

Research Design

Katelyn Diaz, Sarah Fleming, Shevaughn Holness

4/1/2022

Background

Despite the United States spending 19% of its GDP on healthcare in 2020 – the most of any other advanced country — health insurance rates have been declining unevenly since 2017 (Medicare and Medicaid Services n.d.). Health insurance not only serves an important role in reducing medical debt but it has also been linked to a significant decrease in mortality (Card, Dobkin, and Maestas 2007). Despite the overall benefits to health, nearly 8.6% of the US population did not have health insurance in 2020 (US Census Bureau n.d.d).

While health insurance has undergone a multitude of good reforms over the past decade, not all demographics have felt the reforms evenly. Infact, recent reforms have been detrimental for some. With over half of the population getting employer-provided health insurance, people who face obstacles to employment, such as racial minorities, the disabled, and undocumented immigrants, are historically shown to have higher rates of being uninsured (Damico 2021). With the passing of the Affordable Care Act in 2010, there was significant expansions in terms of healthcare coverage, with racial minorities seeing a significant increase in the insured population. However, later policy changes made by the Trump administration reversed these trends, with most minorities having an increase in the uninsured rates (Damico 2021).

Citizenship status also seems to be a crucial factor in determining uninsured rates. Michael S. Cohen’s and William L. Schpero (2018) found that mixed-citizenship status households were less likely to have health insurance due to the fear that it would warrant further inspection into the legality of household members (Cohen and Schpero 2018).

In light of these shifts occurring in uninsured rates, we aim to investigate possible demographic determinants of county level uninsured rates. The two county level demographic determinants of interest for uninsured rates are county median household income and percentage of non US citizen residents. Median income is a significant factor regardless of employment status because higher-income households can afford private insurance. Undocumented people often barriers to employment, meaning that they may be less likely get a job with benefits, and thus, get health insurance from their employers.

We suspect that as the percentage of non-citizens residents increase, the percent of the county’s population who are uninsured will also increase. We also suspect that as median income increases the percent of the county’s population who are uninsured will decrease.

Methods

Data

The dataset we will be using comes from the 2019 American Community Survey (ACS) collected by the United States Census Bureau; we chose 2019 as the year of interest as it is the most up-to-date and validated ACS (US Census Bureau n.d.a), (US Census Bureau n.d.c). The ACS collects demographic, housing, and economic information on the county level for the United States, Puerto Rico, and Guam. The U.S Census Bureau collected this information through the use of mailed questionnaires, telephone interviews, and visits

from Census Bureau field representatives. This data is collected from 3.5 million household addresses annually and only in counties with at least 65,000 people(US Census Bureau n.d.b).

The unit of analysis for this dataset is county, and there are eight hundred fifty-one total observations. From this dataset, we chose to focus on counties that are in the United States which leads to our dataset having eight hundred twenty-one observations. As we can only include a portion of the counties in the U.S., our dataset is a sample as the population of interest is all counties in the U.S.

Variables

The analysis conducted includes the following variables: median household income, disability status, marriage status, employment status, US citizenship status, uninsured, and majority racial group. The response variable was the percentage of people who are uninsured, which is a continuous variable. The key variables analyzed were the continuous variables median income per county in dollars, and percentage of people who are citizens of the US.

The covariates included the continuous variables percentage of people with a disability, percentage of people married, and percentage of people employed, along with the reference binary variable of if the county is majority white. For our analysis, race was mutated into a categorical variable consisting that stated which racial group held the majority in that county. The reference group is `majority_white` (which indicates if the county is majority white). The other groups in the model are `majority_black`, `majority_hispanic`, and `majority_other`.

Analysis

Null Hypothesis1: The percentage of Non-US citizens in a county has no effect on the amount of people who are uninsured for a county. **Alternative Hypothesis1 :** The percentage of Non-US citizens in a county has an effect on the amount of people who are uninsured for a county.

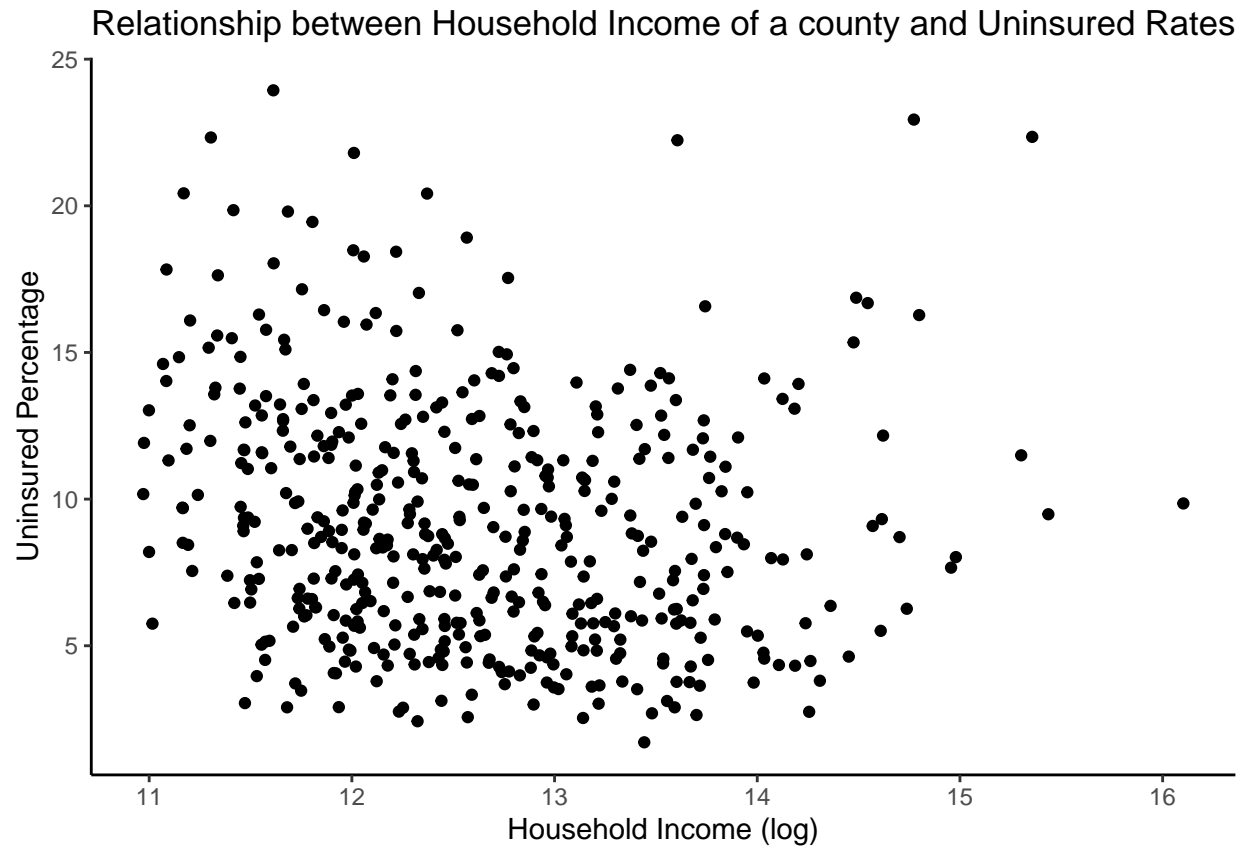
Null Hypothesis2: The median household income in a county has no effect on the amount of people who are uninsured for a county.

Alternative Hypothesis2: The median household income in a county has an effect on the amount of people who are uninsured for a county.

Our primary hypothesis is that the percentage of Non US-Citizens inside a county as well as the median household income of a county have an effect on the percentage of uninsured people in that county. As such, we want to start with simple linear regression to show the relationship between each of those variables independently before moving to multiple regression to view their relationship with uninsured rates when controlling for other variables. Two of our models will be simple linear regression models that have either percentage of non US-citizens inside the county or median household income of the county as an explanatory variable for percentage of people who are uninsured. This would show if there was a relationship between the explanatory and the response variable when not controlling for other factors.

We would then create numerous multiple regression models. The third model we create will have only the percentage of Non US-Citizens inside a county and the median household income as explanatory variables. We will include a full model that has all the covariates as well as the independent variables as a focus. The last few models will be the model produced after accounting for any potential multicollinearity.

All models will be shown; the model that we will choose to discuss is the model that is indicated to be best fitted model by its R-Squared Adjusted, AIC, and BIC. We will check regression diagnostics for each model before running them. ### Results



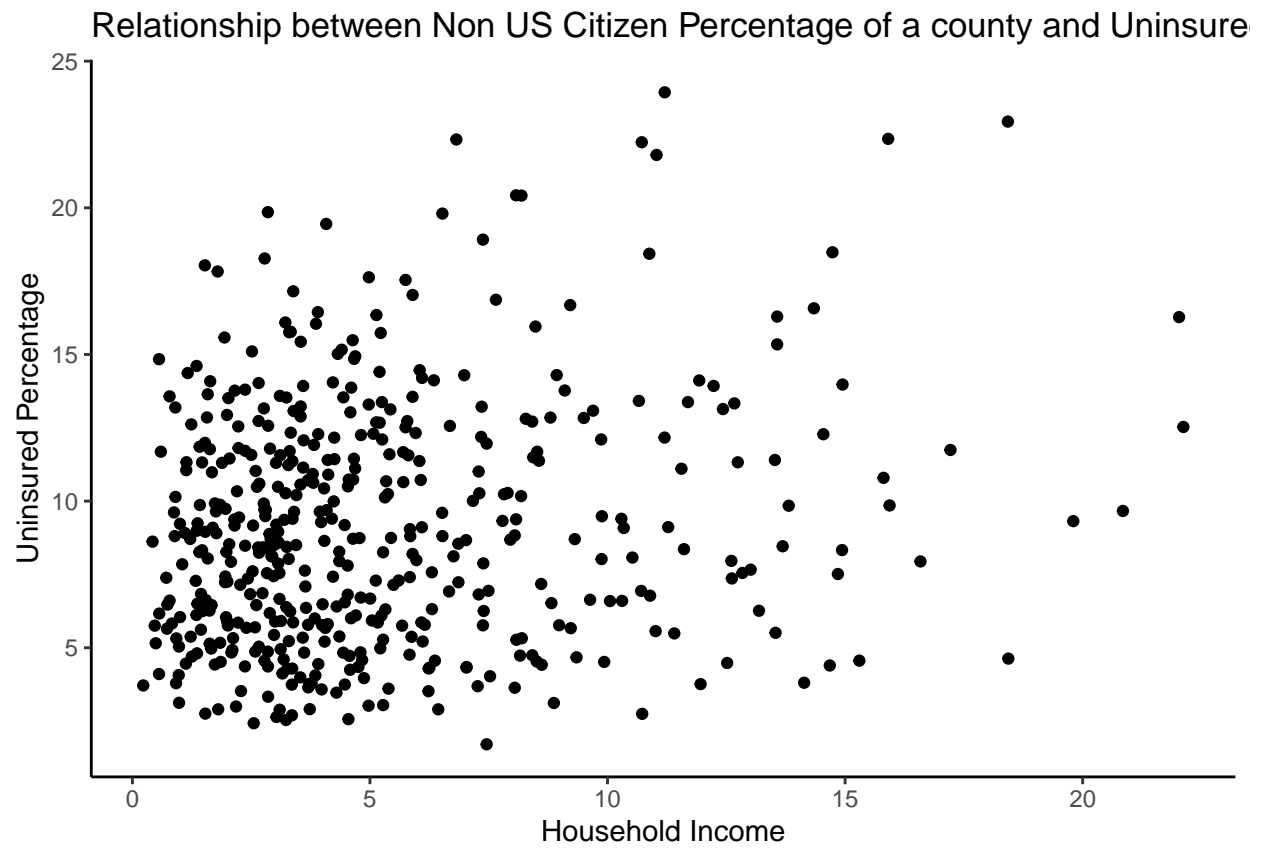


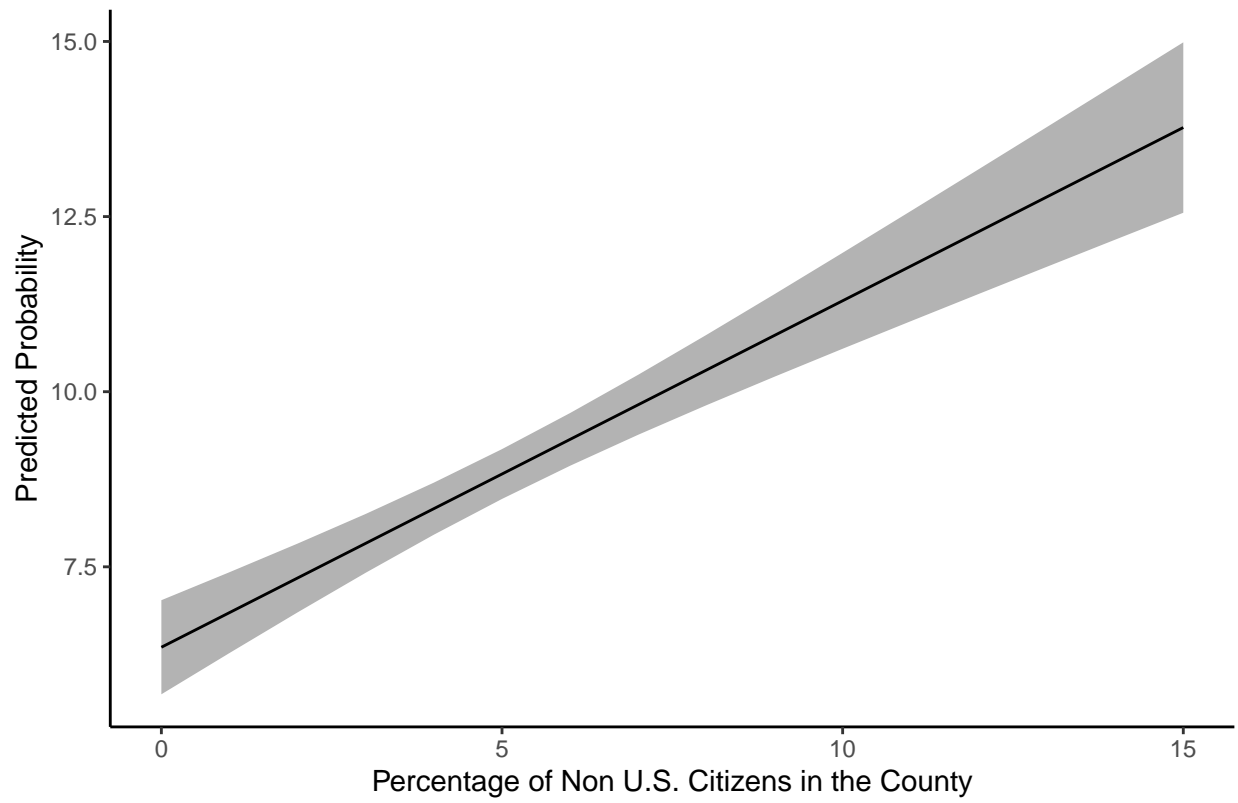
Table 1: Predicting County Uninsured Rates

	<i>Dependent variable:</i>					
	uninsured_pct					
	(1)	(2)	(3)	(4)	(5)	(6)
Citizenship Percentage	0.191*** (0.048)		0.423*** (0.056)		0.346*** (0.056)	0.495*** (0.059)
Household income (log)		-0.711*** (0.209)	-1.752*** (0.240)	-0.627*** (0.228)		-1.429*** (0.233)
Employment Percentage				-0.266*** (0.078)	-0.307*** (0.076)	-0.272*** (0.073)
Disabled Percentage				-0.012 (0.082)	0.270*** (0.083)	0.197** (0.081)
Marriage Percentage				-0.049 (0.030)	-0.038 (0.029)	-0.049* (0.028)
Racial Majority: White				-3.125*** (0.623)	-1.716*** (0.627)	-1.748*** (0.604)
Constant	8.147*** (0.316)	18.119*** (2.637)	29.015*** (2.875)	38.653*** (7.085)	25.804*** (6.270)	42.564*** (6.626)
Adjusted R ²	0.030	0.022	0.127	0.112	0.166	0.227
Akaike Inf. Crit.	2,676.759	2,680.671	2,627.998	2,639.074	2,609.245	2,574.523
Bayesian Inf. Crit.	2,689.230	2,693.141	2,644.626	2,668.173	2,638.344	2,607.779

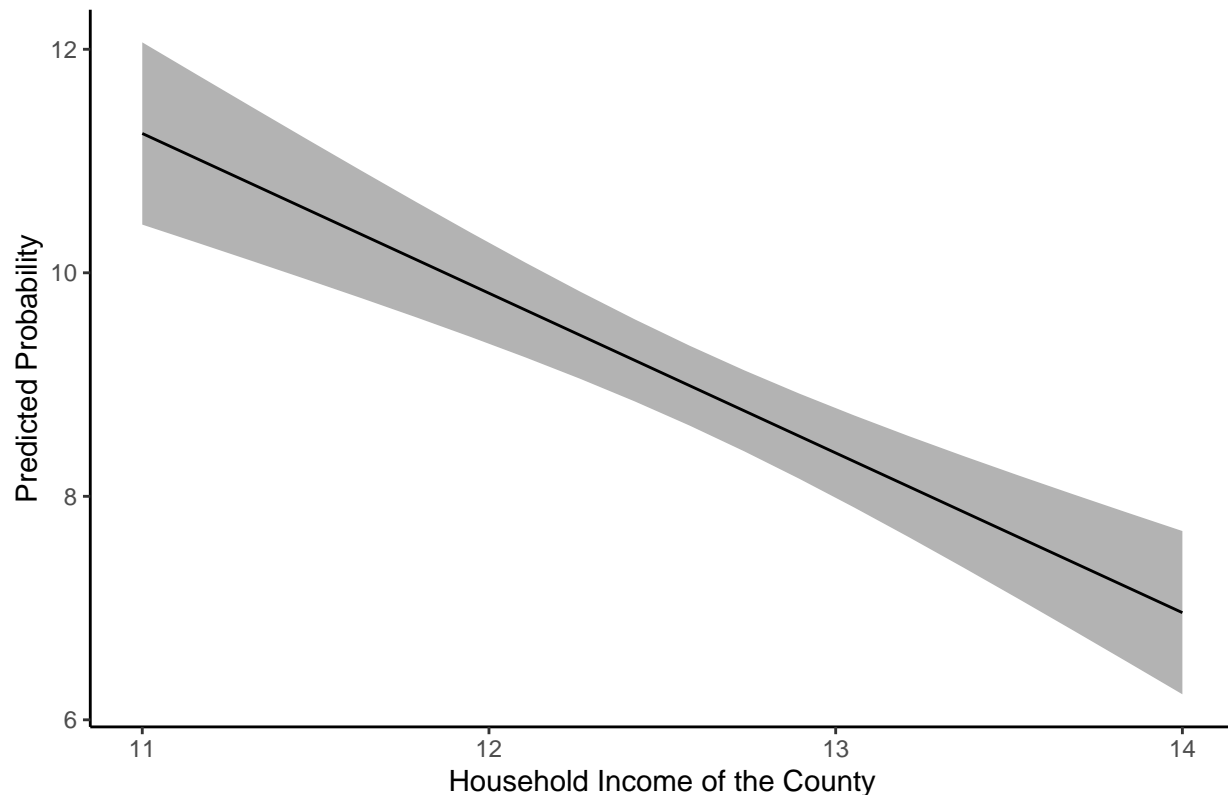
Note:

*p<0.1; **p<0.05; ***p<0.01

Relationship between U.S. Citizenship Percentage of a county and Uninsu



Relationship between household of a county and Uninsured Rates



Appendix

Histograms of variables of interest

Variance Inflation Factor tests

Testing models for assumptions

Bibliography

- Card, David, Carlos Dobkin, and Nicole Maestas. 2007. "Does Medicare Save Lives?" Working {Paper} 13668. National Bureau of Economic Research. <https://doi.org/10.3386/w13668>.
- Cohen, Michael S., and William L. Schpero. 2018. "Household Immigration Status Had Differential Impact On Medicaid Enrollment In Expansion And Nonexpansion States." *Health Affairs* 37 (3): 394–402. <https://doi.org/10.1377/hlthaff.2017.0978>.
- Damico, et. al, Anthony. 2021. "Health Coverage by Race and Ethnicity, 2010-2019." *KFF*. <https://www.kff.org/racial-equity-and-health-policy/issue-brief/health-coverage-by-race-and-ethnicity/>.
- Medicare, Center for, and Medicaid Services. n.d. "NHE Fact Sheet CMS." Accessed April 1, 2022. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet>.
- US Census Bureau. n.d.a. "2020 Data Release Schedule." *Census.gov*. Accessed April 1, 2022. <https://www.census.gov/programs-surveys/acs/news/data-releases/2020/release-schedule.html>.

- . n.d.b. “American Community Survey (ACS) Questions and Answers.” Accessed April 1, 2022. <https://www.bls.gov/lau/acsqa.htm>.
- . n.d.c. “Explore Census Data.” Accessed April 1, 2022. <https://data.census.gov/cedsci/table?q=United%20States&t=Health%20Insurance&g=0100000US%240500000&tid=ACST5Y2019.S2701>.
- . n.d.d. “Health Insurance Coverage in the United States: 2020.” *Census.gov*. Accessed April 1, 2022. <https://www.census.gov/library/publications/2021/demo/p60-274.html>.