Ha Nguyen & Yizhou Li

Professor Kebede

ECON375E

Final Research Paper: Idea Submission

### 

### **Empirical Research Question**

Our main research question is: How does health insurance coverage affect healthcare services utilization?

More particularly, we want to explore the impact of publicly subsidized health insurance (such as Medicaid) on the utilization of healthcare services, such as outpatient physician visits, hospitalization, or spending on healthcare, among non-elderly, low-income adults in the US.

Based on our real-life experience and the literature we have read so far, our initial hypothesis is that higher/wider health insurance coverage leads to an increase in healthcare service utilization. This implies the effectiveness of subsidized healthcare insurance payers such as Medicaid for the underserved, low-income population.

Our topic is still open to more definitions of the dependent variable.

### **Feasibility and Data Availability**

Here is the available health insurance data we've collected, which includes both **CPS** and **ASEC** data, combined into the **IPUMS CPS**. Additional datasets, such as NHIS, will also be included in the Initial **References for the Literature Review section**, as they are also referenced in relevant healthcare utilization research.

#### 

#### **Data Summary:**

**Health Insurance: Tables 2018-forward (All People)(**[Source](https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-hi/hi.2022.html#list-tab-1157843587))

This dataset offers valuable information on health insurance coverage over time, allowing for an examination of trends in healthcare utilization post-2018. It provides detailed breakdowns by demographic and socioeconomic groups. Variables include:

* **Family status**
* **Race and Hispanic origin**
* **Age** (12-65)
* **Sex**
* **Nativity**
* **Region**
* **Household income**
* **Work experience**
* **Disability status**
* **Education**
* **Firm size**
* **Income-to-poverty ratio**
* **Health status**
* **Marital status**
* **Residence**

Each category is accompanied by variables such as the **count of individuals**, **margins of error**, and **percentages of the total population** covered by health insurance. Health insurance types are divided into subcategories:

* **Any health insurance**
* **Private health insurance** (including employer-based, direct-purchase, and TRICARE)
* **Public health insurance** (including Medicare, Medicaid, and VA or CHAMPVA)
* **Uninsured**

*These datasets are already summarized and do not provide individual or case-by-case observations.*

#### **Raw Datasets:**

**IPUMS CPS (Current Population Survey)**

“Among these supplemental surveys, the Annual Social and Economic Supplement (ASEC) is the most widely used by social scientists and policymakers. The ASEC, along with basic monthly data from the CPS, provides the data for IPUMS-CPS.”([Source](https://cps.ipums.org/cps-action/faq#:~:text=Among%20these%20supplemental%20surveys%2C%20the,the%20data%20for%20IPUMS%2DCPS.))

**Current Population Survey 2024 Annual Social and Economic (ASEC) Data**:

This is the source dataset for the Health Insurance: Tables 2018-forward (all people) data summary. ([Source](https://www.census.gov/data/datasets/time-series/demo/cps/cps-asec.html))

**More CPS data** is available [here](https://cps.ipums.org/cps/index.shtml), healthcare related variables include:

* **PRVTCOUTLY**: Private insurance coverage through someone outside the household last year
* **PRVTCOVNW**: Currently covered by private insurance
* **PRVTDEPNW**: Currently have private health insurance through a household member
* **PRVTOWNNW**: Policyholder for current private insurance
* **PRVTCOUTNW**: Current private coverage provided by a person outside the household
* **GRPCOVLY**: Covered by employment-based group health last year

And many more, mostly ranging from 2019~2024.

These are all valuable pieces of information that can help assess healthcare insurance and identify the types of coverage available.

**IPMUS Health Survey NHIS dataset:**

The National Health Interview Survey (NHIS) is an annual cross-sectional survey that collects data on the health status, healthcare access, and health behaviors of the civilian population, with data available from 1963 to the present. It includes healthcare-related variables such as:

* **Demographic Information**
* **COVID-19 Related Variables(Covid may have affected healthcare utilization as well， perhaps positive effect)**
  + **CVDTEST**: Ever been tested for COVID-19
  + **COVIDEVR**: Ever told they had or likely had COVID-19
* **Health Behaviors(Where Card et al (2004) researched)**
  + **ALCANYNOE**: Frequency of alcohol consumption in the past year
* **Healthcare Information accessibility(May also may be an factor influencing health care utilization)**
  + **PCLOOKHELYR**: Searched for health information online in the past 12 months
  + **PCACS**: Has internet access
  + **MCAIDPREM**: Enrollment fee or premium for Medicaid plan

In the **Card et al (2004)** research paper, two measures of healthcare utilization are whether the individual had at least one doctor visit in the past year and whether they had one or more overnight hospital stays during the same period, all of which are available in NHIS data(perhaps not the exact same variables):

* **DVINT**: Interval since the last doctor visit (available from 1963, except for 1965 and 1966)
* **HOSPNITE**: Number of nights spent in the hospital in the past 12 months (available from 1999-2018)

These are all candidate variables that could be potentially used for our health utilization measurements

Other specific metrics that measure how insurance coverage may assist with the utilization of particular (surgeries) or general healthcare services that benefit from insurance coverage, include:

* **MCPRIHORS**: Medicare and/or private insurance coverage for hospital and/or surgical services
* **MCPRIHS**: Medicare and/or private insurance coverage for both hospital and surgical services
* **MCSUMM**: Summary of Medicare coverage for hospital and/or doctor services

# **Initial References for Literature Review:**

1. **Card et al (2004): The Impact Of Nearly Universal Insurance Coverage On Healthcare Utilization And Health: Evidence From Medicare**

**Key findings:**

1. **Improved Access to Medical Care**: Medicare eligibility at age 65 significantly reduces disparities in access to medical care, especially for groups with large increases in insurance coverage. These groups are less likely to delay or forgo care and are more likely to have regular doctor visits and checkups.
2. **Increased Hospitalizations**: Medicare eligibility leads to a 10% increase in hospital stays, primarily for non-emergency and elective procedures like joint replacements. Interestingly, this increase is more pronounced among whites and people in areas with higher pre-65 insurance coverage rates.
3. **No Effect on Health Behaviors**: Medicare eligibility does not have a noticeable impact on smoking, exercise, or obesity rates.
4. **Slight Improvement in Self-Reported Health**: There are small but noticeable improvements in self-reported health, especially among groups that saw the largest increases in insurance coverage after Medicare eligibility.
5. **No Immediate Impact on Mortality**: There is no immediate effect of Medicare eligibility on mortality rates at age 65, nor any clear change in the growth rate of mortality after age 65.

**DataSet:** Data from NHIS (https://nhis.ipums.org/nhis/) and the Behavioral Risk Factor Surveillance System (BRFSS) dataset

**Empirical Approach:**

**Using a two-stage least square estimation method**

(1)

Where：

is the health outcome for individual i in group j at age a.

​ represents measured characteristics with group-specific coefficients .

is a smooth function representing the age profile of the outcome, such as a low order polynomial

is a dummy variable indicating whether an individual has insurance coverage.

is an unobserved error component.

**Endogeneity** arises as may be correlated with the error term . To address this, 2SLS is employed, with the following first-stage regression:

(2)

Where is the instrumental variable for insurance coverage. By substituting ​ from equation (2) into equation (1), we get:

⇒ (3)

Where:

By substituting each variable into equation (3) the **reduced form is** :

(4)

Given Hence we have population instrumental variable estimate /

**Age as a Confounder**: Age may influence both insurance coverage and health outcomes, making it a potential confounder in the model.

1. **Ghosh et al (2019): The Effect of Health Insurance on Prescription Drug Use Among Low-Income Adults: Evidence from Recent Medicaid Expansions**

**Key Findings:**

1. **Prescription Utilization Increase:** The study shows that ACA Medicaid expansions led to a 19% rise in Medicaid-paid prescriptions among low-income, non-elderly adults within the first 15 months since new health insurance became available.
2. **Chronic Condition Medications:** The largest increases were observed for medications treating chronic conditions such as diabetes and cardiovascular diseases. Demand for diabetes-related medications rose by 24%, birth control by 22%, and cardiovascular medications by 21%.
3. **Generic vs. Brand-name Drugs:** The increase in prescription utilization was more pronounced for generic drugs than brand-name drugs, suggesting a shift towards lower-cost medications.
4. **Price Sensitivity:** Medicaid expansions revealed a high price sensitivity among low-income populations. States with higher Medicaid drug copayments saw a smaller utilization increase, meaning cost-sharing might impact medication utilization.

**Primary** **Dataset:** A national, all-payer (regardless of payer type) pharmacy transactions database, aggregated by units of geography (both state level and Core Based Statistical Areas–CBSA). To control for the unemployment rate, the authors merge this dataset with CBSA-level unemployment rates from the Bureau of Labor Statistics (BLS) Local Area Unemployment Statistics (LAUS). Additionally, they include county-level 2013 uninsured rates for non-elderly adults at or below 138 percent FPL from the Census Bureau’s Small Area Health Insurance Estimates (SAHIE) program, and racial/ethnic composition data from the Area Health Resource Files (AHRF).

**Empirical Model:**

1. **Difference-in-Difference**: The primary econometric approach employed was a difference-in-difference model. This model compared changes in prescription drug use between states that expanded Medicaid and those that did not, before and after the expansions.
2. The OLS regression:

Yst = α + βPostt × Expansions + ∂UEst + τt + θs + ϵst

Yst​: logged total prescriptions in the state, with s indexing state and t indexing each quarter in the data

β: measures the change in Medicaid scripts in expansion states net of the change in non-expansion states.

Postt​: a binary variable indicating the post-expansion period

Expansions​: a binary variable indicating whether the state expanded Medicaid

→ Postt × Expansions: whether a state has expanded Medicaid at time t

UEst: state quarterly unemployment rate, control for change in economic climate

θs: state fixed effects to account for time-invariant state-specific differences in prescription drug use and

τt: time dummies for each quarter

ϵst​: the error term.

1. **Additional Analyses**: The study also conducted event studies and placebo tests to verify the robustness of the results. They examined heterogeneous effects across therapeutic classes and drug types (generic vs. branded).
2. **O'Connor (2015): The Impact Of Insurance Coverage On Consumer Utilization Of Health Services: An Exploratory Study**

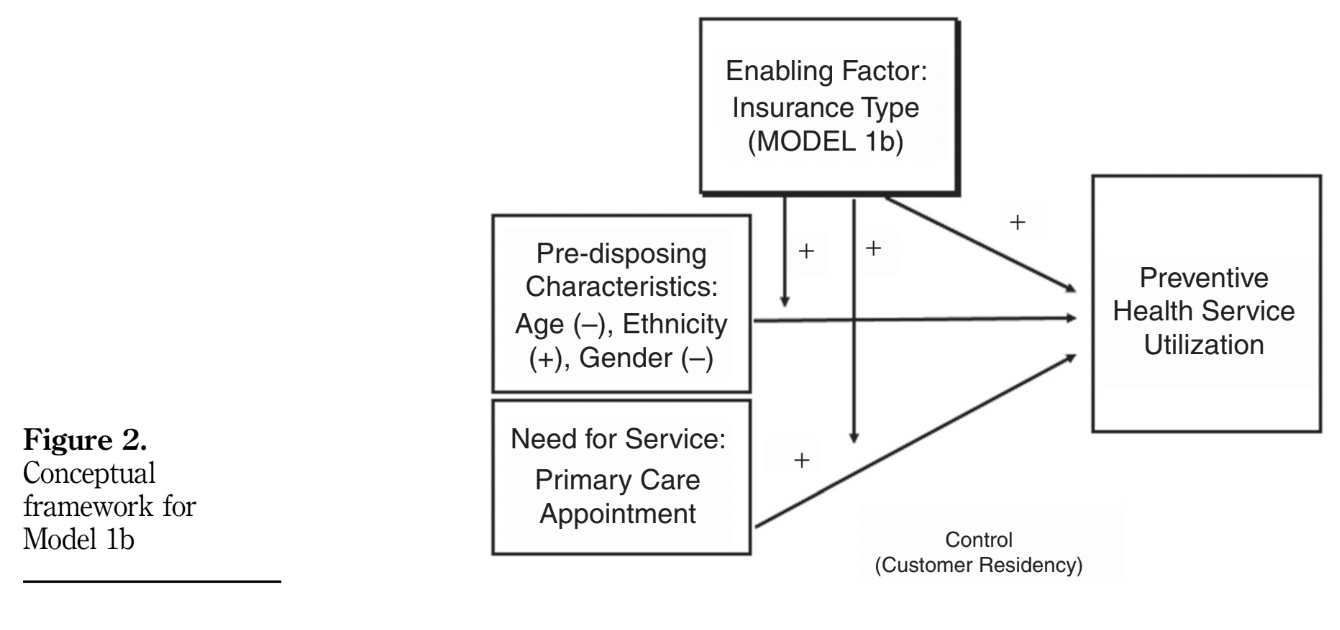
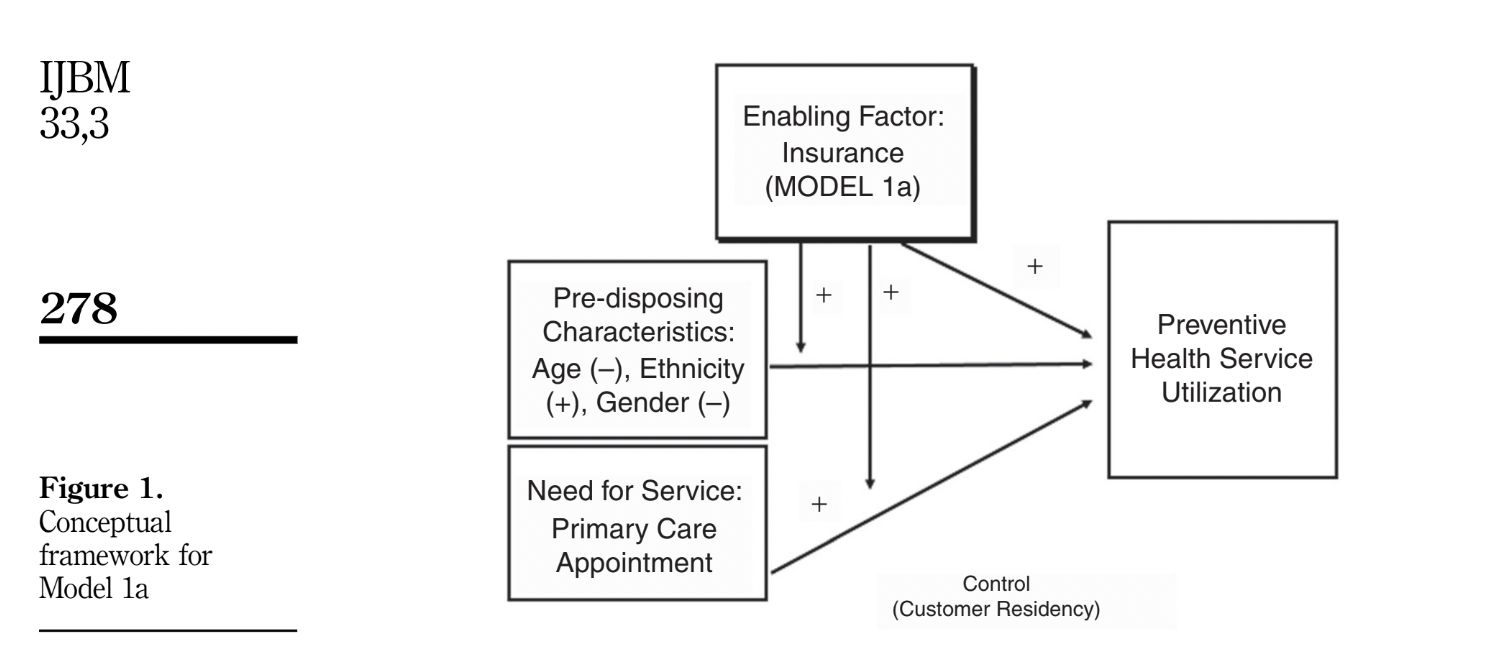
**Key Findings:**

1. **Direct Impact of Insurance:** The study finds that health insurance coverage significantly increases the likelihood of utilizing health services. Insured individuals are more likely to schedule preventive care appointments than uninsured individuals.
2. **Demographic Factors:** The impact of insurance coverage varies by factors such as age, gender, and ethnicity. E.g: younger, female individuals are more likely to utilize preventive services.
3. **Type of Insurance:** Different types of insurance (private vs. government-sponsored) also influence healthcare utilization. Private insurance enrollees are more likely to use preventive services than those with government insurance.

**Dataset:** The author uses primary data source of proprietary data from a major metropolitan healthcare system in the northeastern US, detailing information on ~500,000 urban and suburban patients’ outpatient appointments over one year.

**Empirical Model:** The author utilizes 2 logistic regression models. A logistic regression model estimates how explanatory variables affect the likelihood of scheduling a preventive health service appointment. The analysis is based on individual customer appointments, with logistic regression chosen because the outcome, preventive health service utilization, is binary.

1. Conceptual Frameworks:



1. General regression model:

*E{Yi} = exp(X’1)/1 + exp(X’1)*

Y: preventive health service utilization (acute care appointments = 0, preventive care appointments = 1)

X: vector of independent variable (both enabling factor and pre-disposing characteristics)

E: likelihood of choosing one of the response outcomes (preventive care appointment or acute care appointment)

*:* represent the change in log-odds of selecting preventive care for a one-unit increase in the independent variables.

*i:* number of appointments

1. **Other literature:**
2. Hahn (1994) examines and discusses that extending private insurance to the uninsured significantly increased their healthcare utilization while shifting Medicaid recipients to private insurance decreased their utilization. The magnitude of these changes varied by type of care, with notable differences in physician visits and hospitalizations. The study emphasized the distinct healthcare needs and utilization patterns of the uninsured and Medicaid populations. The research used data from the 1987 National Medical Expenditure Survey (NMES).
3. Freeman et al (2008) is a systematic review of studies about the impact of healthcare insurance on access to healthcare, utilization, and quality of health. This paper gives us a detailed look into more sources that we could use as references for the final literature review.

Bibliography

# 

Card, David, Carlos Dobkin, and Nicole Maestas. 2008. “The Impact of Nearly Universal Insurance Coverage on Health Care Utilization: Evidence from Medicare.” *American Economic Review* 98 (5): 2242–58. https://doi.org/10.1257/aer.98.5.2242.

Freeman, Joseph D., Srikanth Kadiyala, Janice F. Bell, and Diane P. Martin. 2008. “The Causal Effect of Health Insurance on Utilization and Outcomes in Adults.” *Medical Care* 46 (10): 1023–32. https://doi.org/10.1097/mlr.0b013e318185c913.

Ghosh, Ausmita, Kosali Simon, and Benjamin D. Sommers. 2019. “The Effect of Health Insurance on Prescription Drug Use among Low-Income Adults:Evidence from Recent Medicaid Expansions.” *Journal of Health Economics* 63 (January): 64–80. https://doi.org/10.1016/j.jhealeco.2018.11.002.

Hahn, Beth. 1994. “Health Care Utilization: The Effect of Extending Insurance to Adults on Medicaid or Uninsured.” *Medical Care* 32 (3): 227–39. https://doi.org/10.1097/00005650-199403000-00004.

O’Connor, Genevieve Elizabeth. 2015. “The Impact of Insurance Coverage on Consumer Utilization of Health Services.” *International Journal of Bank Marketing* 33 (3): 276–97. https://doi.org/10.1108/ijbm-05-2014-0061.