CAUSAL STUDY OF EFFECT OF AFFORDABLE CARE ACT ON HEALTH OUTCOME

Shubhankar Mitra
Department of Data Science
Indiana University
Bloomington, IN
shmitra@iu.edu

Ankit Swarnkar
Department of Data Science
Indiana University
Bloomington, IN
ankswarn@iu.edu

Saurabh Kumar
Department of Data Science
Indiana University
Bloomington, IN
kumarsau@iu.edu

Affordable care act (ACA) has been the topic of many debates. Studies have been conducted in the past to understand its impact on insurance enrollment and labor markets. Easier access to health insurance via ACA might improve public's ability to get health checkups and improve chronic disease diagnosis rate. Here we try to understand the impact of ACA enforcement on chronic disease detection and health outcomes. ACA was adopted in 32 states by end of 2013 while 19 states decided to not adopt ACA. This creates an opportunity to understand the effect of ACA via a difference in difference methodology. Behavioral Risk Factor Surveillance System (BRFSS) data is used to understand the effect of ACA.

Keywords: ACA, DID, BRFSS, Obama care act effect.

1 Introduction

In January 2014, under the Affordable Care Act (ACA), twenty-five states and the District of Columbia expanded Medicaid eligibility to individuals ages 19–64 with family incomes below 138 percent of the federal poverty level. The ACA Medicaid expansion has greatly increased insurance coverage among low-income nonelderly adults in states that implemented the expansion.

Expansion of insurance has the possibility of improving the accessibility of health services for low-income individuals. Subsequently, it may lead to increased detection of chronic health diseases like diabetes. This study aims to understand the effect of ACA expansion on the same. Since ACA was adopted in certain states on January 2014, it provides an opportunity to apply the difference in difference (DID) methodology to understand ACA's impact. Since the work by Ashenfelter and Card (1985)[1], the use of difference-in-differences methods has become very widespread. We combine DID along with matching on covariates to provide a robust estimate of ACA's impact. We also tested sensitivity of our claim of the estimate from unobserved covariates using Rosenbaum's sensitivity analysis. The particular outcomes which were tested for ACA's

impact are: Visit to doctor in last one year, health care coverage, diagnosis of diabetes and diagnosis of kidney disease.

2 Data

"The Behavioral Risk Factor Surveillance System (BRFSS) is the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services".[2] This data is used to study the health conditions as well as an impact of federal decisions on them.

2.1 Data Collection

The survey questionnaire has standard core questions, rotating core questions, optional modules and state added questions. States can either employ private company or university to collect the data, which is done using BRFSS protocols. These protocols help to maintain consistency across different states. An interview is considered complete if at least half of the core questions are answered. BRFSS uses 2 samples, one for landline telephones and one for cellular telephones. Landline samples are considered multiple adult household while cellular samples are considered as single adult households. To remove bias in the data, data weighting is done. The BRFSS weighting has 2 steps: design weighting and iterative proportional fitting.

2.2 Data Description

The combined landline and cellular data consists of 355 variables. Most of the variables are survey questions. The rest of the variables are derived variables. Not every state has all the variables, but they still do have the core questions. The survey collects data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Most of variables are categorical.

2.3 Data Processing

We extracted the data for years 2006 to 2015. First, we filtered out 47 relevant variables from the entire dataset. Unrelated and low-quality columns with a high percentage of non-response were filtered. The chronic diseases which were not part of the outcome were combined to form a data column indicating the count of chronic disease. We ended up with 19 covariates and 4 treatment variables for kidney stones, diabetes, checkups and health plan. We processed the categorical variables to convert the entries to a standard form of 0 for 'no' and 1 for 'yes'. An additional categorical variable was created to specify ACA expansion states. Fig 1 shows the states that implemented ACA.

The final list of covariates which might act as confounders and were used for matching and control are: number of adults in household, annual income of household, employment

information, lack of access to healthcare due to cost, physical activity, military duty, general health of respondent, existence of personal doctor, using any medical equipment, number of suffering chronic disease excluding the outcome, education level, gender, renting or having bought a house, type of insurance being expanded in the state (state based marketplace, state partnership or federally facilitated).

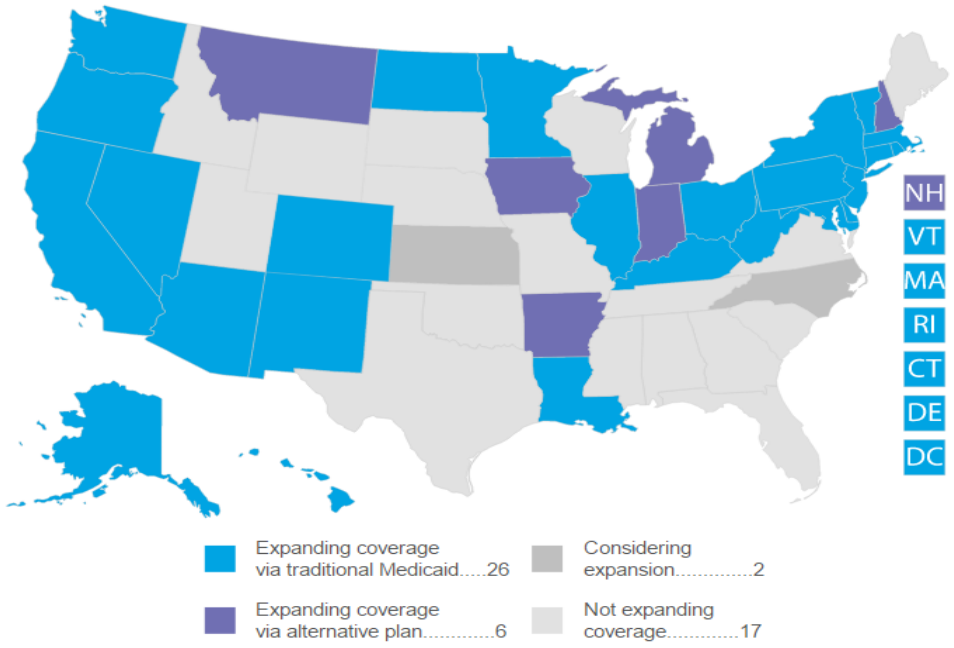


Fig 1: States that implemented Affordable Health Care Act[3]

2.4 Data Exploration

ACA was implemented in year 2014. So, yearly data before 2014 becomes part of the control group. For the 4 treatment variables, we plot the percentage of people affected, i.e. the percentage of people that answered yes, across different years. Fig 2(a), 2(b) and 3(b) show plots for Health Plan enrollment, Checkup in last one year and Diabetes diagnosis respectively. These are plotted for years 2006 to 2015. Due to limited data for Kidney disease, we plot the percentage graph from years 2011 to 2015. The blue line represent the treated group while green represents the control group. The vertical divider in the graph represents the year after which ACA came into effect in certain states.

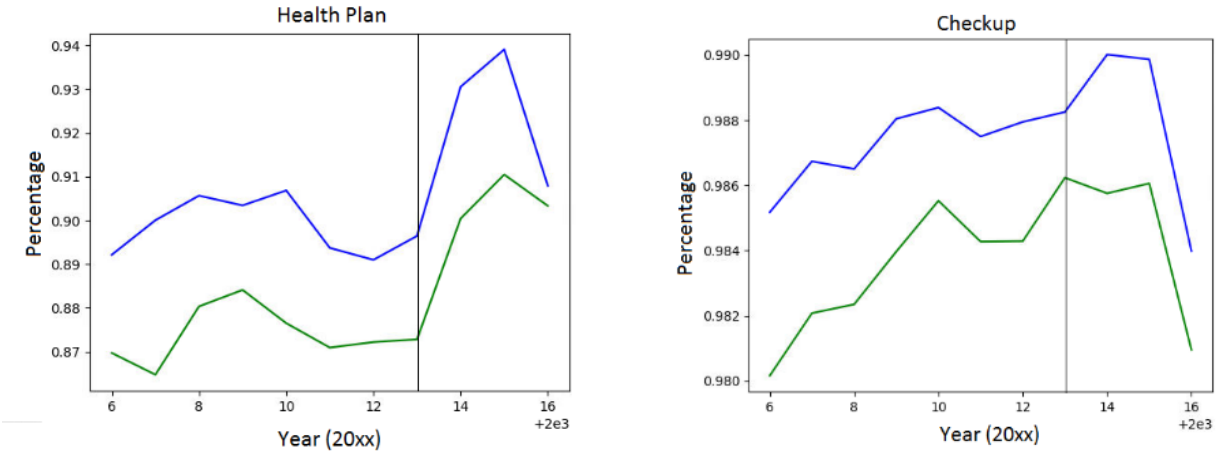


Fig: 2a) Health Plan enrollment percentage over years 2b) Checkup in last one year

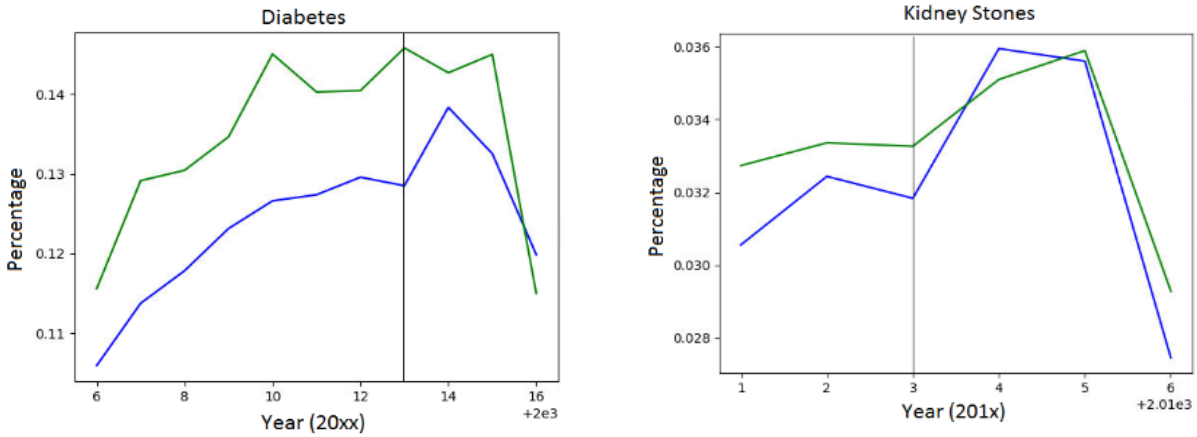


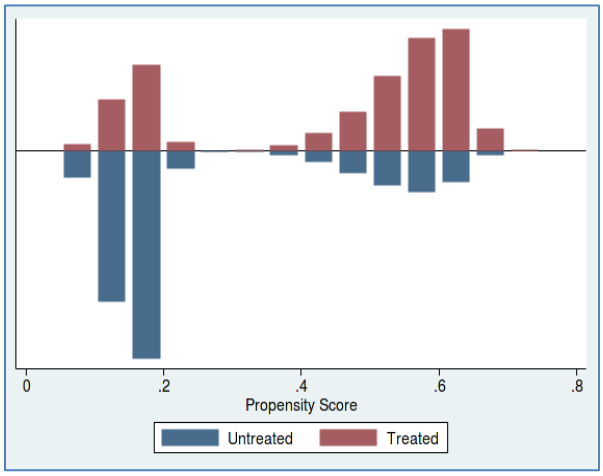
Fig 3: Percentage of Population reported having a) Diabetes b) Kidney disease

In Fig 3, Diabetes and Kidney disease detection show an uptake in disease diagnosis following the year 2013 in the ACA expansion states. In Fig 2, Health plan enrollment and doctor's checkup also show a considerable higher uptake for the expansion states compared to non-expansion states. The pre-treatment trends looks to be similar in the four outcomes. This provides us an indication to explore further the impact of ACA in the above discussed outcomes.

3 Causal Analysis

We first perform a simple multivariate regression analysis to understand the impact of treatment controlling for all the covariates. We then perform propensity score matching for the 4 treatment variables to check if covariates are balanced.

Fig 4 shows the presence of counterfactuals at almost all levels of propensity scores.



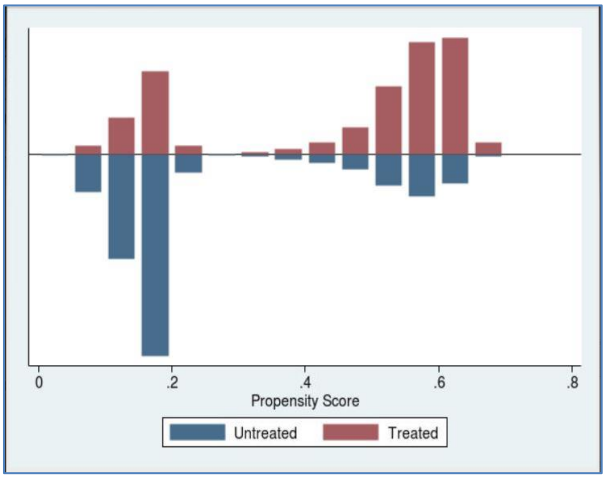
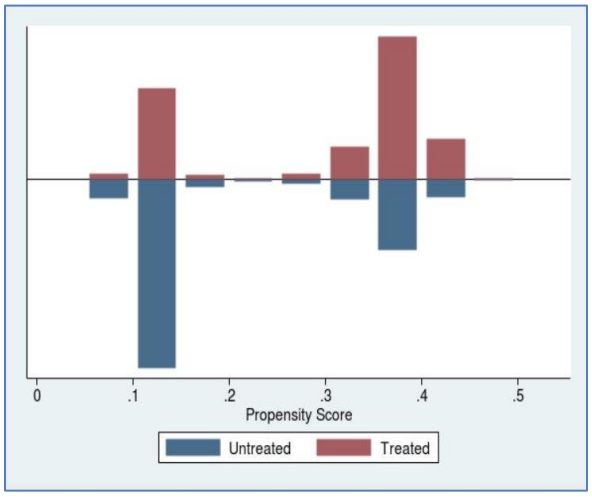


Fig 4: Propensity Score matching for a) Health Plan b) Checkups



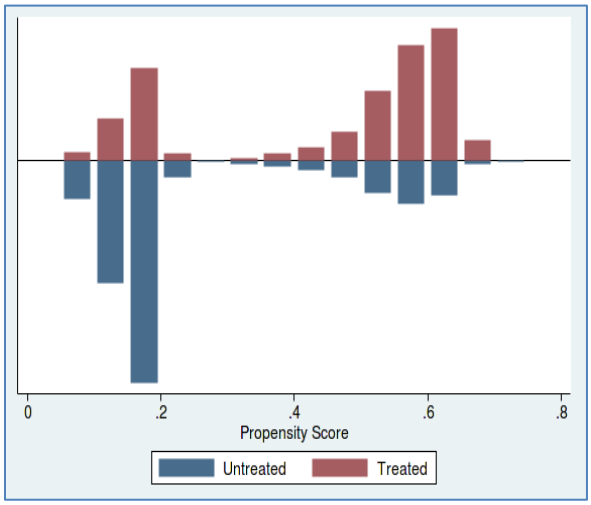


Fig 4: Propensity Score matching for c) Kidney Stones d) Diabetes

3.1 Causal Model

Given our data is panel data, we selected to follow difference in difference framework[4] to find the average treatment effect. Difference in difference is a type of fixed affect estimation model which relies on a less strict exchangeability assumption, i.e., in absence of treatment, the unobserved differences between treatment and control groups are the same overtime. In this modelling, we assume that in absence of ACA changes, the impact on health outcomes and accessibility is determined by sum of time invariant state effect (Υ_s) and a year effect (α_t) that is common across all states i.e.

$$\mathbf{E[Y^{t1}_0|s,t] = \Upsilon_s + \alpha_t}$$

We use the same specification as in Waldinger and Havnes(2010)[5] modeling. Our baseline regression model, estimated by OLS over the sample during the period 2009–2015, can be defined as:

$$\mathbf{Y_{DS}^{t} = \beta_0 + \beta_1 D_s + \beta_2 T_t + \beta_3 (T_t * D_s) + \beta_4 (X) + \varepsilon}$$

Let Y_{DS}^{t} be the outcome variable under consideration, where D_S indicate whether it is an ACA or non-ACA state, T_t is the time factor to indicate whether ACA is implemented or not and X is the collection of covariates.

We used three methods in conjunction with DID framework. For baseline estimates we started with DID framework and logit regression. We then used two frameworks based on whether propensity score is used or not used for matching framework with DID. First one based on simple quasi experiment without using propensity score mechanism. For second framework, we used propensity score matching, As Jenni (1998)[6] mentioned in her study difference in difference matching estimates are optimal way to study the longitudinal data. Using the similar framework as used by Jenni (1998), let t and t' be the two-time period, aca and non-aca time indicator. Y^{t}_{0} is the outcome observed at time t. Condition which can justify the difference in difference estimators is:

$$\mathbf{E[Y^{t}_{0} - Y^{t'}_{0} \mid P(X), D=1] = E\,[Y^{t}_{0} - Y^{t'}_{0} \mid P(X), D=0]}$$

If the above condition are validated we can estimate the effect as follow:

$$\mathbf{Effect = n_{1t}^{-1} \sum_{i\ \{D=1\}} \{Y_1{}^{t}_{i}\,(X_i) - E(Y_0{}^{t}_{i} | P(X_i, D_i=0))\} - n_{1t'}^{-1} \sum_{j\ \{D=1\}} \{Y_1{}^{t'}_{i}\,(X_j) - E(Y_0{}^{t'}_{j} | P(X_j, D_j=0))\}}$$

Where n_{1t} and $n_{1t'}$ are the number of observation in both the group.

4 Output

	Healthcare Plan ATE	SE	Diabetes Detection ATE	SE	Kidney disease Detection ATE	SE	Check up ATE	SE
DID (OLS)	**0.08334****	**0.001**	**.04039*****	**0.012**	**.2735****	**0.04**	**.0468*****	**0.052**
DID (quasi-experimental)	**0.018*****	**0.002**	**0.002**	**0.002**	**0.007*****	**0.001**	**0.001****	**0.01**
Did with propensity score matching (2)	**.023959068**	**0.001**	**.01306**	**0.013**	**.009**	**0.011**	**.0005**	**.001**

After matching on propensity score it was seen that the covariates were balanced across the treatment variable.

P -Value :

* Significance pattern: P < 0.1. ** Significance pattern: P < 0.05. *** Significance pattern: P < 0.01.

Sensitivity Analysis

Rbound

	Health	**Kidney**	**Diabetes**	**Check Up**
Rbound Gamma (ṙ) +	**1.5**	**2**	**2**	**1**
Rbound Gamma (ṙ) -	**< 1**	**<1**	**1.2**	**1.1**

5 Conclusion

The DID framework with multivariate logistic regression (controlling for covariates) and basic quasi experiment showed promising results with significant treatment effects for all four outcomes. When we used propensity scores to match the covariates across the treatment none of the outcomes were seen to be significantly impacted by ACA. This shows that after matching counterfactuals ACA does not have a significant impact. The sensitivity analysis also shows that the results are sensitive to unobserved factors. The reason for non-existence of ACA impact on matching counterfactuals could be due to the long pre-treatment timeline which might have other events impacting the outcomes. In Fig 2 and Fig 3, regarding the post treatment we observe that after the sudden spike in 2014 the percent difference between expansion and non-expansion state reduces. An analysis focusing on years close to the ACA implementation might show significant effects.

6 References

Orley Ashenfelter, 1985, Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs, The Review of Economics and Statistics, Vol. 67, No. 4 (Nov., 1985), pp. 648-660

https://www.cdc.gov/brfss/annual_data/annual_data.html

https://www.advisory.com/daily-briefing/resources/primers/medicaidmap

Tarjei Havnes, 2010, Is Universal Child Care Leveling the Playing Field? Evidence from Non-Linear Difference-in-Differences, IZA Discussion Paper No. 4978

Fabian Waldinger, 2009, Differences-in-Differences lecture ppt

Pratical guide to Matching Estimates, 1998

Difference-in-Difference Estimation, Columbia University Mailman School of Public Health

Lechner, Michael. 1999. ‘‘Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumptions.’’ IZA Discussion Papers no. 91, University St. Gallen.

Heckman, James, H. Ichimura, and P. Todd. 1998. ‘‘Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program.’’ Review of Economic Studies 64:605–654.

King, Gary. 1991. ‘‘‘Truth’ Is Stranger than Prediction, More Questionable than Causal Inference.’’ American Journal of Political Science 35(4):1047–1053. http://gking.harvard.edu/files/abs/truth-abs.shtml.