Final Project Report

Yiyue Wang (yw3750)

Identification Strategy in General: Instrumental Variable

The purpose of the project is to replicate Taubman et al (2014) and identify the causal effect of Medicaid coverage on average emergency department usage per person.

Taubman et al (2014) adopted lottery selection as an instrumental variable as part of the identification strategy. Since not everyone won the lottery were eventually enrolled in Medicaid, directly regressing ED use on medical coverage would incur selection bias, i.e., some who won the lottery did not apply, and some who applied did not get approval. To minimize selection bias, this analysis continued adopting the approach of using lottery selection as an instrumental variable for Medicaid coverage, utilizing lottery's random assignment to isolate the causal effect of Medicaid coverage. Specifically, this analysis estimated a local average treatment effect capturing the causal effect of Medicaid for those who were covered due to the lottery, under the assumption that the lottery only affects ED utilization through Medicaid coverage. In Taubman et al's earlier work, they did not find cause for concern regarding potential threats to this assumption, and they tested that the nonrandom take-up of Medicaid among those selected in the lottery reduced statistical power but did not confound the causal interpretation of the effect of Medicaid.

Hence, after sample selection, data preprocessing, and balance check, this analysis first examined whether the strong first stage assumption and the exclusion restriction held and then proceeded to identify the causality using 2SLS models. The analysis was conducted at individual level. However, it is worth noting that as the state randomly selected individuals from the lottery list while covering all the selected individuals' household members, an individual's probability of winning the lottery varied by the number of the individual's household members on the list. To account for this, all analyses controlled for variable 'numhh' which had been factorized from 'numhh_list' to specify the individual's number of household members on the list.

Sample Selection

To replicate the study of Taubman et al (2014), I adopted the sample consists of individuals in Portland-area postal codes (N = 24,646), also referred to as the emergency-department (ED) analysis sample from dataset 'oregonhie_ed_vars'.

Data Pre-processing

I merged the datasets into a data frame named 'ed_sample' based on the unique 'person_id' from the 'oregonhie_ed_vars' dataset and got dummies for all the categorical variables of interest. I further dropped the rows whose outcome variables include N/A value. The sample size after pre-processing reduced to N = 24605.

Balance Check

To ensure that the assignment to treatment or control is well-randomized, I performed OLS balance check to observe whether the treatment and control groups are identically distributed and thus comparable, conditioned on the number of people in the household. The results are listed in Table 1.

Table 1. Treatment-Control Balance.

	Medicaid Coverage		Lottery Selection	
Feature	**t-statistics**	**p-value**	**t-statistics**	**p-value**
Requested English Materials	-0.406	.684	2.098	.036
Signed self up	0.295	.768	0.295	.768
Signed up for lottery on the first day	6.869	<.001	1.572	.116
Gave Phone Number	1.147	.251	0.650	.516
POBOX	0.295	.768	0.534	.593
Female	15.623	<.001	-1.480	.139
Birth Year	3.255	.001	0.602	.547
Pre-randomization Categorical Outcome Variables	15.433	<.001	1.166	.244
Pre-randomization Real-Valued Outcome Variables	10.211	<.001	-1.361	.173

Note. The feature is considered imbalanced if p-value < .05. Pre-randomization Categorical Outcome Variables of each person is measured by the sum of the categorical values of the pre-randomization categorical outcome variables. Pre-randomization Real-Valued Outcome Variables of each person is measured by the sun of the standard-scaled values of the pre-randomization real-valued outcome variables.

As indicated by the above results, it is worth noting that because the treatment Medicaid Coverage not only included the program decided by lottery selection, many of the features were imbalanced due to the nonrandom nature of the other Medicaid programs, i.e., women and the elderly are more likely to be enrolled in other Medicaid programs. However, the data supported that the selection bias was minimized in the process of lottery selection.

Treatment Variable, Instrumental Variable, and Outcome Variables

a) **Treatment Variable**

I used 'ohp_all_ever_matchn_30sep2009' from dataset 'oregonhie_stateprograms' as the treatment variable, which specifies if the person has ever been enrolled in Medicaid from matched notification date to 30 Sep 2009, and this variable was used as the definition of insurance coverage in estimating the effect of Medicaid coverage.

b) Outcome Variables

The outcome variables are from dataset 'oregonhie_ed_codebook'. Main outcome variables include 'any_visit_ed' (Any ED visit in the study period) and 'num_visit_cens_ed' (Number of ED visits in the study period (Censored)). Other outcome variables include Emergency-department use by hospital admission, timing, type of visit, etc.

c) Instrumental Variable

Taubman et all (2014) used treatment selection (i.e., whether the person was selected by the lottery or not) as an instrumental variable to identify the causal effect of Medicaid on ED utilization. To replicate this approach, I used the dummy variable created from the variable 'treatment' in dataset 'oregonhie_descriptive_vars' named 'treatment_Selected' as the instrumental variable. To testify its validity, I performed the following analysis.

Strong First Stage Assumption: Lottery selection and Medicaid Coverage should be correlated. In this study, people won lottery to get Medicaid, so there should exist a strong correlation. To test this assumption, I performed an OLS weak instrument test, reporting F-statistics = 1026 > 10, which suggested that Lottery Selection as the instrumental variable satisfied the strong first-stage assumption.

Exclusion Restriction: Conditioned on Medicaid Coverage and the number of people in the household, Lottery Selection should be uncorrelated with the outcome variables and other confounders. In other words, only winning the lottery but not getting Medicaid Coverage should have no effect on the Emergency Department Utilization. This hypothesis is difficult to test without a randomized experiment but should be self-evident.

2SLS

To estimate the causal effect of Medicaid coverage on ED utilization using lottery selection as an instrumental variable, a series of 2SLS models were fitted to identify the causal relationships between lottery selection, Medicaid coverage and different outcome variables. The results are reported in Table 2.

Table 2. 2LS Results.

	Percent with any visits				Number of visits			
	N	**Mean value in control group**	**Effect of Medicaid coverage**	**P-value**	**N**	**Mean value in control group**	**Effect of Medicaid coverage**	**P-value**
				Overall				
All Visits	24605	30.16	9.07 (2.45)	<.001	24591	0.81 (2.14)	0.37 (0.12)	.002
				By emergency-department use in the pre-randomization period				
No visits	16930	19.89	8.32 (2.74)	.002	16930	0.36 (1.01)	0.28 (0.08)	<.001
One visit	3880	43.07	12.07 (5.96)	.042	3880	0.99 (1.85)	0.68 (0.25)	.006
Two+ visits	2874	63.49	11.63 (6.55)	.075	2871	2.08 (2.91)	0.16 (0.42)	.700
Five+ visits	921	86.77	1.08 (8.61)	.900	910	5.94 (6.04)	2.60 (1.71)	.128
Two+ outpatient visits	4126	100.00	7.55 (1.33)	<.001	4125	4.47 (3.84)	0.68 (0.43)	.108
				By hospital admission				
Inpatient Visits	24605	5.97	-0.75 (1.34)	.573	24605	0.09 (0.45)	-0.02 (0.03)	.508
Outpatient Visits	24605	28.2	10.25 (2.41)	<.001	24591	0.72 (1.97)	0.40 (0.11)	<.001
				By timing of visit				
On-hours visits	24605	21.8	6.47 (2.25)	.004	24594	0.45 (1.30)	0.19 (0.07)	.013
Off-hours visits	24605	18.9	7.71 (2.16)	<.001	24599	0.36 (1.16)	0.21 (0.07)	.001

Required immediate care				
Emergent, not preventable	24597	0.17	0.05	.128
(Required ED care, could not have been prevented)		(0.64)	(0.03)	
Emergent, preventable	24603	0.06	0.03	.052
(Required ED care, could have been prevented)		(0.31)	(0.02)	
Primary care treatable	24592	0.28	0.19	<.001
(Did not require ED care)		(0.86)	(0.05)	
Did not require immediate care				
Nonemergent	24597	0.16	0.09	.007
		(0.61)	(0.03)	
Unclassified				
Unclassified	24602	0.15	0.06	.124
		(0.61)	(0.04)	

Note. This table reports the estimated effect of Medicaid on emergency-department use over the study period (10 March 2008 to 30 September 2009) in the entire sample and in subpopulations based on pre-randomization emergency-department use. For each subpopulation, it reports the sample size, the control means of the dependent variable (with standard deviation for continuous outcomes in parentheses), the estimated effect of Medicaid coverage (with standard error in parentheses), and the P value of the estimated effect. Sample consists of individuals in Portland-area postal codes after dropping those containing null value in the outcome variables (N = 24,605) or specified subpopulation (N in table). For the percent-with-any-visits measures, the estimated effects of Medicaid coverage are shown as percentage points. Visits were on hours if they occurred from 7 a.m. to 8 p.m. Monday through Friday and off hours otherwise.

As shown in Table 2, Medicaid increased overall emergency-department use. The above results are consistent with Taubman et al (2014). In the control group, 30.16% of individuals had an emergency-department visit during the 18-month study period. Medicaid increased the probability of having a visit by 9.07 percentage points (SE = 2.45; P < 0.001). Likewise, Medicaid coverage increased the number of emergency-department visits by 0.37 visits (SE = 0.12; P = 0.002) compared to the control mean of 0.81 visits.

Table 2 also illustrates the effect of Medicaid on emergency-department use for subgroups with no visits, one visit, two or more visits, five or more visits, and two or more outpatient visits in pre-randomization. In all groups, Medicaid increased ED use despite that some of the results are not statistically significant.

Visits were further subgrouped by hospital admission and timing of visit. By hospital admission, Medicaid coverage increased the percentage of people having outpatient visits by 10.25 (SE = 2.41; P < 0.001) and the number of visiting by 0.4 (SE = 0.11; P < 0.001). There was no statistical significance found regarding the effect of Medicaid on ED inpatient visits. The increase in emergency-department use from Medicaid was thus solely in outpatient visits.

By timing of visit, I separated visits into on-hours visits (7 a.m. to 8 p.m. Monday through Friday) and off-hours visits (nights or weekends). Medicaid significantly increased both type of visits, and it increased off-hour visits relatively more in both percentage and actual terms.

In addition, Medicaid coverage significantly increased the percentage of people having any ED visit to a low uninsured volume hospital in the study period by 4.09% (SE = 2.01; P = 0.04) and that of any ED visit to a high uninsured volume hospital by 7.59% (SE = 2.20; P < 0.001).

Replicating Taubman et al (2014), this analysis also classified visits by the primary diagnosis code for the visit into five types—emergent & not preventable, emergent & preventable, primary care treatable, nonemergent, and unclassified. As shown in Table 2., Medicaid increased visits in all classifications despite statistical insignificance for some. The increases were most notable in people classified as primary care treatable (0.19 visits; SE = 0.05; P < 0.001) and nonemergent (0.09 visits; SE = 0.03; P = 0.007).

Following the original study, this analysis further examined how Medicaid changed the ED visits for various medical conditions including injuries, headaches, backaches, depression, etc. Albeit insignificance, the vast majority of the causal effect were positive. Notably, Medicaid significantly increased emergency-department use for headache by 1.63% (SE = 0.007, P = 0.028) and injury by 5.56% (SE = 0.0184, P = 0.003).

Conclusion & Discussion

This analysis was aimed at replicating the results of Taubman et al (2014) regarding the causal effect of Medicaid coverage on emergency-department use via the instrumental variable lottery selection, conditioned on the number of people in the household. After conducting a series of balance check, weak instrument test, and 2SLS model fitting, this analysis reached the same conclusion as the original study, which was that Medicaid coverage increased emergency-department use both in general and in specific conditions. Although the estimated results were not the same as the original study, the difference was slight and reasonable, probably due to nuances in feature selection and data-preprocessing.

To further improve the analysis, the Difference-in-Difference tactic could be incorporated to eliminate the potential influence of omitted variables bias on causal inference, such as the long-term effect (i.e., changes in omitted variables overtime might confound the results).