Household Income and Chronic Disease Diagnoses: a County Level Analysis

Srija Somaka

Econ 3161: Dr. Shatakshee Dhongde

Abstract:

Increasing chronic disease burden is one of the most pressing problems the United States faces today despite maintaining the highest levels of health expenditure globally. This paper attempts to discern the relationship between county level diabetes diagnosis rates and median household income using regression analysis. A variety of other social determinants of health are considered in this study in order to account for factors that can impact chronic disease outcomes: density of Doctors of Medicine, insurance coverage, age, education levels, poverty levels, unemployment rates, and metropolitan classifications. Ultimately a negative association between income levels and diabetes diagnoses was found; as the median household income of a county increased, the percentage of the population diagnosed with diabetes decreased.

I. Introduction

Despite having one of the highest expenditure rates on preventative healthcare measures globally, the United States consistently ranks in the highest percentile for chronic disease rates and number of avoidable hospitalizations from preventative causes. This shocking juxtaposition of health expectations and realities in America can generally be boiled down to one main concern: access to care. Increasing chronic disease prevalence proves to be one of the largest strains on the American health system today, making access to healthcare an increasingly relevant topic in evaluating the efficiency and effectiveness of the U.S. health systems in place. A variety of uncontrollable factors play a significant role in determining an individual's life expectancy, treatment outcomes, and overall quality of life. These disparities are increasingly prevalent in the large and growing gap between rural and metropolitan regions where healthcare inequality is stark.

The factors that feed the inequity of health access are Social Determinants of Health; these are the breakdown of the invisible social, economic, and geographic factors that silently manipulate the outcomes of an individual or community's overall health and quality of life (CDC). The scope and intensity of these determinants can vary from region to region, but it is increasingly clear how their pervasiveness can impact rural America at a greater scale than its urban counterpart.

Although there are a variety of determinants available for study, this paper will focus on the economic factors that impact health outcomes and widen the health equity gap. In particular, this paper will analyze the effect that household income levels have on the diagnosed diabetes prevalence. Diabetes prevalence was chosen as the dependent variable in this study to reflect one of the growing chronic disease burdens the United States faces and to better understand what may contribute to it.

This paper will test the hypothesis that lower median household incomes corresponds to a higher diagnosed diabetes rate. There are two frameworks that support this notion. At an individual level, a lower income will correspond to less money available to spend on healthcare. A stricter healthcare budget would lead to delayed care seeking and thus the exacerbation of existing treatable disease. At an institutional level, lower income would result in lower income taxes which is correlated to lower health infrastructure expenditure. With less money funneled towards hospitals and ambulance services, there would be less hospitals and doctors, leading to lower access to care and this higher chronic disease incidence. With these two frameworks as motivation,

this paper aims to identify and explain the relationship between income levels and diabetes prevalence. A variety of other social determinants will also be included in this assessment in order to paint a fuller and less biased image that better reflects the true relationship between the independent variable of interest and the associated county level health outcomes.

II. Literature Review

The lack of equitable access to care is a prevailing issue in rural health. In their paper, Baffour (2017) outlines a variety of social determinants and explains how these factors can impact health outcomes. Although not direct economic analysis, the paper provides a valuable insight into the evaluation of social determinants of health and provides a logical framework for how they impact a community's access to care. These findings helped to shape and direct the motivations and studies of this paper. In their paper, it is clear how the challenges posed by social determinants is particularly high in rural communities, particularly in ethnic and minority groups in those communities. There are a few factors of interest; the first one evaluated is insurance coverage. Lower insurance coverage leads to increased cost of care and higher wait times for patients, both directly decreasing an individual's access to care and resultant health outcomes. Other economic factors such as poverty and unemployment rates are also analyzed. With higher poverty and unemployment rates, less money is spent on health infrastructure, leading to a shortage of doctors and treatment options. Factors such as education and health literacy levels decrease the likelihood of people seeking out or receiving preventative care, which ultimately leads to more expensive treatment once it is finally sought out. A combination of all these factors is present in each community, but the rural community faces a dire shortage of factors that positively impact access to care. As a result of this, rural communities are more likely to bear the burden of chronic diseases than urban communities are.

Streeter, Snyder, Kelly, Stahl, Li and Washko (2020) further study the impact of social determinants on health access by juxtaposing markers for social determinants with Health Professional Shortage Areas to analyze trends. This was achieved by using aggregated census data as well as the Health Resources and Services Administration's cross sectional Area Health Resource Files, which provide a measure of the different social determinants as well as a count of the health professionals at a county level. This provided them with a county level analysis of how the primary care Health Professional Shortage Areas (pcHPSAs) interact with a variety of other

determinants of health like economic and geographic factors that impact access to care in rural areas. Though this paper enumerated many at risk counties in the States, there were varying levels of risk, each associated with different levels of health professional shortage. The primary demographic markers used in this study were low birth weight rates, median household income, poverty, education, rural distinctions of counties. Moreover, although various social determinants of health were identified nearly ubiquitously in counties nationwide, the magnitude of social determinants prevalent made a difference in risk levels. The selected markers were used to analyze and create a vulnerability score and see how it aligned to the pcHPSA. The methodology is simple: the higher the marker count, the higher the vulnerability of the region. Ultimately, the paper concludes that there is a relationship between higher vulnerability and higher health professional shortages.

Overall life expectancy has increased over time, but not necessarily for people in rural areas who are impacted by some key health indicators. Singh, Daus, Allender, Ramey, Martin, Perry, Reyes, and Vedamuthu (2017) look into the socioeconomic and geographic disparities between poor rural areas and their comparatively richer counter parts. In order to study this, they use cross sectional data from the CDC's National Health Interview Survey and American Community Survey to investigate social determinants and geographic inequalities in order to determine a health differential between richer and poorer rural regions. The health differentials were calculated using life tables, prevalence, and risk ratios. Ultimately, the paper concludes that race had a large impact on the health outcomes of different populations. This can be tied to socioeconomic factors as well because of minority groups being more likely to live in areas with higher poverty rates, higher unemployment, and lower access to care. These differences lead to statistically significant differences in life expectancies at birth across race and gender.

Rather than providing a broad scope of general health outcomes, the focus of this paper will specifically be on a growing concern in U.S. healthcare: chronic disease. In particular, impact of different economic social determinants on the prevalence of diabetes will be analyzed. This paper will also use more granular county level data, as well as more recent data than the papers listed so far. The differences in models between urban and rural counties will also be considered.

III. Data

In order to investigate the hypothesized relationship between household income and diabetes prevalence, cross sectional county data gathered from the CDC Diabetes county indicators, Health Resources and Service Administration (HRSA) Area Health Resources, and the American Community Survey from the US census was used. This information was conveniently aggregated by the Rural Health Information Hub. RHI used the data in order to map these indicator measurements to each county. While it included algebraic manipulations to deduce ratios and percentages, the dataset remained clean and provided complete information for 3,140 counties. Some data cleaning was used to remove missing values and merge on county and state before use. The data used for analysis was collected in 2018. The table below provides a description of the data and where it was collected from.

Table I: Data Description

Variable Name	Description	Year	Units	Source
diabetes	Percentage of county population diagnosed with Diabetes	2018	Percentage	CDC Diabetes County Data Indicators
MDs10k	Ratio of Doctors of Medicine per 10,000 people, rates calculated against US census data	2018	Ratio	HRSA Area Health Resource Files
Centuninsur	percentage	2018	Percentage	US Census Small Area Health Insurance Estimates
medAge	Median age of the county	2018	Years	US Census Population and Housing Unit Estimates
popNoHSDip	Percentage of the county population without a high school diploma	2019	Percentage	US Census, ACS
medHouseIncome	Median Household Income	2018	Dollars	US Census Small Area and Poverty Estimates
logMedHouseInc	Log of previous median household income variable	2018	Log(dollars)	calculated
Poverty	County poverty rate	2018	Percentage	US Census Small Area and Poverty Estimates
Urate	County unemployment rate	2018	Percentage	USDA Economic Research Service
Metro	Metro or non-metro classification of the county based on population	2018	Dummy	National Center for Health Statistics

The main independent variable in this study is the median household income. This is motivated in part from a study in the literature review by Streeter et al. (2020) which shows a high correlation between median household incomes on general health outcomes. This paper aims to analyze the impact of this main variable on the dependent variable of diabetes prevalence. In analysis, the log of the income was used in order to linearize the model and better interpret the results of the regression. Diagnosed diabetes prevalence is calculated as the percentage of the population diagnosed with diabetes using the CDC dataset which provides the number of diagnoses per county and the census data which provides county population as a whole. The ratio of Doctors of Medicine per 10,000 people in each county is included in order to quantify the shortage of doctors in a region as this factor can play a large role in determining access to care. Calculations of the percentage of the population that is insured were included in this analysis due to insurance coverage often impacting an individual's ability to afford and seek out care. The median age variable is included in order to control for any relationship between age and health status. In order to account for education and its potential impact on an individual's propensity to seek care and maintain overall health levels, a variable that represents the percentage of the county that held a high school diploma was included. Poverty and unemployment rates were also included as economic indicators for each county. Lastly, a dummy variable was included to determine if there was a significant difference in urban and rural counties.

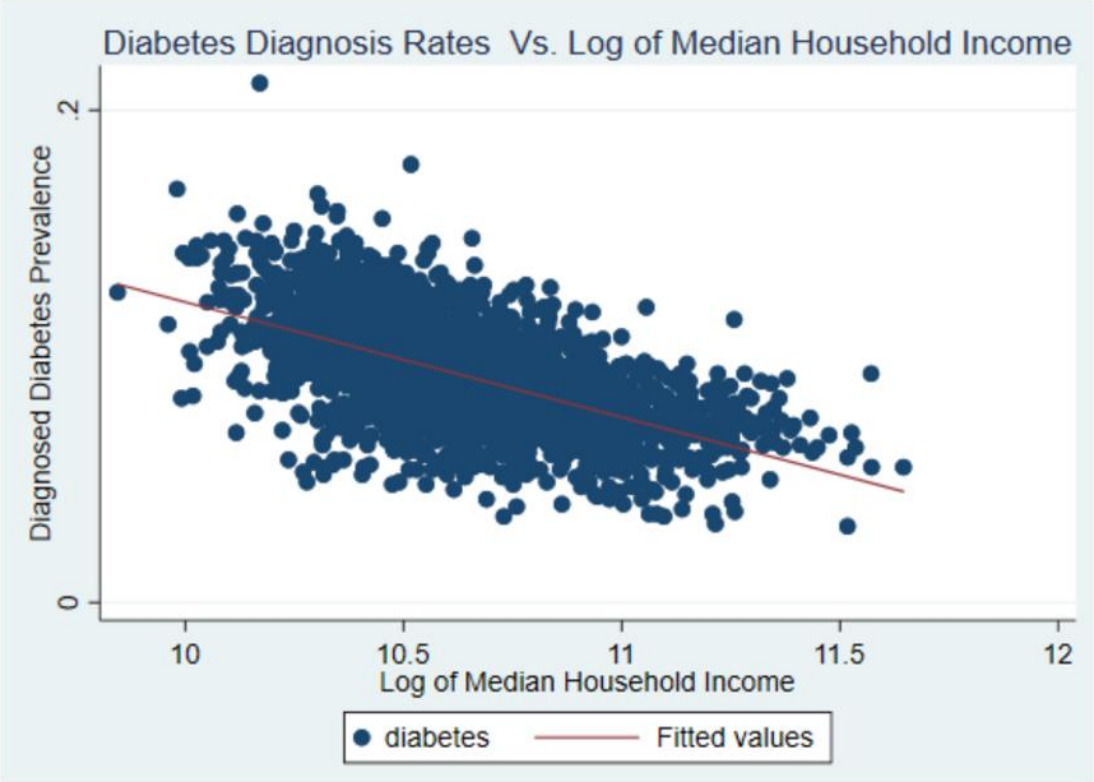
In order to better understand the impact of the main independent variable of interest, median household income, other Social Determinants of Health have been included in the dataset used for analysis. Since household income and the chronic health outcomes this paper aims to correlate do not exist in a vacuum, it is important to account for the other factors that play a part in determining long term quality of life.

Table 2: Summary Statistics

Variable	Obs	Mean	Standard Deviation	Min	Max
Diabetes	3140	0.09	0.02	0.031	0.21
MDs10k	3140	0.001	0.001	0	0.03
Centuninsur	3140	0.21	0.07	0	0.51
medAge	3140	40.4	5.08	22	62.9
popNoHSDip	3140	0.18	0.08	0.01	0.58
logMedHouseInc	3140	10.64	0.24	9.84	11.65
Poverty	3140	0.16	0.06	0.03	0.82
Urate	3140	0.05	0.02	0.01	0.20
metro	3140	0.38	0.48	0	1

Below is simple scatter plot of the log of median household income and diagnosed diabetes prevalence in order to get a quick visual idea of the data and what the trends will be. Here it shows a relatively strong negative relationship; as median household income of a county increases, diabetes prevalence is on a downward trend.

Figure 1: Scatterplot of Diabetes Diagnosis Rates vs. the Log of Median Household Income



Having presented the data it is important to ensure that the data meets the requirements of the Classical Linear Model so that we can fairly assess the validity of the regression results.

Assumption 1: The model is linear in parameters so that $y = B_0 + B_1x_1 + B_2x_2 + \ldots + B_kx_k + u$

This assumption is satisfied because the results in part IV are linear.

Assumption 2: Data selection was achieved via Random Sampling

All the data sources use some variation of county data and no counties were excluded in the data collection. There was no premeditated data selection technique implemented, therefore the data is proven to be random.

Assumption 3: Explanatory variables are not perfectly collinear

In order to prove this condition, a correlation table was calculated on STATA. As shown in the matrix below, there are no perfectly collinear variables.

Table 3: Correlation Matrix

	Diabetes	MDs10k	Centuninsur	medAge	popNoHSDip	logMedInc	Poverty	Urate	metro
Diabetes	1								
MDs10k	-0.19	1							
Centuninsur	0.26	-0.24	1						
medAge	0.13	-0.19	-0.09	1					
popNoHSDip	0.54	-0.25	0.61	-0.16	1				
logMedInc	-0.57	0.30	-0.48	-0.12	-0.61	1			
Poverty	0.54	-0.15	0.50	-0.19	0.68	-0.84	1		
Urate	0.34	-0.13	0.12	0.003	0.35	-0.43	0.49	1	
metro	-0.10	0.32	-0.25	-0.29	-0.20	0.46	-0.27	-0.13	1

Assumption 4: Zero Conditional Mean

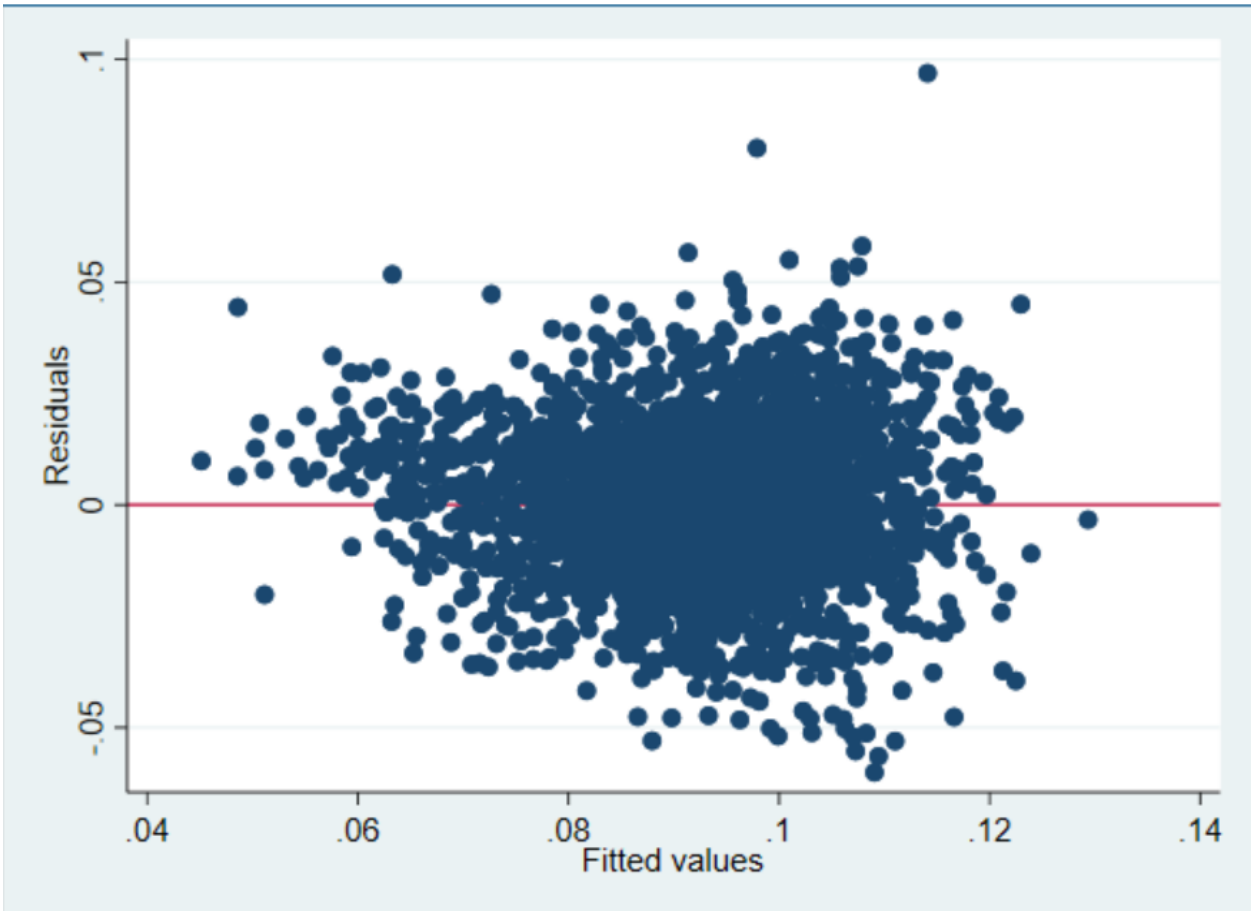
The zero conditional mean assumption assumes that the values of the explanatory values in the model do not include any information from unobserved factors. Since it is known that there are a variety of social determinants that impact chronic illnesses and not just economic determinants as we test here, interpretation of the results should be assessed accordingly. It is difficult to make this assumption as omitted variables may have information about the error term. One example of an omitted variable is the density of fast food restaurants in a county. Availability of unhealthier and cheaper fast food is common in low income areas and could also contribute to the diagnosed diabetes percentages of a county. Unfortunately, data for this was unavailable and

was not used in this analysis. This and other such missing variables should be considered when interpreting the results of this study.

Assumption 5: Homoscedasticity

The homoscedasticity assumption assumes that the explanatory variables do not have any influence on the variance of the error term. To check for this assumption, a plot was generated to show the residuals and fitted values of the model. The residuals here are relatively low and are clustered along the line with few outliers.

Figure 2: Plot of Residuals and Fitted Values



Assumption 6: Normality

This assumption requires that the population error, u, be independent of the explanatory variables included and that u is normally distributed. This is difficult to prove because as listed in Assumption 4, there are likely many terms which are not included in this analysis that would

influence the error term. The fact that this study cannot encompass all of the potential factors that impact diabetes outcomes makes it unfeasible to prove the normality assumption for the error term.

IV. Results

Several regression models were used to test this hypothesis. The complete STATA output can be found in the Appendix.

Model 1:

The first model is a simple linear regression that is used to represent the relationship between a county's diagnosed diabetes rate and the log of the county's median household income.

$$\text{diabetes} = B_0 + B_1(\text{logMedHouseInc}) + u$$

The STATA output from running this regression provides the following model.

$$\text{diabetes} = 0.590 - 0.047(\text{logMedHouseInc}) + u$$

The interpretation of this coefficient shows that a 1% increase in the median household income will lead to a 0.00047% decrease in diagnosed diabetes percentage in a county. This regression has an R-squared value of 0.327; while low, it shows that this variable alone accounts for a significant portion of the relationship. This coefficient was significant at the 1% level as signified by p-values of 0 and large t-values.

Model 2:

The initial multiple linear regression includes all of the variables to better explain the relationship between the social determinants listed that affect diabetes prevalence in an area.

$$\text{diabetes} = B_0 + B_1(\text{MDs10k}) + B_2(\text{centuninsur}) + B_3(\text{medAge}) + B_4(\text{popNoHSDip}) + B_5(\text{logMedHouseInc}) + B_6(\text{poverty}) + B_7(\text{urate}) + B_8(\text{metro}) + u$$

The STATA output from running this regression provides the following model:

$$\text{diabetes} = 0.321 - 0.336(\text{MDs10k}) - 0.045(\text{centuninsur}) + 0.001(\text{medAge}) + 0.098(\text{popNoHSDip}) - 0.027(\text{logMedHouseInc}) + 0.047(\text{poverty}) + 0.025(\text{urate}) + 0.009(\text{metro}) + u$$

Predictably, the inclusion of other variables increases the strength of the relationship defined by the model as evidenced by a higher R-squared of 0.4666. The variables varied in significance however, with the number of doctors available per 10,000 people, MDs10k, being significant at the 5% level, the unemployment rate insignificant at even the 10% level, and all other variables being significant at the 1% level. These significance levels are determined by the p-values; the urate in particular has a high p-value of 0.152, indicating its insignificance. In this model, the impact of the log median household income decreases as other variables start to explain the relationship better; a 1% increase in median household income corresponds to a 0.00027% decrease in diagnosed diabetes percentage.

Model 3:

With the knowledge that some of the variables are less significant in their contribution to the diagnosed diabetes rate, the third multiple linear regression model excludes variables MDs10k and the urate.

$$\text{diabetes} = B_0 + B_1\,(\text{centuninsur}) + B_2(\text{medAge}) + B_3(\text{popNoHSDip}) + B_4(\text{logMedHouseInc}) + B_5(\text{poverty}) + B_6(\text{metro}) + u$$

The STATA output from running this regression provides the following model.

$$\text{diabetes} = 0.329 - 0.045(\text{centuninsur}) + 0.001(\text{medAge}) + 0.100(\text{popNoHSDip}) - 0.027(\text{logMedHouseInc}) + 0.048(\text{poverty}) + 0.009(\text{metro}) + u$$

While the R-squared of this model decreased slightly to 0.4656, this model proves to be a better model than the previous one as all of the variables turn out to be significant at the 1% level with all of the p-values being 0.000. The coefficient on the main dependent variable of median household income remains the same in this model.

Table 4: A Summary of the Regression Models

Independent Variable	Model 1	Model 2	Model 3
logMedInc	-0.0467*** (0.001)	-0.0267*** (0.003)	-0.0274*** (0.002)
MDs10k		-0.3362** (0.005)	
centuninsur		-0.0447*** (0.005)	-0.0453*** (0.005)
medAge		-0.0009*** (0.000)	0.0009*** (0.000)
popNoHSDip		0.0981*** (0.005)	0.1004*** (0.005)
Poverty		0.0470*** (0.010)	0.0477*** (0.009)
Urate		0.0249 (0.017)	
Metro		0.0087*** (0.001)	0.0085*** (0.001)
Intercept	0.5895*** (0.013)	0.3212*** (0.029)	0.3289*** (0.029)
Observations	3140	3140	3140
R^2	0.3270	0.4666	0.4656
Adjusted R^2	0.3268	0.4653	0.4645

** = significant at 5% *** = significant at 1%

V. Extension

Upon running the regressions, it is interesting to see that the dummy variable defining the county's rural or urban status is highly significant with correspondingly high t-values. The next steps for this paper involves determining if there is a difference in each of the urban or rural groups. A Chow test will be utilized to note if there is a statistically significant difference in the coefficients for each split of the data. In order to execute this test, three new regressions were done: a regression with all the significant variables without the dummy included, a regression done on rural counties only, and a regression done on urban counties only.

The pool model equation is as follows:

$$\text{diabetes} = 0.221 - 0.053(\text{centuninsur}) + 0.001(\text{medAge}) + 0.103(\text{popNoHSDip}) - 0.017(\text{logMedHouseInc}) + 0.064(\text{poverty}) + u$$

The null hypothesis then is that there is no statistically significant difference between the coefficients of the models for rural counties and urban counties.

The equation used to conduct the Chow test is:

$$F = \frac{[SSR_{pool} - (SSR_{rural} + SSR_{urban})]}{SSR_{rural} + SSR_{urban}} * \frac{n - 2(k+1)}{k+1}$$

The resultant F value from this equation is 37.47. The critical value at the 10% level with 3140 observations is 1.77. With the F value from the Chow test being much larger than this number, the null hypothesis that the coefficients between the rural and urban models are the same is rejected. This Chow test proves that there is a statistically significant difference between the models for urban and rural counties.

VI. Conclusion

After interpreting the regression results and the models, there is sufficient evidence to support the hypothesis that an increase in median household income significantly corresponds to a decrease in diagnosed diabetes rates for a county. In the final model where all of the variables are significant, the coefficient on the main independent variable is 0.027. The interpretation of this coefficient reveals that a 1% increase in median household income corresponds to a 0.00027% decrease in diagnosed diabetes percentage. Although this may seem low, it is important to note that the diagnosed diabetes rates tend to range between 0 and less than 20% of the population and there is not much room for variance, making even small percentage changes meaningful. While all of the variables used in the final model were highly significant with miniscule p-values, the R-squared value was 0.04656, meaning the model accurately presented a little less than half of the full picture. There are certainly a plethora of other social determinants that could significantly impact the results of this study that were excluded due to lack of data. This omitted variable bias may be impacting the results gleaned; the coefficients are likely overestimating the impact of each of the variables as they take on the weight of the variables missing from the model.

This paper also shows that there is a statistically significant difference in the models between urban and rural counties by using the Chow test. This has important policy implications; understanding that the different metropolitan regions face different models describing the impact

social determinants have on diabetes prevalence means that a blanket solution or policy will be ineffective in meeting the needs of the different types of populations. Further should be conducted upon each subsection of the overall pool of all counties in order to understand the needs of different metropolitan regions and better tackle their individual issues.

VII. References

Baffour, T. D. (2017). Addressing the Social Determinants of Behavioral Health for Racial and Ethnic Minorities: Recommendations for Improving Rural Health Care Delivery and Workforce Development. Journal of Best Practices in Health Professions Diversity, 10(2). https://www.jstor.org/stable/26554276

Singh, G., Daus, G., Allender, M., Ramey, C., Martin, E., Perry, C., Reyes, A., & Vedamuthu, I. (2017). Social Determinants of Health in the United States: Addressing Major Health Inequality Trends for the Nation, 1935–2016. International Journal of MCH and AIDS (IJMA), 6(2). https://doi.org/10.21106/ijma.236

Streeter, R. A., Snyder, J. E., Kepley, H., Stahl, A. L., Li, T., & Washko, M. M. (2020). The geographic alignment of primary care Health Professional Shortage Areas with markers for social determinants of health. PLOS ONE, 15(4), e0231443. https://doi.org/10.1371/journal.pone.0231443

Tikkanen, R., & Abrams, M. (2020, January 30). U.S. Health Care from a Global Perspective, 2019: Higher Spending, Worse Outcomes? The Commonwealth Fund. https://www.commonwealthfund.org/publications/issue-briefs/2020/jan/us-health-care-global-pers pective-2010

Appendix A: Correlation Table

```
. correlate diabetes MDs10k centuninsur medAge popNoHSDip logMedHouseInc poverty urate metro
(obs=3,140)
```

	diabetes	MDs10k	centun~r	medAge	popNoH~p	logMed~c	poverty	urate	metro
diabetes	1.0000								
MDs10k	-0.1891	1.0000							
centuninsur	0.2607	-0.2449	1.0000						
medAge	0.1267	-0.1943	-0.0929	1.0000					
popNoHSDip	0.5448	-0.2505	0.6086	-0.1591	1.0000				
logMedHous~c	-0.5718	0.2997	-0.4834	-0.1174	-0.6135	1.0000			
poverty	0.5356	-0.1531	0.5054	-0.1889	0.6750	-0.8411	1.0000		
urate	0.3358	-0.1332	0.1183	0.0029	0.3516	-0.4321	0.4863	1.0000	
metro	-0.0951	0.3230	-0.2457	-0.2862	-0.2032	0.4588	-0.2674	-0.1332	1.0000

Appendix B: STATA output

Model 1:

```
. regress diabetes logMedHouseInc
```

Source	SS	df	MS		
Model	.396426879	1	.396426879	Number of obs	= 3,140
Residual	.815925067	3,138	.000260014	F(1, 3138)	= 1524.63
				Prob > F	= 0.0000
				R-squared	= 0.3270
				Adj R-squared	= 0.3268
Total	1.21235195	3,139	.000386222	Root MSE	= .01612

diabetes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logMedHouseInc	-.0467445	.0011971	-39.05	0.000	-.0490918	-.0443972
_cons	.5895244	.0127383	46.28	0.000	.5645481	.6145006

Model 2:

```
. regress diabetes MDs10k centuninsur medAge popNoHSDip logMedHouseInc poverty urate metro
```

Source	SS	df	MS	Number of obs	=	3,140
				F(8, 3131)	=	342.42
Model	.565736064	8	.070717008	Prob > F	=	0.0000
Residual	.646615882	3,131	.000206521	R-squared	=	0.4666
				Adj R-squared	=	0.4653
Total	1.21235195	3,139	.000386222	Root MSE	=	.01437

diabetes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
MDs10k	-.3361674	.1695881	-1.98	0.048	-.6686825	-.0036523
centuninsur	-.0447155	.0050428	-8.87	0.000	-.0546031	-.0348279
medAge	.0008611	.000062	13.89	0.000	.0007395	.0009827
popNoHSDip	.0981254	.0049323	19.89	0.000	.0884545	.1077963
logMedHouseInc	-.0266714	.0025057	-10.64	0.000	-.0315843	-.0217585
poverty	.0470471	.0095343	4.93	0.000	.0283529	.0657412
urate	.0248615	.0173327	1.43	0.152	-.0091231	.0588462
metro	.0087223	.0006356	13.72	0.000	.0074761	.0099686
_cons	.3211682	.0290582	11.05	0.000	.2641931	.3781433

Model 3:

```
. regress diabetes centuninsur medAge popNoHSDip logMedHouseInc poverty metro
```

Source	SS	df	MS	Number of obs	=	3,140
				F(6, 3133)	=	454.86
Model	.564413802	6	.094068967	Prob > F	=	0.0000
Residual	.647938145	3,133	.000206811	R-squared	=	0.4656
				Adj R-squared	=	0.4645
Total	1.21235195	3,139	.000386222	Root MSE	=	.01438

diabetes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
centuninsur	-.0452956	.0049143	-9.22	0.000	-.0549311	-.0356601
medAge	.0008799	.0000616	14.29	0.000	.0007592	.0010006
popNoHSDip	.1003542	.0048558	20.67	0.000	.0908334	.109875
logMedHouseInc	-.0274207	.0024802	-11.06	0.000	-.0322837	-.0225576
poverty	.0476568	.0091033	5.24	0.000	.0298079	.0655058
metro	.0085282	.0006272	13.60	0.000	.0072984	.0097581
_cons	.3288838	.0288581	11.40	0.000	.272301	.3854665

Appendix C: State output for Chow Test

Pool Regression Output

```
. regress diabetes centuninsur medAge popNoHSDip logMedHouseInc poverty
```

Source	SS	df	MS	Number of obs	=	3,140
				F(5, 3134)	=	480.66
Model	.526183036	5	.105236607	Prob > F	=	0.0000
Residual	.686168911	3,134	.000218943	R-squared	=	0.4340
				Adj R-squared	=	0.4331
Total	1.21235195	3,139	.000386222	Root MSE	=	.0148

diabetes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
centuninsur	-.0526866	.0050253	-10.48	0.000	-.0625399	-.0428333
medAge	.0007437	.0000625	11.90	0.000	.0006212	.0008663
popNoHSDip	.1034944	.0049905	20.74	0.000	.0937094	.1132794
logMedHouseInc	-.0166532	.0024184	-6.89	0.000	-.0213949	-.0119115
poverty	.0635219	.0092892	6.84	0.000	.0453084	.0817355
_cons	.2214318	.0285574	7.75	0.000	.1654387	.2774249

Regression Output for Rural Countries (metro==0)

```
. regress diabetes centuninsur medAge popNoHSDip logMedHouseInc poverty if metro==0
```

Source	SS	df	MS	Number of obs	=	1,960
				F(5, 1954)	=	310.81
Model	.320295466	5	.064059093	Prob > F	=	0.0000
Residual	.402727617	1,954	.000206104	R-squared	=	0.4430
				Adj R-squared	=	0.4416
Total	.723023083	1,959	.000369078	Root MSE	=	.01436

diabetes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
centuninsur	-.0393364	.0058849	-6.68	0.000	-.0508778	-.027795
medAge	.0006513	.0000777	8.38	0.000	.0004989	.0008036
popNoHSDip	.0891166	.0057576	15.48	0.000	.077825	.1004083
logMedHouseInc	-.0263153	.0036826	-7.15	0.000	-.0335376	-.019093
poverty	.052241	.0113168	4.62	0.000	.0300467	.0744354
_cons	.3267397	.0424281	7.70	0.000	.2435306	.4099488

Regression Output for Urban Countries (metro==1)

```
. regress diabetes centuninsur medAge popNoHSDip logMedHouseInc poverty if metro==1
```

Source	SS	df	MS	Number of obs	=	1,180
				F(5, 1174)	=	239.43
Model	.241516288	5	.048303258	Prob > F	=	0.0000
Residual	.236846544	1,174	.000201743	R-squared	=	0.5049
				Adj R-squared	=	0.5028
Total	.478362833	1,179	.000405736	Root MSE	=	.0142

diabetes	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
centuninsur	-.0567289	.0090155	-6.29	0.000	-.0744172	-.0390406
medAge	.0012997	.0001087	11.96	0.000	.0010865	.0015129
popNoHSDip	.1151748	.0096025	11.99	0.000	.0963349	.1340147
logMedHouseInc	-.0279179	.0037172	-7.51	0.000	-.0352111	-.0206247
poverty	.0483557	.0183529	2.63	0.009	.0123476	.0843638
_cons	.3262964	.0442265	7.38	0.000	.2395246	.4130682