Medical Expenditures and Chronic Illness

Jackson Van Vooren

November 4, 2022

Introduction

This report seeks to understand what relationship exists, if any, between the total number of chronic diseases an individual has and their total medical cost. The primary question is if the quantity of chronic illnesses significantly increases a person's total medical expenditure. Much of the current literature focuses broadly on the relationship between total healthcare cost (for a government) and chronic illness prevalence across a population. For example, the CDC reports 90% of the US's healthcare expense is due to chronic disease[1]. Similarly, the NIH calls this issue the "chronic illnesses epidemic," offering a high level view of the issue for US policy[2]. However, this brief report focuses on the individual-level, comparing the number of diseases to that individual's medical expenditures. As most of the current literature states, chronic illnesses in the United States is a growing issue, with such a disease affecting roughly 50% of the population[3]. Pursuing this question will ideally offer insights into a person's expense, which could be used by insurance companies and hospital enterprises when making pricing decisions for policies and services. While it is generally accepted that increased illnesses increase expenditures, it is especially interesting to see the extent to which one extra illness affects cost.

Methodology

I used the health.dta file (health dataset) to address this research question. Because I am interested in how total cost depends on the number of chronic illnesses, I viewed expenditure as the dependent variable (on the y-axis) and the number of illnesses as the independent variable (placed on the x-axis). Cost is continuous, but the number of chronic diseases is discrete. This makes a scatterplot an effective way to visualize the data, as there will be many cost points at each individual illness number. To check for any relationship between my variables, I used Stata's regression capabilities to check for confidence intervals and correlation coefficients.

Results

When correlating active chronic illnesses and total medical expenditures, I found a correlation coefficient of $R = 0.3547$. This suggests a somewhat moderate correlation between the two variables, but generally, it is not too strong. However, as one variable increases in magnitude, the other should also increase.

For the regression, I let expenditure be the dependent variable and the number of active chronic illnesses be the independent one. This was because I was most interested in seeing how cost depended on a factor that is entirely not controllable.

For the linear regression between the two variables, I found a p-value of 0, so the results are significant at a 5% significance level. The 95% confidence interval yields $[3208, 3361]$. As the number of chronic illnesses increase by 1, we should expect the total cost to increase by $3285.14. The R^2 value of 0.1258 suggests that only 12.5% variability observed in total medical expenditure is explained by the regression model. The scatterplot affirms this as well.

---

[1] https://www.cdc.gov/chronicdisease/about/costs/index.htm
[2] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7077778/
[3] Ibid

Conclusions

There is a moderate positive correlation between total medical expenditures and active chronic illnesses. However, this correlation is not particularly strong. As number of chronic illnesses increase, the total cost does increase, but not completely linearly. Interestingly, the highest cost of medical expenditures is from an individual with no chronic illnesses. This is likely due to there being other emergency situations and illnesses that are not chronic, but rack up large bills. An emergency trip to the hospital could be significantly more expensive than a chronic disease that is not very acute. It is also a possibility that individuals with lots of chronic diseases are unable to afford all the care necessary to truly treat them. From a policy standpoint, it might be advantageous to subsidize medical costs, which would encourage people with chronic illnesses to get the care necessary. There is certainly some positive correlation, so costs to the individual generally could be decreased by subsidizing some healthcare expenses.

I learned that it is important to dispel initial assumptions about certain correlations. I thought there would be a strong correlation between the variables. It seemed intuitive that an increasing number of chronic diseases would necessitate that people spend more money. However, there are many other factors that the data does not capture. It is essential to keep this in mind when considering correlation because confounding factors can affect the conclusions made.

Appendix

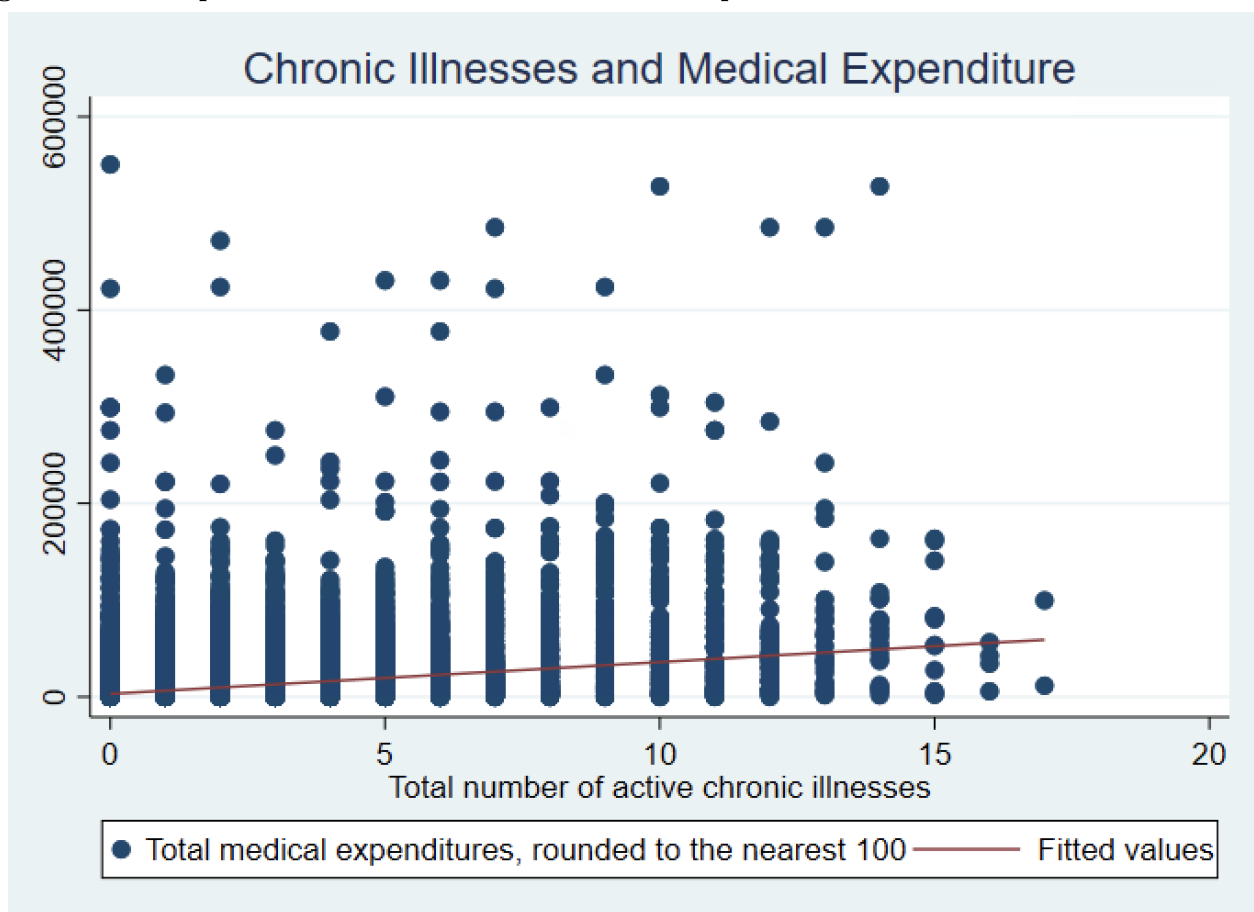**Figure 1. Scatterplot of chronic illnesses and medical expenditure**

Figure 2. Correlation Coefficient

```
             |  cost_t gagne_~t

      cost_t |  1.0000
 gagne_sum_t |  0.3547    1.0000
```

Figure 3. Linear Regression

```
      Source |       SS           df       MS      Number of obs   =     48,784
-------------+----------------------------------   F(1, 48782)     =    7021.43
       Model |  1.9865e+12          1  1.9865e+12   Prob > F        =     0.0000
    Residual |  1.3801e+13     48,782   282921620   R-squared       =     0.1258
-------------+----------------------------------   Adj R-squared   =     0.1258
       Total |  1.5788e+13     48,783   323637265   Root MSE        =      16820


      cost_t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 gagne_sum_t |   3285.135   39.20491    83.79   0.000     3208.293    3361.977
       _cons |   3210.064   92.84035    34.58   0.000     3028.095    3392.032
```