The impact of insurance coverage rates on obesity rates

Kaleb Shah

May 7th, 2020

Abstract

Obesity is one of the most prevalent diseases in the US that is yet to be tackled by the healthcare system. I use Government sourced data to analyze if the percent of the population covered under health insurance impacts the obesity rates. I measure obesity by looking at every US States' obesity rates for 5 years. Similarly, my variable of interest is the percentage of people without insurance in each state within the aforementioned timespan. Primary results indicate that there is positive relationship between the percent uninsured and obesity rates. However, after conducting a subsample analysis, we find no significance. Ultimately, the findings are inconclusive.

I. Introduction

The US has the one of the highest rates of obesity in the world. With over a 42% of the population suffering from obesity, the annual cost of obesity and its related diseases amount to $147 Billion. The association between mental disorders and obesity also prompts more attention (Becker et al, 2001). Even though obesity has become a national crisis, there has been little to no change in obesity rates across the US. It is suggested that there is a link between poverty and obesity in the US as access to healthy and nutritious food is often not available to people living in poverty. Similarly, people in impoverished neighborhoods cannot exercise effectively as gyms are rare and, working out outdoors may not be safe due to street violence (Levine, 2011). Since a huge population of the US living in poverty do not have health insurance, we must investigate obesity rates for populations like these. Thus, this paper examines whether the percentage of the population with health insurance impacts the obesity rates.

There are a lot of factors that influence obesity rates, these include socio-economic status, education, diet, culture and more (Kim, 2019). Most national and state level data for obesity rates are gathered by the Centre for Disease Control (CDC) since it is a disease. Obesity is measured my body mass index (BMI). An individual with a BMI of 30 or higher is considered to have obesity. By sampling approximately 400,000 US citizens across all states, the CDC is able to collect data on percentage of people in each state with obesity every year. By exploiting this data, we can begin to answer some questions.

For the empirical model, I use the data mentioned above to measure obesity rates and I use data from the Kaiser Family Foundation (KFF) to measure the percent of population without health insurance coverage. Since the data spans 5 years across 50 states and the district of Columbia, we have 251 observations. I further incorporated state level GDP per capita, unemployment rate, education and exercise as controls in the model. All this data was sourced from Government sources. Finally, we apply fixed effects on states and years to control differences that remain constant across these variables.

The primary results suggest that there is a positive relationship of 0.150 between the percent uninsured and the obesity rates with a p-value < 0.05. However, after conducting a subsample analysis by clustering each state into its respective region (Northeast, Midwest, South and West), we find no significance between the variable of interest and the obesity rates. Thus, the results remain inconclusive. Because clustering the data according to regions reduces the number of observations within each subsample, it is plausible that we can find significance with a larger number of observations spanning a larger time span.

In addition, the paper other sections of the paper are as follows. The second section discusses existing literature on the research question while the third section explains the sources of data and its collection. The fourth section explains the empirical model and the fifth section interprets the results of the models. The sixth and final section concludes the paper.

II. Literature Review

There is various pre-existing research on financial and social predictors of obesity rates. The following papers offer an opposing view to this paper by positing that health insurance has negative impact on obesity rates. “Incentives in Obesity and Health Insurance.” (Rashad et al, 2009) examines the interesting effect of ex ante moral hazard behavior exhibited by people after purchasing or obtaining health insurance. The paper finds an association between health insurance and body mass but finds no significance between the variable of interest and obesity. Similarly, "Smoking, Obesity, Health insurance and Health Incentives in the Affordable Care Act (ACA)" (Schmidt, 2013) tries to extrapolate how customer protection methods afforded to people by the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and the ACA incentivize obesity instead. Because the additional cost of insuring someone with obesity is absorbed by the population, and because people with pre-existing conditions are protected under the ACA, there are no disincentives against obesity.

On the other hand, "Insurance coverage for weight loss" (Bennet et al, 2013) takes another path. They examine the personal opinions of people with obesity by surveying a sample of 600 obese people who visited a primary care provider in the past year. The majority, 83% of respondents consider health insurance beneficial in treating obesity and its related diseases.

Lastly articles and papers such as "Poverty and Obesity in the US" (Levine, 2011) and "Links between Poverty and Obesity during the Transition to Young Adulthood" (Lee et al, 2009) argue that poverty and a lack of resources increase the probability of obesity. The first article explains that as the level of poverty by county rises, the obesity rates rise as well. This article further highlights that poverty restricts access to fresh and healthy food, thus increasing rates of obesity. Moreover, the latter paper finds an association between poverty and obesity in young females. There are other papers that examine the impact of socio-economic factors on obesity across all age groups. However, studies on health insurance and obesity are limited but continue to grow.

III. Data

The data for this project comes from various Government sources with the exception of the KFF. Due to a lack of existing datasets with the necessary measures, compiling the primary dataset involved extracting individual variables from these different sources and composing them manually. Because of this, the dataset and its observations are constrained. The data is state level and spans 5 years, with the inclusion of the 50 states and the District of Columbia. The result is 255 observations. Each observation is a state and its annual variables. The first variable collected was the variable of interest, the percentage of the population without health insurance. This is the only variable collected from a non-government website. However, their dataset is also based on the Census Bureau's American Community Survey (ACS). This survey samples 1% of the US population resulting in what is widely considered precise state-level estimates. However, criticisms on this data's reliability and margin of error have been raised (Spielman et al, 2014). Although 1% of the total US population may seem to be a sufficiently large enough sample size, some state estimates are based on a sample size in the thousands which could create sampling error.

Furthermore, the next variable collected was the response variable, obesity rates. This data was sourced from the CDC. The national survey as mentioned earlier samples roughly 400,000 US citizens to estimate the entire US population. Similar to the ACS data mentioned above, this data is not only prone to sampling errors. According to the Census (ACS

Accuracy of the Data, 2018), the integrity of the data can be threatened by non-sampling errors as well. This includes coverage error and item non-response error. The latter is especially critical, as people with obesity may feel embarrassed to answer personal questions regarding their health and weight. As such, any conclusion derived there from are similarly questionable. Equally important is the definition of Obesity. According to this method of collection, any individual with a Body Mass Index (BMI) of 30 or greater is deemed obese. This measure is often criticized as poor indicator of obesity as it is measured using height and weight. Furthermore, an article in the Journal of Obesity argues "the use of BMI as a measure of obesity can introduce misclassification problems that may result in important bias in estimating the effects related to obesity." Because "It does not, for example, take age, sex, bone structure, fat distribution or muscle mass into consideration." (Rothman, 2008). These concerns should be considered when interpreting the results.

To avoid omitted variable bias, control variables for the model were necessary. Based on existing literature (Kim et al, 2019), the model needed controls for the following: Income, Poverty, education and fitness. For income and poverty, we added each state's GDP per capita collected from the Bureau of Economic Analysis (BEA) and state unemployment rates gathered from the Bureau of Labor Statistics (BLS). Likewise, for education we measured quality and quantity. On one hand, to measure the quality of data, we chose to add each state's average grade 8 school scores for reading and math from the National Assessment of Educational Process (NAEP). On the other hand, to measure quantity, we sourced the percentage of population over 25 of every state that have completed high school and a bachelors. This data was gathered by the National Center for Education Statistics (NCES).

Lastly, to measure fitness, we looked at the percentage of adults that engage in no leisure-time physical activity of every state from the CDC.

Table 1 displays the summary statistics of every variable. Notice that the mean uninsured rate is 8.69% and the mean obesity rate is 30.04%. We can notice that the standard deviations are small relative to the mean value of the summary statistic. On the contrary, the range for the summary statistics are substantial. The large range could be attributed to the expansive and diverse population trying to be estimated by the relatively small sample size of the survey data.

IV. Empirical Model

To answer the research question, we create a linear regression with multiple regressors. The empirical model for this procedure is as follows:

$$Y_{it} = \beta_0 + \beta_1 X_1 + [\beta_i]\left[X_i\right] + \delta_i + \gamma_t + \varepsilon_{it},\ i = 1, ..., n,\ t = 1,\ ...T,$$

where Y_{it} is an entity that denotes the obesity rate in state i at time t. There are n = 51 states. Additionally, in this model, X_1 is the percent of the population uninsured while β_1 is its coefficient. The term $[X_i]$ is the vector of all the control variables while $[\beta_i]$ is the corresponding vector of coefficients for each of the control variables: uninsured rate, unemployment rate, percent of adult population with no leisure-time exercise, each state's average grade eight reading and math scores, the high school completion rate and the bachelors' completion rate for adults over the age of 25 years.

The term ε_{it} denotes the error term of an entity in time t. To ensure the integrity of the OLS regressors, we ensure the least square assumptions are satisfied. We assume that $E(\varepsilon_{it}) = 0$. That is the conditional distribution of the error term given X_i is 0. By using robust standard errors in every model and later using fixed effects, the threats of heteroskedasticity are eased. Based on the sample selection process of the Census data, we assume that (Y_i, X_i) are independently and identically distributed. Alike, based on the summary statistics, we can notice that there are no large outliers. Lastly, the model was created on Stata. Since Stata did not drop any of the variables from the model, we can assume there is no perfect multicollinearity. Thus, our least square assumptions are satisfied.

In order to reduce omitted variable bias, we added the controls in the vector $[X_i]$ mentioned above. However, due to a lack of time constraints and availability of data, certain controls such as culture/eating habits and demographics etc. cannot be added. Thus, to control for omitted variable bias across states and time, we added the state and time fixed effects δ_i and γ_t respectively. On one hand, the state fixed effects absorb the influences of omitted variables that vary across time but are constant across all states. On the other hand, time fixed effects absorb the influences of omitted variables that change from state to state but are constant over time.

Similarly, to ensure the robustness of the estimates from the regression model, we conducted a subsample analysis. By clustering the states their respective regions: Northeast, Midwest, South and West. We create 4 subsamples that we further look at individually.

V. Results

To measure the impact of the uninsured rate on obesity rates, we first create a simple OLS model by adding only one regressor, the variable of interest, to the first model. We observe a strong positive correlation of 0.299 with a p-value < 0.001 between the uninsured rate and obesity. This strong significance warrants further investigation. However, as numerous research highlights, there are several factors influencing obesity rates. In ordered to mitigate the issues that arise due to omitted variable bias, we add control variables. By adding control variables to the model, we not only reduce the high standard errors caused by using only one regressor, but we also reduce risk of confounding variables. Therefore, we introduce controls into the next model.

In the second model, we introduce controls for poverty, income and fitness. We do this by adding GDP per capita, unemployment rate and the no exercise rate. As seen in Table 2, after adding these controls, we find that the variable of interest loses its significance. Additionally, the strength of the coefficient also reduces. Thus, we add more controls to improve the model.

In the third model, we add controls for quality and quantity of education. For quality of education, we add each state's average grade eight reading and mathematics scores. To measure quantity of education, we look at each state's high school completion rate and bachelor's completion rate. In doing so, our coefficient for the uninsured rate becomes negative for the first time. However due to a lack of significance, we can ignore this change.

In the fourth model, we further reduce omitted variable bias and control for heterogeneity. We add fixed effects for time and fixed effects for state. In this model, as seen in Table 2, our variable of interest, the uninsured rate has some significance with a coefficient of 0.151 and a p-value < 0.05. Although this implies variable of interest and our response have some association, we can further test for robustness of the model and its conclusions by conducting a subsample analysis.

We cluster each state into their respective regions including Northwest, Midwest, South and East. After applying the same controls and fixed effects as model 4 to each of these clusters, we can observe the results in Table 3. It is clear across all subsamples that there is no significance in the association between uninsured rates and obesity rates. This could be attributed to the decrease in the number of observations since the number of observations reduced from 255 to an average of ~64 observations in each subsample. Despite this fall in the number of observations, the lack of significance across the subsample analysis calls the results from the last model into question. Thus, ultimately the results are inconclusive.

VI. Conclusion

Obesity rates across the US have been rising constantly over time reaching an all-time high of 42%. Forecasts predict that by 2030, over 50% of the adult US population could suffer from obesity adding $550 Billion to annual healthcare costs during that time. On a similar note, health insurance and universal healthcare reform has been discussed recently. Since a huge population of the US lives without insurance, we look at the impact of insurance coverage rates on Obesity rates across the US.

By using data collected by the Census and its ACS, we analyze state level uninsured rates and obesity rates across five years from 2014 to 2018 including. Furthermore, we compile data from distinct Government data sets to add control variables for the model based on existing literature on the factors affecting obesity rates. In our case, these include controls for income, exercise and education.

We try to predict the response variable by using the variable of interest and the control variables. Furthermore, we add fixed effects for state and time to complete the full model and ensure its robustness. Primary results indicate some significance regarding the relationship between uninsured rates and obesity. The model suggests a coefficient of 0.151 with a p-value < 0.05. Theory would suggest that we observe these results because a greater percent of the population with health insurance means more access to healthcare and health education from primary care providers. To further validate these findings, we conducted a subsample analysis by clustering each of the states into their respective geographical regions. Results from this model show no significance of the uninsured rate on obesity rates across all the regions. As a result, this paper's findings are inconclusive. However, it is plausible to find significance with a larger number of observations across a longer time span.

The external validity of this study is threatened by several factors. These factors include differences in population and differences in setting. The obesity rates of the US are unique to a developed country with such a vast population. Furthermore, the access to fast food restaurants and the culture surrounding food in the US may be unique to its population and setting. Furthermore, the healthcare markets of countries vary greatly. For example, out of pocket medical costs in the US are unaffordable making health insurance the only reasonable way to a acquire health care. Similarly, this study is not externally valid to countries with universal healthcare because their entire population has health insurance. However, since this study was designed with US in mind, external validity is not a grave concern.

This paper adds to exiting literature on this area of study. Based on the primary results of the paper, the positive impact of uninsured rates on obesity, policy recommendations would include universal health care as increase in the percent of the population with health insurance would reduce the obesity rate. However, the lack of significance warrants a more expansive study. An example of this would be studying county level data over a longer time span. To conclude, the lack of consensus on the impacts of percent insured on obesity rates further advocates the necessity of academic inquiry and discussions.

VII. References

Kelly, Inas Rashad, and Sara Markowitz. “Incentives in Obesity and Health Insurance.” *Inquiry*, vol. 46, no. 4, 2009, pp. 418–432. *JSTOR*, www.jstor.org/stable/29773445. Accessed 6 May 2020.

Madison K, Schmidt H, Volpp KG. Smoking, Obesity, Health Insurance, and Health Incentives in the Affordable Care Act. *JAMA.* 2013;310(2):143–144. doi:10.1001/jama.2013.7617

Jarlenski, Marian P et al. “Insurance coverage for weight loss: overweight adults' views.” *American journal of preventive medicine* vol. 44,5 (2013): 453-8. doi:10.1016/j.amepre.2013.01.021

Spielman, Seth E et al. “Patterns and causes of uncertainty in the American Community Survey.” *Applied geography (Sevenoaks, England)* vol. 46 (2014): 147-157. doi:10.1016/j.apgeog.2013.11.002

Rothman, K. BMI-related errors in the measurement of obesity. *Int J Obes* 32, S56–S59 (2008). https://doi.org/10.1038/ijo.2008.87

Centers for Disease Control and Prevention. National Center for Chronic Disease Prevention and Health Promotion, Division of Nutrition, Physical Activity, and Obesity. Data, Trend and Maps [online]. [accessed May 08, 2020].
URL: https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html.

Becker, E., Margraf, J., Türke, V. *et al.* Obesity and mental illness in a representative sample of young women. *Int J Obes* 25, S5–S9 (2001). https://doi.org/10.1038/sj.ijo.0801688

Levine, James A. “Poverty and Obesity in the U.S.” *Diabetes*, American Diabetes Association, 1 Nov. 2011, diabetes.diabetesjournals.org/content/60/11/2667.full.

Lee, H., Harris, K.M. & Gordon-Larsen, P. Life Course Perspectives on the Links Between Poverty and Obesity During the Transition to Young Adulthood. *Popul Res Policy Rev* 28, 505 (2009). https://doi.org/10.1007/s11113-008-9115-4

Kim D, Hou W, Wang F, Arcan C. Factors Affecting Obesity and Waist Circumference Among US Adults. Prev Chronic Dis 2019;16:180220.
DOI: http://dx.doi.org/10.5888/pcd16.18022

VIII. Appendix

Table 1 – Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
obesityper~e	255	30.04039	3.814515	20.2	39.5
uninsuredr~e	255	8.694125	3.335885	2.494445	19.07565
unemployment	255	4.583529	1.288274	2.4	7.9
noexercise~e	255	24.53765	4.185727	15.7	36.8
gdppercapita	255	49911.51	18004.04	31522	163274
gr8reading	255	265.3725	5.777913	247	278
gr8math	255	282.2392	7.139887	263	301
highschool~e	255	88.82865	2.938244	81.737	93.5966
bachelorsc~e	255	30.45037	6.220977	18.8091	57.4125

Table 2 – Primary Results (Models 1, 2,3 and 4)

	(1) obesityper~e	(2) obesityper~e	(3) obesityper~e	(4) obesityper~e
uninsuredr~e	0.299***	0.113	-0.0303	0.151*
	(4.45)	(1.83)	(-0.48)	(2.10)
unemployment		-0.204	-0.387**	-0.836***
		(-1.43)	(-2.68)	(-6.20)
gdppercapita		-0.0000397***	0.0000289	-0.0000302
		(-4.82)	(1.58)	(-1.06)
noexercise~e		0.535***	0.386***	0.00690
		(12.31)	(8.20)	(0.16)
gr8reading			-0.0804	-0.0946
			(-1.11)	(-1.71)
gr8math			0.0134	-0.0790
			(0.30)	(-1.69)
highschool~e			0.104	0.0882
			(1.02)	(0.67)
bachelorsc~e			-0.365***	-0.163*
			(-6.92)	(-2.06)
_cons	27.44***	18.84***	40.59***	78.41***
	(42.04)	(14.51)	(3.44)	(5.20)
N	255	255	255	255

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Table 3 – Subsample Analysis

	(1) Northeast	(2) Midwest	(3) South	(4) West
uninsuredr~e	0.0377	0.338	0.168	0.0500
	(0.17)	(1.50)	(1.54)	(0.47)
unemployment	-0.574	-0.634	-0.209	-0.157
	(-2.28)	(-1.48)	(-1.07)	(-0.41)
noexercise~e	-0.0443	-0.000293	-0.0967	-0.0483
	(-0.50)	(-0.00)	(-1.72)	(-0.66)
gr8reading	0.0325	-0.0452	0.0245	-0.0887
	(0.12)	(-0.32)	(0.32)	(-1.12)
gr8math	-0.0484	0.175	0.0519	-0.170
	(-0.49)	(0.98)	(0.58)	(-1.42)
highschool~e	-0.229	0.342	0.985***	0.176
	(-0.60)	(0.65)	(4.03)	(0.35)
bachelorsc~e	0.183	0.603***	0.366*	0.325
	(0.89)	(6.06)	(2.82)	(1.79)
_cons	49.89	-55.02	-81.97	74.31
	(0.83)	(-0.59)	(-1.76)	(1.02)
N	45	60	85	65

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001