly impact mental health through increased medication use and reduced unmet mental health care needs. In another paper, they also concluded that coverage significantly reduced the share of respondents reporting unmet dental care needs and increased visits to emergency departments for dental issues. However, it had no detectable effect on uncovered dental care or out-of-pocket spending, illustrating the limitations of Medicaid coverage in addressing all aspects of dental health (Baicker, Allen, Wright, et al., 2017a). With regards to medication use, Baicker, Allen, Wright, et al., 2017b found Medicaid's effect on increasing prescription medication usage, and provided evidence of enhanced access to necessary medications for managing mental health and diabetes among the newly insured, indicating Medicaid's crucial role in improving medication adherence.

In Baicker et al., 2014 exploring Medicaid's impact on employment and program participation, the authors found no significant effect of Medicaid on employment, earnings, or most government benefit receipts, aside from an increase in food stamp uptake. This suggested Medicaid's role in providing health insurance does not disincentivize work among low-income adults. In adjacent research Baicker and Finkelstein also concluded that respondents to follow-up surveys reported that Medicaid virtually eliminated out-of-pocket catastrophic medical expenditures and reduced the probability of having to borrow money or skip paying other bills because of medical expenses by more than 50%.

When it comes to the methodological literature we reviewed in regards with the use of machine learning methods towards answering questions pertaining to the causal impact of health insurance, two recent works came to light[1] – Hattab et al., 2024 and Goto et al., 2024. While the former used the in-person survey data to estimate CATE using instrumental forest and found weak heterogeneity in effect across subgroups based on gender, age, and race, the latter used the causal forest approach to detect heterogeneous effects of Medicaid coverage on depression and found reduced

[1] Other related research particularly combining methods such as instrumental variables with machine learning has been conducted by Johnson et al., 2022, Chernozhukov et al., 2018, Frölich, 2007

risk of depression and greater impacts for populations at-risk at baseline.

4 Data

In 2008, Oregon implemented a limited expansion of its Medicaid program for low-income adults through a lottery, selecting names from a waiting list to fill a limited number of available spots. Those selected had the opportunity to apply for Medicaid and to enroll if they met eligibility requirements. One instrumental variable of this selection process is the Randomized Assignment. The lottery system's providing the opportunity to apply for Medicaid serves as a natural experiment. This randomization ensures that the treatment and control groups are comparable on both observed and unobserved characteristics, addressing potential selection bias that often complicates observational studies. The randomness of the lottery thus allows for a clean comparison between those offered Medicaid and those not, isolating the impact of Medicaid access from other factors.

This lottery presented a unique opportunity to gauge the effects of Medicaid.

- In January 2008, Oregon opened a waiting list for its Medicaid program for low-income adults that had previously been closed to new enrollment. Approximately 90,000 people signed up. The state deemed random selection by lottery the fairest way within federal guidelines to allocate its limited number of openings.
- Between March and September 2008, the state drew approximately 30,000 names from this waiting list by lottery to fill the 10,000 available spots.
- This random selection allows us to gauge the many effects of health insurance itself, isolating it from the types of confounding factors that can plague observational studies.
- We have data on all the individuals that were deemed eligible for this program, those that were selected in the lottery, and those that eventually ended up enrolling for coverage after they were selected in the lottery and deemed eligible to enroll.

The OHIE data is publicly available.[2] It contains eight data files with very few overlapping variables and provides information on all 74,922 individuals that were determined to be eligible for the study. The four files that especially concern us are the ones pertaining to demographic characteristics, enrollment status in state programs, responses from the initial mail survey and responses from the twelve-month survey. Since the follow up surveys with our outcome variables and informing the treatment assignment are self-reported in nature, our analysis seeks to take this into account.

[2]Public data link: https://www.nber.org/research/data/oregon-health-insurance-experiment-data

4.1 Empirical Strategy

We propose a two-pronged investigation strategy in this paper – to first investigate the conditional local ATE using a causal forest approach[3] and then calculate the local ATE in general using Double Machine Learning (DML) while formally taking into account missing values from follow-up surveys[4].

5 Methodology

5.1 Causal Forest

One of our approaches for analyzing the Heterogeneous Treatment Effect in OHIE was the use of a Random Forest Double Machine Learning, or Causal Forest, due to its inherent ability to model complex, non-linear relationships and interactions between covariates, enabling the precise estimation of individual-level treatment effects across a diverse population.

5.1.1 Data Processing

We selected variables from the descriptive, ED, in-person, initial, and post-treatment survey datasets. Given the large amount of data available, we decided to use XGBoost to pre-select a handful of covariates and confounders in order to reduce dimensionality for further processing. We used XGBRegressor with the objective of minimized reg:squarederror. We ran the regressor on each outcome vector considering all data obtained before or at the onset of the lottery determined Medicaid coverage in September 2008. We selected 'gain', an average of a feature's contribution to the improvement in the model's predictive ability, as feature selection criteria due to its similarity to entropy used in Random Forests. Excluding the outcome and identification variables we reduced the number of covariates to 54 by considering only those in the top 75th percentile of their calculated gain score.

We then conducted a K-Nearest Neighbour (K-NN) imputation to fill in the missing values on both all features of the subsetted data. We selected the number of neighbors as 5 through cross validating the imputed data against Finkelstein et al., 2012 estimates. This imputation had very insignificant effects on the descriptive statistics of the subsetted data and maintained the original balance of covariates.

The Random Forest Double Machine Learning (Causal Forest) estimator was developed to target heterogeneous treatment effects. The method requires unconfoundedness that all covariates that

[3]This will allow us to compare our results with Hattab et al., 2024, who used same technique but carried out the analysis on different subgroups

[4]This allows us to compare results with Finkelstein et al., 2012, who used a Two Stage Least Squares (2SLS) approach.

could simultaneously have affected the treatment and the outcome to be observed. The data preprocessing approach we took attempted to satisfy this condition. Now, the method aims to solve the following local residual on residual moment condition:

$$\hat{\theta}(x) = \operatorname{argmin}_\theta \sum_{i=1}^{n} K_x\left(X_i\right) \cdot \left(Y_i - \hat{q}\left(X_i\right) - \theta \cdot \left(T_i - \hat{f}\left(X_i\right)\right)\right)^2 \tag{1}$$

For the first stage we use a Random Forest Classifier for treatment and a Random Forest Regressor for the outcome. The model then regresses a second stage on those residuals.

The two nuisance estimates are fitted on the global objective and the similarity metric that was potentially used to fit these estimates is not coupled with the similarity metric used in the final effect estimation.

The purpose of this analysis is to identify the heterogeneous treatment effects of health insurance on a variety of outcomes. The Random Forest Double Machine Learning Estimator provides multiple concrete reasons for its use on the Oregon Health Insurance Experiment data.

5.2 Double Machine Learning Accounting for Missing Values

To apply DML to LATE estimation while accounting for missing values, we based our approach on the procedure outlined in Chernozhukov et al., 2018. First, we need to construct an appropriate Neyman orthogonal score function. We preserve the usual set-up and assumptions associated with LATE estimation, as adapted from Lecture Notes 14-15 and Frölich (2007).

Assumption 1. Let $Z \in \{0, 1\}$ be a binary instrument, $D(0), D(1) \in \{0, 1\}$ denote potential treatments, and $Y(0), Y(1)$ denote potential outcomes. Let $\tau \in \{n, c, d, a\}$ denote the subpopulation belong, taking on four possible values: never takers n, compilers c, defiers d, and always takers a. Furthermore, let S be a binary variable indicating observation status, i.e. whether Y and D can be observed for a specific unit i. In addition, let W and X be two sets of observable covariates.

Suppose that the following assumptions hold.

1. Monotonicity: the subpopulation of defiers has probability measure 0, i.e.

$$\mathbb{P}(\tau = d) = 0.$$

2. Existence of compliers: the subpopulation of compliers has positive probability:

$$\mathbb{P}(\tau = c) > 0.$$

3. Validity of instrument conditional on W:

$$Y(0), Y(1), D(0), D(1) \perp Z \mid W.$$

4. Common support:

$$\text{supp}(W|Z = 0) = \text{supp}(W|Z = 1).$$

5. Missing at random conditional on X:

$$(Y, D, Z, W) \perp S \mid X.$$

Assumptions (1) through (4) allows identification of LATE as

$$\frac{\mathbb{E}[m(1, W) - m(0, W)]}{\mathbb{E}[p(W) - p(0, W)]},$$

where

$$m(Z, W) = \mathbb{E}[Y \mid Z, W], \quad p(Z, W) = \mathbb{E}[D \mid Z, W].$$

Along with Assumption (5), this gives us the following result.

Theorem 5.1. *Let $A = (Y, D, Z, W, S, X)$. Under the setting stated in Assumption 1, one can show that the score*

$$g(A; \tau_{LATE}, \eta) = (g^{1,1;o}(A; \eta) - g^{1,0;o}(A; \eta)) - (g^{2,1;o}(A; \eta) - g^{1,1;o}(A; \eta))\tau_{LATE}$$

is Neyman orthogonal with respect to η, where

$$\begin{aligned}
g^{1,1;o}(A; r, q, a_1, b_1) &= \frac{YZS}{rq} - (Z - r) \cdot \frac{a_1}{r^2} - (S - q) \cdot \frac{b_1}{rq}, \\
g^{1,0;o}(A; r, q, a_0, b_0) &= \frac{Y(1 - Z)S}{(1 - r)q} - (Z - r) \cdot \frac{a_0}{(1 - r)^2} - (S - q) \cdot \frac{b_0}{(1 - r)q}, \\
g^{2,1;o}(A; r, q, f_1, g_1) &= \frac{Y(1 - Z)S}{rq} - (Z - r) \cdot \frac{f_1}{r^2} - (S - q) \cdot \frac{g_1}{rq}, \\
g^{2,0;o}(A; r, q, f_0, g_0) &= \frac{Y(1 - Z)S}{(1 - r)q} - (Z - r) \cdot \frac{f_0}{(1 - r)^2} - (S - q) \cdot \frac{b_0}{(1 - r)q},
\end{aligned}$$

and $\eta = (r, q, a_1, a_0, b_1, b_0, f_1, f_0, g_1, g_0)$ *is such that*

$$
\begin{aligned}
r(W) &= \mathbb{E}[Z \mid W] = \mathbb{P}[Z = 1 \mid W], & & \\
q(X) &= \mathbb{E}[S \mid X] = \mathbb{P}[S = 1 \mid X], & q(X,W) &= \mathbb{E}[S \mid X, W] = \mathbb{P}[S = 1 \mid X, W], \\
a_1(W) &= \mathbb{E}[YZ \mid S = 1, W], & a_0(W) &= \mathbb{E}[Y(1-Z) \mid S = 1, W], \\
b_1(X,W) &= \mathbb{E}[YZ \mid S = 1, X, W], & b_0(X,W) &= \mathbb{E}[Y(1-Z) \mid S = 1, X, W], \\
f_1(W) &= \mathbb{E}[DZ \mid S = 1, W], & f_0(W) &= \mathbb{E}[D(1-Z) \mid S = 1, W], \\
g_1(X,W) &= \mathbb{E}[DZ \mid S = 1, X, W], & g_0(X,W) &= \mathbb{E}[D(1-Z) \mid S = 1, X, W].
\end{aligned}
$$

The proof can be found in Appendix 7.1. In particular, note that the second nuisance parameter takes two form, i.e. $q(X, W) = q(X)$, given the independence assumption $W \perp S \mid X$ (since joint independence implies marginal independence). Thus this serves as a good robustness check: if the independence assumption holds, then the two sets of propensity scores should yield similar estimates. Note that the reverse direction does not apply.

For the main results, we focus on working with the sample analog of this score function, combined with 5-fold cross fitting to control variance (Chernozhukov et al., 2018). Then asymptotic normality holds under some moment restrictions and rate requirements, which are possible to verify but remains beyond the scope of this paper. We will then use bootstrapping to obtain standard errors.

5.2.1 Data Processing

As with other field surveys, the OHIE experiment witnessed two types of non-response: item non-response, where a responder fails to answer particular items on the survey, and unit non-response, when a potential participant declines to answer the survey as a whole. As seen from Section 2, while item non-response is certainly not negligible, we are primarily concerned with the potential systematic bias in the LATE estimate caused by unit non-response. This calls for some method to control for item non-responses among the responders.

Since item non-responses are dispersed but pervasive in the data set, appearing not only in outcomes of interest but also in potential covariates (especially those that belong to the missing value validating covariates X_i), we decide to apply k-nearest neighbors to impute these missing data using KNNImputer library from Python. In other words, we repeat some steps from data pre-processing for the causal forest, except that we exclude the initial mail survey to avoid potential unrelated bias. This approach is suitable because the algorithm imputes the data globally (i.e. over the dataset as a whole). However, the associated imputer function does have one drawback: it only works with numerical values and thus do not directly impute categorical variables, especially binary indicators. Instead of setting $k = 1$, which would resolve this issue but also potentially introduce more bias in the case of outliers, we tried larger values of k and then manually adjust the binary variables based on a threshold of 0.5. We worked with both the imputed and adjusted variables during the initial stage of our investigation and dropped the adjusted one after compared results during our first two-stage-least-squares estimates in Table 7.3 below.

To choose the appropriate k, we impute the data over several values of k and performed the same 2SLS procedure for LATE estimation as Finkelstein et al., 2012 did. Results can be found in Table 7.3 in the Appendix. In particular, note that the procedure allowed us to conduct two important robustness checks. First, we would like to see how much of the potential bias should be attributed to item non-response as opposed to unit non-response. For if controlling for item non-response already introduce significant bias, there is no reason to expect that the LATE estimates would be the same as Finkelstein et al. estimates after further adjustment for unit non-response. In addition, we would like to ensure that our imputation results are sensitive to different values of k.

5.3 Variable, Model & Feature Selection

Given the independence assumption in Assumption 1, the choices of W and X becomes crucial. In alignment with Finkelstein et al., 2012 and the set-up of the experiment, we take W to be the set of household size indicators, survey wave indicators and their interactions, as these should guarantee the first four assumptions and the proper conduction of the randomized controlled trial, barring any unobserved contamination. Finkelstein et al., 2012 includes an extensive discussion on the validity of the instrumental variable, so we will focus more on the selection of X here.

Since the independence assumption in Assumption 1 is inherently untestable, we rely on other approaches to find potential covariates for X. First, we focus on marginal correlation through plotting correlation heatmaps between the response indicator and all demographic characteristics and state program variables. Then we ran a random forest, again of the response indicator against all the demographic characteristics and state program variables. This allows us to identify important influencers of response rate under a joint distribution setting. We then again plot graphs of response indicator against top variables in terms of feature importance not identified above and stopped when the graph no longer shows imbalance from new variables. Interestingly, this corresponds to a feature importance threshold of 0.05. The procedure combined yield 8 covariates in total, which include follow-up protocol intensity, age, gender, phone possession, lottery sign up week, and benefits from other previous and current governmental programs. The response status graphs can be found in Appendix 7.2.

Given the size of our dataset and high dimensionality of covariates, we decided to use random forests to estimate the nuisance parameters. Due to computational feasibility involving 11 random forests, we were forced to use 100 trees for each random forest (and not anything higher), but we selected tree depth and maximum features using 5-fold cross-validation on the entire dataset.

5.4 Empirical Results

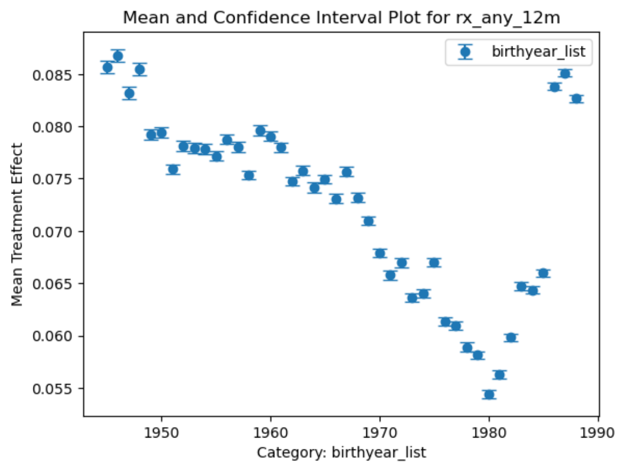

Figure 1: Mean HTE by Birth Year on Prescriptions

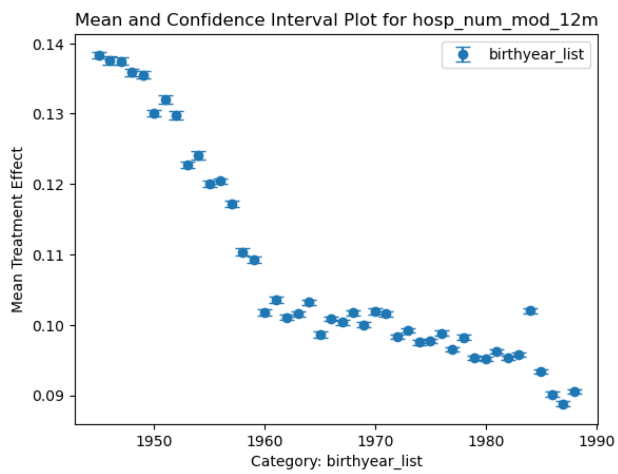

Figure 2: Mean HTE by Birth Year on Hospital Visits

Variable	(1)	(2)	(3)
	Health Utilization		
Prescription drugs, currently	0.088	-0.043	-0.050
(extensive margin)	(0.029)	(0.037)	(0.037)
Prescription drugs, currently	0.347	-0.100	-0.109
(total utilization)	(0.176)	(0.144)	(0.145)
Outpatient visits, last six months	0.212	0.133	0.130
(extensive margin)	(0.025)	(0.035)	(0.034)
Outpatient visits, last six months	1.083	0.649	0.660
(total utilization)	(0.182)	(0.188)	(0.184)
ER visits, last six months	0.022	-0.034	-0.035
(extensive margin)	(0.023)	(0.030)	(0.029)
ER visits, last six months	0.026	-0.064	-0.065
(total utilization)	(0.056)	(0.060)	(0.059)
Inpatient hospital admissions, last six months	0.008	-0.010	-0.010
(extensive margin)	(0.014)	(0.015)	(0.015)
Inpatient hospital admissions, last six months	0.021	-0.002	-0.000
(total utilization)	(0.021)	(0.023)	(0.023)
	Financial Strain		
Any OOP medical expense, last six months	-0.200	-0.272	-0.272
	(0.026)	(0.035)	(0.035)
Any OOP medical expense, last six months	-0.180	-0.273	-0.272
	(0.026)	(0.034)	(0.034)
Borrowed money or skipped bills for medical expense	-0.154	-0.217	-0.216
	(0.025)	(0.032)	(0.032)
Refused treatment due to medical debt	-0.036	-0.050	-0.051
	(0.014)	(0.019)	(0.019)
	Self-Reported Health		
Self-reported health (overall)	0.133	0.089	0.092
	(0.026)	(0.031)	(0.031)
Self-reported health (change)	0.113	0.048	0.050
	(0.023)	(0.037)	(0.037)

Table 1: Double Machine Learning LATE Estimates using $\dim(X) = 8$

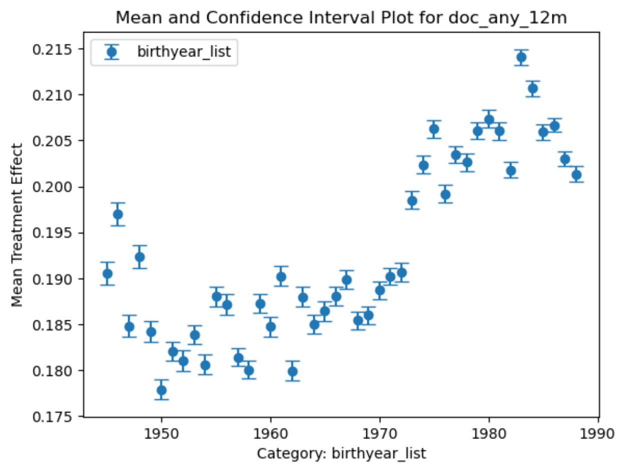

Figure 3: Mean HTE by Birth Year on Doctor's Visits

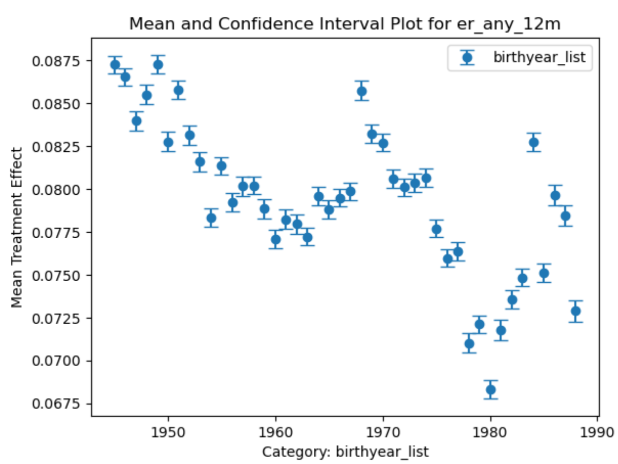

Figure 4: Mean HTE by Birth Year on ER Visits

6 Conclusion & Discussion

From our causal forest analysis for the computation of the HTEs by various subgroups. Most interestingly, we found that age plays a crucial role in the effects of insurance coverage on health outcomes. Our research revealed significant variations in these effects across different age groups. These variations could be attributed to the fundamental differences between these groups, such as varying health needs and risk factors associated with age. Our analysis of the results from this experiment showed that insurance coverage led to increased healthcare utilization, reduced financial strain, and improved mental health outcomes. However, the magnitude at which these effects impacted each age group exhibited very non-linear trends with respect to age and positively affected younger population more. These findings underscore the complex interplay between insurance coverage, age, and health outcomes, highlighting the need for policies that consider these heterogeneous effects. We did not see any significant differences between men and women. We found that people from smaller households exhibited similar behaviors to those from larger households when it came to the effects of insurance coverage on health outcomes. However, medium sized households showed a contrasting trend, tending to have an opposite (diminished or enhanced) reaction to treatment. This could be due to a variety of factors unique to medium-sized households, and further research is needed to fully understand these dynamics. There was no clear correlation between the treatment effect on health outcomes and the number of hours worked. This suggests that the impact of insurance coverage on health outcomes is not directly tied to changes in work hours. Further research is needed to explore these intricate dynamics, and we have identified the existence of these heterogeneous treatment effects.

Through our Double Machine Learning implementation after accounting for missing values, the first thing to note is that the differences between the two sets of LATE estimates exhibit different patterns based on the category of outcomes. In particular, for the health utilization outcomes, the estimates we obtained through DML are either diminished in magnitude or of opposite signs to those observed in previous estimates of the effect. From Table 7.3, we can see that item non-response only account for the effect partially, since the coefficients are mostly only diminished in magnitude. In contrast, the DML estimates show greater, significant effect for reduction in financial stress compared to Finkelstein et al., 2012's estimates. Combined, this means that unit non-response are more concentrated on those who would not benefit much directly from the health services covered by the insurance, but rather indirectly by the implied financial security. This makes intuitive sense by the imbalance traits shown in our motivation graphs (for instance, young people are likely to be healthier, thus needing less of the healthcare services, but nonetheless benefit from the associated financial security).

In addition, for the item non-response table, we imputed data with $k = 10$ and $k = 20$ and the differences between the 2SLS estimates are minimal (negligible). This shows that our results are more or less robust to different imputations and already partially compensate for the non-response bias.

Lastly, we found that the bootstrapped standard errors that the Double Machine Learning analysis is larger than Finkelstein et al., 2012's estimates. This caught us by surprise, as Neyman orthogonality is known to produce smaller standard errors. We attribute this to the variation introduced by accounting for the missing values. In the hypothetical scenario where the non-responding sample can be observed, the variance in response would likely result in even larger standard errors. However, this remains our theory, and we plan to explore it further as a next step.

For other potential directions, we plan to examine Finkelstein's one-dimensional X (which is a complicated set of inverse propensity scores) and apply the same procedure, to see if the one-dimensional covariate would be sufficient for the independence assumption to hold. Further, we plan to advance our analysis by using a different set of treatment and covariates than so far (such as using the lottery assignment itself as the treatment variable, an approach that is very common in literature). We can also check for longer run outcomes by leveraging the follow-up surveys in the next time periods.

References

Baicker, K., Allen, H. L., Wright, B. J., et al. (2017a). The effect of medicaid on dental care of poor adults: Evidence from the oregon health insurance experiment. *Health Service Research*, *53*, 2147–2164.

Baicker, K., Allen, H. L., Wright, B. J., et al. (2017b). Effect of medicaid on medication use among poor adults: Evidence from oregon. *Health Affairs*, *36*(12), 2110–2114.

Baicker, K., Allen, H. L., Wright, B. J., Taubman, S. L., & Finkelstein, A. N. (2017c). The effect of medicaid on management of depression: Evidence from the oregon health insurance experiment. *The Milbank Quarterly*, *96*, 29–56.

Baicker, K., Finkelstein, A., Song, J., & 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.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68. https://doi.org/https://doi.org/10.1111/ectj.12097

Denteh, A., & Liebert, H. (2023). Who increases emergency department use? new insights from the oregon health insurance experiment. http://arxiv.org/abs/2201.07072

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J. P., Allen, H., Baicker, K., & Oregon Health Study Group. (2012). The oregon health insurance experiment: Evidence from the first year. *The Quarterly Journal of Economics*, *127*(3), 1057–1106. https://doi.org/10.1093/qje/qjs020

Finkelstein, A., Taubman, S., Allen, H., Wright, B., & Baicker, K. (2016). Effect of medicaid coverage on ed use – further evidence from oregon's experiment. *New England Journal of Medicine*, *375*(16), 1505–1507.

Frölich, M. (2007). Nonparametric iv estimation of local average treatment effects with covariates. *Journal of Econometrics*, *139*(1), 35–75. https://EconPapers.repec.org/RePEc:eee:econom:v:139:y:2007:i:1:p:35-75

Goto, R., Inoue, K., Osawa, I., Baicker, K., Fleming, S. L., & Tsugawa, Y. (2024). Machine learning detects heterogeneous effects of medicaid coverage on depression [Advance online publication]. *American journal of epidemiology*. https://doi.org/10.1093/aje/kwae008

Hattab, Z., Doherty, E., Ryan, A. M., & O'Neill, S. (2024). Heterogeneity within the oregon health insurance experiment: An application of causal forests. *PLoS ONE*, *19*(1), e0297205. https://doi.org/10.1371/journal.pone.0297205

Johnson, M., Cao, J., & Kang, H. (2022). Detecting heterogeneous treatment effects with instrumental variables and application to the oregon health insurance experiment. *The Annals of Applied Statistics*, *16*(2), 1111–1129. https://doi.org/10.1214/21-AOAS1535

7 Appendix

7.1 Proof of Theorem 5.1

The idea is to construct each of $g^{1,1;o}$, $g^{1,0;o}$, $g^{2,1;o}$, and $g^{2,2;o}$ such that each component is Neyman orthogonal to its own nuisance parameters. We will only show the proof for $g^{1,1;o}$, as the others follow from the exact the same process (with repeating calculations). First,

$$
\begin{aligned}
\mathbb{E}[m(1,W)] = \mathbb{E}[\mathbb{E}[Y \mid Z=1,W]] &= \mathbb{E}\left[\frac{YZ}{r(W)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{YZ}{r(W)} \,\Big|\, W, X\right]\right] \\
&= \mathbb{E}\left[\frac{1}{r(W)}\mathbb{E}[YZ \mid W, S=1, X]\right] \\
&= \mathbb{E}\left[\frac{1}{r(W)} \cdot \frac{YZS}{\mathbb{P}[S=1 \mid W, X]}\right] \\
&= \mathbb{E}\left[\frac{YZS}{r(W)q(W,X)}\right],
\end{aligned}
$$

where the second equality follows from the law of iterated expectations (LIE).

We will keep working $q(X,W)$ and switch it out for $q(X)$ at the very end, as this is much more convenient. The above yields

$$
g^{1,1}(A; r, q) = \frac{YZS}{r(W)q(X,W)}.
$$

Then

$$
g^{1,1}_{2,1}(A; r, q) = \frac{YZS}{q(X,W)} \cdot \left(-\frac{1}{[r(W)]^2}\right),
$$

and

$$
\begin{aligned}
\mathbb{E}\left[g^{1,1}_{2,1}(A,;r,q)\right] &= \mathbb{E}\left[-\frac{YZS}{[r(W)]^2 q(X,W)} \;\Big|\; W\right] \\
&= -\frac{1}{[r(W)]^2}\mathbb{E}\left[\mathbb{E}\left[\frac{YZS}{q(X,W)} \;\Big|\; X,W\right] \;\Big|\; W\right] \\
&= -\frac{1}{[r(W)]^2}\mathbb{E}[\mathbb{E}[YZ \mid S=1,X,W] \mid W] \\
&= -\frac{1}{[r(W)]^2}\cdot \mathbb{E}[YZ \mid S=1,W] =: \frac{a_1(W)}{[r(W)]^2}.
\end{aligned}
$$

On the other hand,

$$
g^{1,1}_{2,2}(A;r,q) = \frac{YZS}{r(W)}\cdot\left(-\frac{1}{[q(X,W)]^2}\right),
$$

so

$$
\begin{aligned}
\mathbb{E}\left[g^{1,1}_{2,2}(A;r,q)\right] &= \mathbb{E}\left[-\frac{YZS}{r(W)[q(X,W)]^2} \;\Big|\; X,W\right] \\
&= -\frac{1}{r(W)[q(X,W)]^2}[\mathbb{E}[YZS \mid X,W] \\
&= -\frac{\mathbb{E}[YZ \mid S=1,X,W]}{r(W)[q(X,W)]} =: -\frac{b(X,W)}{r(W)q(X,W)} = -\frac{b_1(X,W)}{r(W)\cdot q(X)}.
\end{aligned}
$$

Combined, this allows us to define

$$
g^{1,1;o}(A;r,q,a_1,b_1) = \frac{YZS}{r(W)q(X,W)} - [Z-r(W)]\cdot\frac{a_1(W)}{[r(W)]^2} - [S-q(X,W)]\cdot\frac{b_1(X,W)}{r(W)\cdot q(X,W)}.
$$

Next, we need to verify Neyman orthogonality (N.O.). It is not hard to see that

$$
\mathbb{E}\left[g^{1,1,0}(A;r,q,a_1,b_1)\right] = \mathbb{E}[m(1,W)],
$$

so we concentrate on the N.O. part. Notation-wise, we substitute the r in the lecture notes for α to get

$$
\begin{aligned}
h_r(\alpha) &= \mathbb{E}\left[g^{1.1.0}(A;r+\alpha(\tilde{r}-r),q,a_1,b_1)\right] \\
&= \mathbb{E}\left[\frac{YZS}{[r+\alpha(\tilde{r}-r)]\cdot q} - \frac{Z\cdot a_1}{[r+\alpha(\tilde{r}-r)]^2} + \frac{a_1}{[r+\alpha(\tilde{r}-r)]} - \frac{(S-q)\cdot b_1}{[r+\alpha(\tilde{r}-r)]\cdot q}\right],
\end{aligned}
$$

So

$$
\begin{aligned}
h_r'(\alpha)\Big|_{\alpha=0} &= \mathbb{E}\left[\frac{-YZS(\tilde{r}-r)}{q\cdot r^2} + \frac{2Z\cdot a_1(\tilde{r}-r)}{r^3} - \frac{a_1(\tilde{r}-r)}{r^2} + \frac{(S-q)\cdot b_1(\tilde{r}-r)}{q\cdot r^2}\right] \\
&= \mathbb{E}\left[\frac{\tilde{r}-r}{r^2}\mathbb{E}\left[\frac{-YZS}{q} + \frac{2Z\cdot a_1}{r} - a_1 + \frac{(S-q)\cdot b_1}{q} \;\Big|\; W\right]\right] \\
&= \mathbb{E}\Bigg[\frac{\tilde{r}-r}{r^2}\Bigg(\mathbb{E}\left[\frac{\mathbb{E}[YZS \mid X,W]}{\mathbb{P}[S=1\mid X,W]} \;\Big|\; W\right] + \frac{2\mathbb{E}[YZ \mid S=1,W]}{\mathbb{E}[Z\mid W]}\cdot \mathbb{E}\big[Z\mid X,W] \mid W\big] \\
&\quad - \mathbb{E}[YZ\mid S=1,W] + \mathbb{E}\left[\frac{\mathbb{E}[S\mid X,W]\cdot\mathbb{E}[YZ\mid S=1,X,W]}{\mathbb{E}[S\mid W,X]} \;\Big|\; W\right] \\
&\quad - \mathbb{E}\big[\mathbb{E}[YZ\mid S=1,X,W]\mid W]\Bigg)\Bigg] \\
&= \mathbb{E}\left[\frac{\tilde{r}-r}{r^2}\cdot 0\right] = 0
\end{aligned}
$$

where the second-to-last equation follows from expanding the nuisance parameters and applying LIE term by term. Similarly,

$$
\begin{aligned}
h_q(\alpha) &= \mathbb{E}\left[g^{1,1;o}(A; r, q+\alpha(\tilde{q}-q), a_1, b_1)\right], \\
&= \mathbb{E}\left[\frac{YZS}{r[q+\alpha(\tilde{q}-q)]} - (Z-r)\cdot\frac{a_1}{r^2} - \frac{S\cdot b_1}{r[q+a(\tilde{q}-q)]} + \frac{b_1}{r}\right],
\end{aligned}
$$

This yields

$$
h_q'(\alpha)\Big|_{\alpha=0} = \mathbb{E}\left[-\frac{YZS(\tilde{q}-q)}{rq^2} + \frac{s\cdot b_1(\tilde{q}-q)}{r\cdot q^2}\right] = 0
$$

again by applying LIE term by term. Lastly,

$$
\begin{aligned}
h_{a_1}(\alpha) &= \mathbb{E}\left[g^{1,1;o}(A; r, q, a_1+\alpha(\tilde{a}_1-a_1), b_1)\right] \\
&= \mathbb{E}\left[\frac{-(Z-r)\cdot[a_1+\alpha(\tilde{a}_1-a_1)]}{r}\right] = 0 \\
\Longrightarrow h_{a_1}'(\alpha)\Big|_{\alpha=0} &= 0,
\end{aligned}
$$

and

$$\begin{aligned} h_{b_1}(\alpha) &= \mathbb{E}\left[g^{1,1;o}(A; r, q, a_1, b_1 + \alpha(\tilde{b}_1 - b_1)\right] \\ &= \mathbb{E}\left[-(S-q) \cdot \frac{[b_1 + \alpha(\tilde{b}_1 - b_1)]}{r \cdot q}\right] = 0 \\ &\implies h'_{b_1}(\alpha)\Big|_{\alpha=0} = 0. \end{aligned}$$

Repeating this process for all four terms lead to the score function, hence the result.

7.2 Graphs

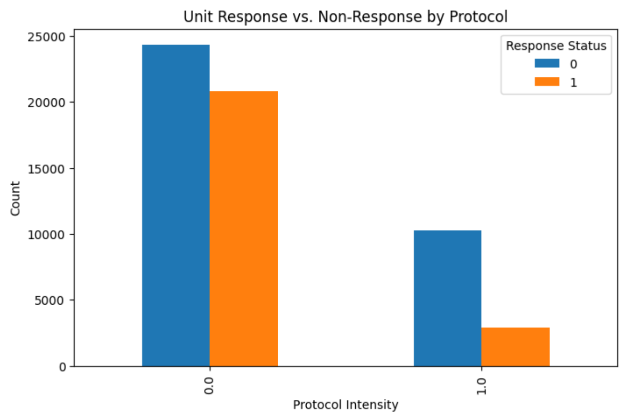

Figure 5: Unit Non Response by Protocol

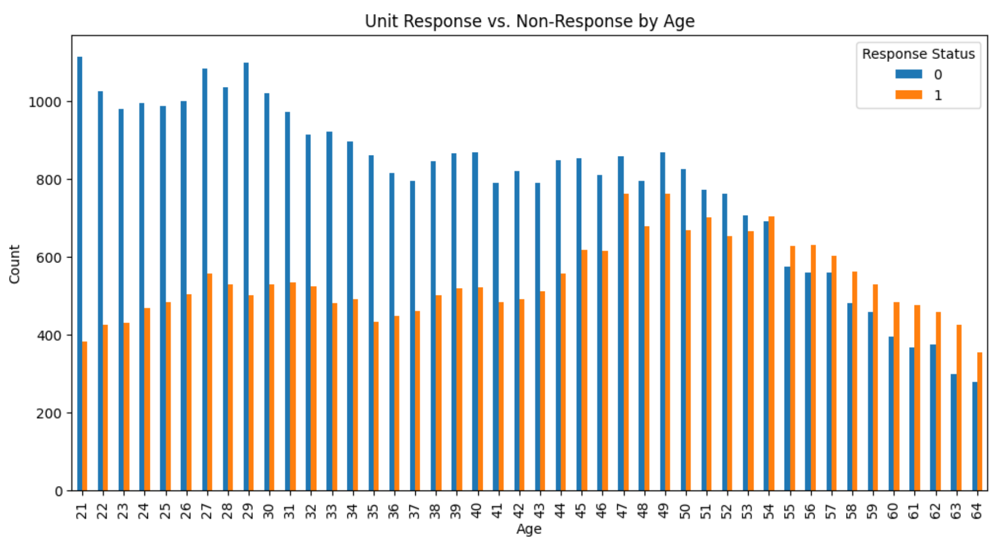

Figure 6: Unit Non Response by Age

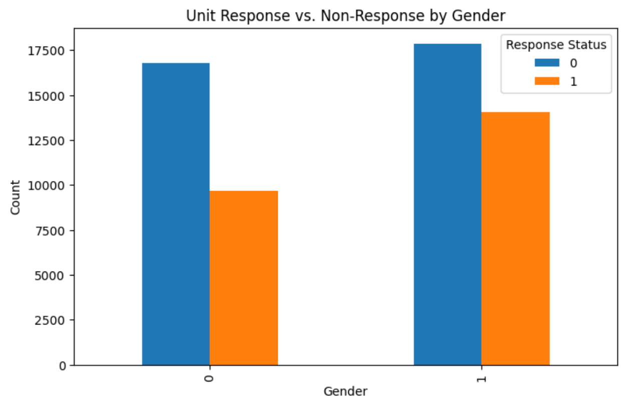

Figure 7: Unit Non Response by Gender

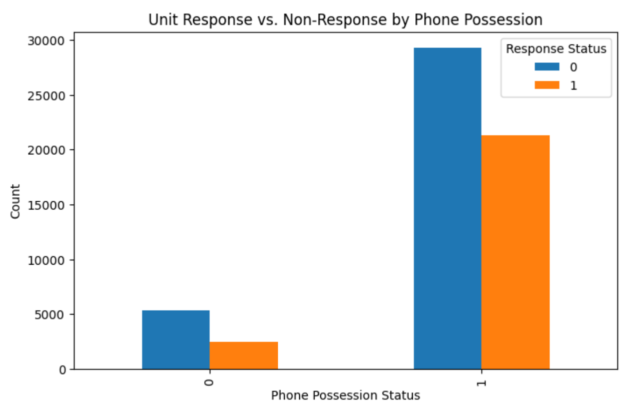

Figure 8: Unit Non Response by Phone Possession

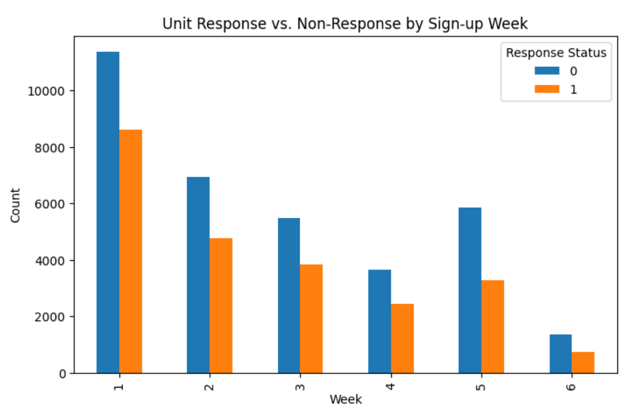

Figure 9: Unit Non Response by Sign-Up Week

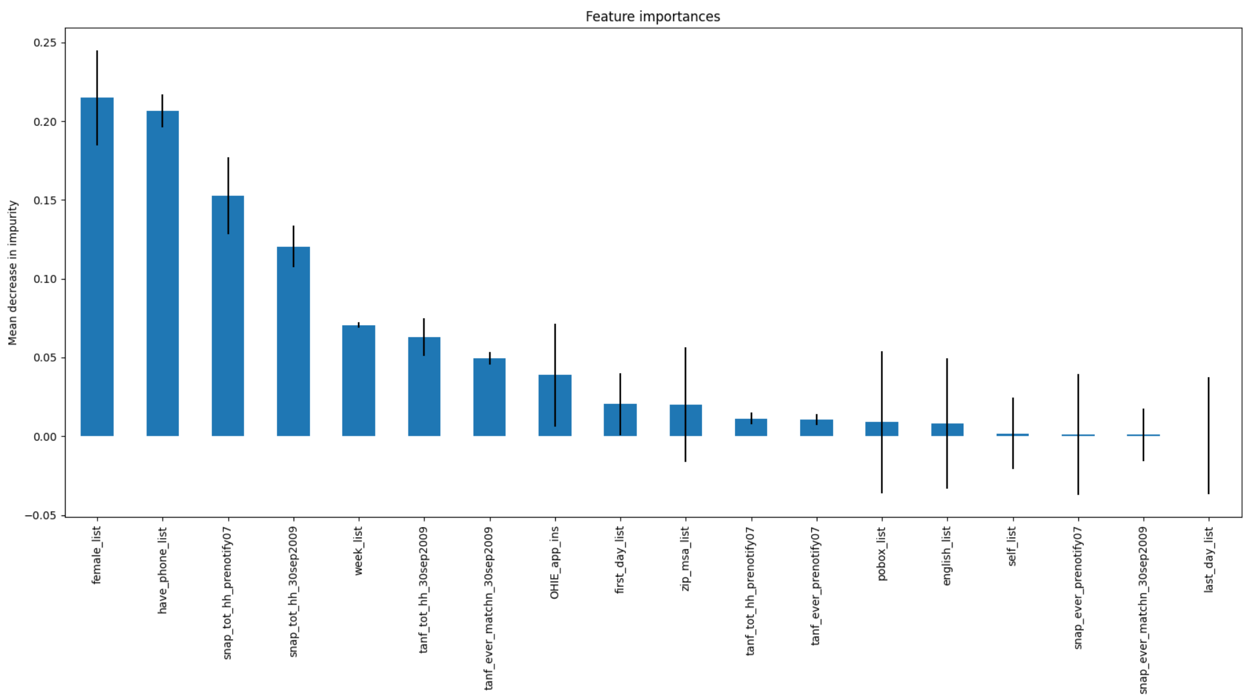

Figure 10: Feature Importance for DML

7.3 Tables

	(1)	(2)	(3)	(4)	(5)	(6)	(7) [5]
	Healthcare Utilization						
Prescription drugs, currently	0.088	0.026	0.015	0.023	0.003	0.026	0.008
(extensive margin)	(0.029)	(0.022)	(0.024)	(0.021)	(0.023)	(0.021)	(0.023)
Prescription drugs, currently	0.347	0.042		0.039		0.062	
(total utilization)	(0.176)	(0.131)		(0.130)		(0.129)	
Outpatient visits, last six months	0.212	0.187	0.188	0.187	0.187	0.188	0.190
(extensive margin)	(0.212)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Outpatient visits, last six months	1.083	0.909		0.903		0.906	
(total utilization)	(0.182)	(0.173)		(0.173)		(0.173)	
ER visits, last six months	0.022	-0.002	-0.004	-0.002	-0.002	-0.002	-0.001
(extensive margin)	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)
ER visits, last six months	0.026	-0.028		-0.029		-0.029	
(total utilization)	(0.056)	(0.053)		(0.053)		(0.053)	
Inpatient hospital admissions, last six months	0.008	-0.000	0.001	0.001	-0.000	0.000	-0.000
(extensive margin)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Inpatient hospital admissions, last six months	0.021	0.015		0.014		0.014	
(total utilization)	(0.021)	(0.020)		(0.020)		(0.020)	
	Financial Strain						
Any OOP medical expense, last six months	-0.200	-0.204	-0.206	-0.204	-0.205	-0.203	-0.203
	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Owe medical expense, currently	-0.180	-0.201	-0.200	-0.202	-0.202	-0.201	-0.202
	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Borrowed money or skipped bills for medical expense,	-0.154	-0.162	-0.161	-0.162	-0.161	-0.162	-0.162
last six months	(0.025)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Refused treatment due to medical debt,	-0.036	-0.044	-0.039	-0.043	-0.038	-0.043	-0.038
last six months	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
	Health						
Self-reported health (overall)	0.133	0.151	0.152	0.152	0.153	0.151	-0.153
	(0.026)	(0.024)	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Self-reported health (change)	0.113	0.124	0.123	0.123	0.122	0.124	0.124
	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)

Table 2: Two-Stage-Least-Squares Results after Controlling for Item Non-Response

[5]Column (1) shows Finkelstein et al. (2012)'s estimates, columns (2)-(7) replicates Finkelstein et al.'s procedure on the dataset post K-NN imputation controlling for item non-response for $K = 5, 10, 20$. Specifically, columns (2), (4), (6) show estimates directly from imputed data, while columns (3), (5), (7) show results for manually adjusted binary outcomes.