Stata Task 4 - Effect of Obesity on Overall Health and Cost of Care

Teddy Foley

March 2023

1 Introduction

In the United States, 39.8% of the population is obese (Fryar, 2018). Obesity's impact on the general health and well being of the public has been extensively studied, and the scientific consensus is that it has significant detrimental effects across the board. According to Cawley, et al. "Research to date has consistently demonstrated a correlation between obesity and higher medical costs for a variety of U.S. subpopulations and specific categories of care." In fact, previous research has likely underestimated the influence obesity has on totoal medical expenditures (Cawley, 2021). In this paper, we aim to confirm these results as well as calculate the negative externalities caused by obesity in the form of higher costs of care. Is obesity a better predictor of cost of care than other easily diagnosable diseases? Provided with this information, medical professionals may be able to give more accurate estimates of cost of care and diagnoses of obesity-related illnesses based only on limited information about a patient.

2 Methodology

The dataset used for this paper, *health.dta* has 155 variables describing a wide set of medical data for each patient. Of particular importance is the binary variable for obesity (tm1_obesity_elixhauser), total medical costs (cost_t), and a range of demographic and health-related variables. In order to investigate the relationship between obesity and total medical costs, we will perform a linear regression analysis, using the obesity indicator as the independent variable and total medical costs as the dependent variable. We will control for potential confounding factors such as age, gender, race, income, and insurance type.

The binary obesity variable (tm1_obesity_elixhauser) takes the value of 1 if an individual is classified as obese and 0 otherwise. Consequently, the results of the linear regression are straightforward to interpret. The constant term represents the average total medical costs for a non-obese individual, while the coefficient on the obesity indicator corresponds to the additional medical costs associated with obesity. For example, if the constant term is 10,000 and the coefficient is 5,000, we can say that the average non-obese individual incurs medical costs of 10,000, while an average obese individual incurs 15,000 in medical costs.

To explore the robustness of our findings, we will perform several sensitivity analyses, such as including additional control variables and using alternative specifications of the dependent variable (e.g., log-transformed medical costs). Furthermore, we will conduct subgroup analyses to investigate whether the relationship between obesity and medical costs varies across different population segments, such as age groups. These subgroup analyses will provide a more nuanced understanding of the impact of obesity on medical costs and may inform targeted interventions and policies to reduce the economic burden of obesity.

It is important to note that our analysis is observational in nature, and we cannot draw causal conclusions from our findings. However, by controlling for confounding factors, we aim to provide valuable insights into the association between obesity and total medical costs.

3 Results

We performed several regressions with obesity as the independent variable: one on the total cost, another on the log(cost), and a final regression on the total number of active chronic illnesses. The first and last are displayed below. In addition, we ran the last regression again while controlling for age.

```
    . reg cost_t tm1_obesity_e~r

      Source |       SS           df       MS      Number of obs   =    48,784
-------------+----------------------------------   F(1, 48782)     =    470.66
       Model |  1.5087e+11         1  1.5087e+11   Prob > F        =    0.0000
    Residual |  1.5637e+13    48,782   320551141   R-squared       =    0.0096
-------------+----------------------------------   Adj R-squared   =    0.0095
       Total |  1.5788e+13    48,783   323637265   Root MSE        =     17904

------------------------------------------------------------------------------
             cost_t | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
--------------------+---------------------------------------------------------------
tm1_obesity_elixhauser |   6058.011   279.2389    21.69   0.000       5510.7    6605.323
             _cons |   7096.931   85.11032    83.39   0.000     6930.114    7263.749
------------------------------------------------------------------------------

. reg gagne_sum_t tm1_obesity_e~r

      Source |       SS           df       MS      Number of obs   =    48,784
-------------+----------------------------------   F(1, 48782)     =   6215.24
       Model |  20801.8783         1  20801.8783   Prob > F        =    0.0000
    Residual |  163269.082    48,782  3.34691243   R-squared       =    0.1130
-------------+----------------------------------   Adj R-squared   =    0.1130
       Total |   184070.96    48,783  3.77326037   Root MSE        =    1.8295

------------------------------------------------------------------------------
          gagne_sum_t | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
----------------------+---------------------------------------------------------------
tm1_obesity_elixhauser |   2.249462   .0285331    78.84   0.000     2.193536    2.305387
               _cons |   1.145508   .0086967   131.72   0.000     1.128462    1.162553
------------------------------------------------------------------------------
```

The results from each regression are highly significant ($p < 0.001$). For the first regression, he coefficient of 6058 and constant of 7096 indicates that typically, an obese person pays almost double to total medical

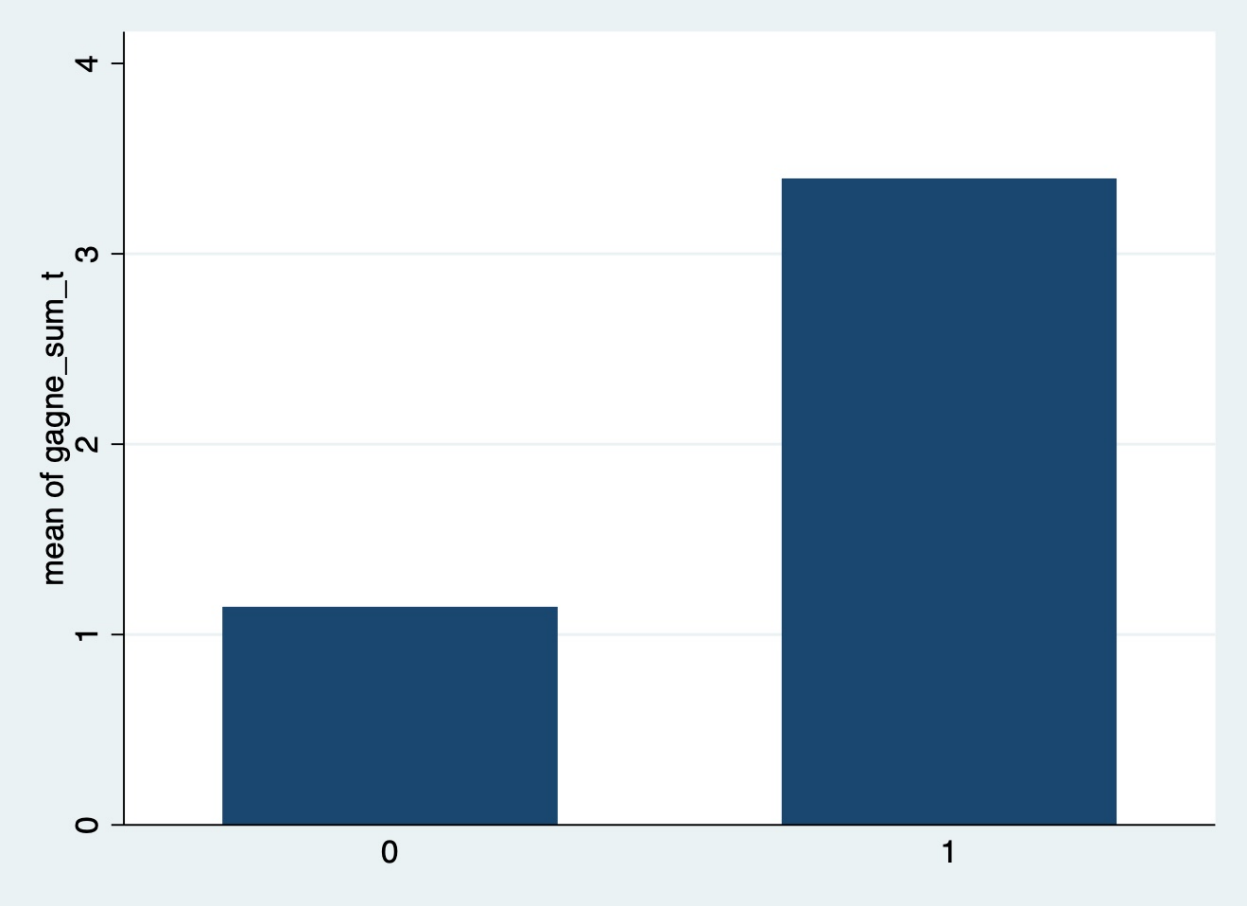

Figure 1: Obesity vs Number of Active Chronic Diseases

expenditures as a non-obese person. However, the R-squared value is only 0.0096, meaning that the model can explain 0.96% of the variance in total medical expenditures. Similar results were found for the regression on the log of costs, making a discussion of those results redundant. These first results suggest that obesity does not play a large role in explaining differences in medical expenditures. On the other hand, based on the second regression above, obese patients average over 3 active chronic diseases to the average patient's 1. Based on an R-squared value of 0.1130, obesity explains about 11.3% of the variance of active chronic diseases. Even when controlling for age, for the regression on the number of chronic illnesses, the coefficient of age was only 0.038 and the R-squared value jumped to 0.219.

4 Conclusion

In light of our findings, we conclude that obesity is significantly associated with higher medical expenditures and a greater number of active chronic diseases. Although the explanatory power of obesity for medical costs is limited, its role in the prevalence of chronic diseases is more pronounced. These results confirm the consensus of the existing literature – obesity should be considered an important factor when assessing healthcare expenditures and managing chronic illnesses.

This study underscores the need for further research to better understand the complex relationship

between obesity, medical costs, and chronic diseases. Investigating potential causal pathways and considering additional factors will be essential to inform targeted interventions and public health policies. By enhancing our understanding of the role of obesity in healthcare and chronic disease management, we hope to contribute to the development of more effective preventative strategies with far-reaching welfare implications.

5 Bibliography

(Fryar, 2018): *https* : //*www.cdc.gov/nchs/data/hestat/obesity_adult*_15_16/*obesity_adult*_15$_1$6.*pdf*

(Cawley, 2021) *https* : //*pubmed.ncbi.nlm.nih.gov*/33470881/