# Evaluating the Effect of Demographic Characteristics on Healthcare Spending

Bradley Rann

April 2023

## 1 Introduction

After Covid-19 the world, and particularly the US, reignited the link between politics and healthcare. The healthcare system in the US is unique for numerous reasons. The most distinctive and politically important is due to the for-profit nature throughout the entire system. Although there is nuance to what is for-profit, essentially all other developed countries have "free" government coverage (like the UK) or a way to purchase affordable insurance (like the Netherlands). Essentially private companies either can't or barely make profit on basic healthcare coverage in other countries, whereas US providers can make profit at every level.

Despite profit incentive, all healthcare systems in growing countries are incurring a larger share their country's GDP share. One explanation for

this is Engel's law. Essentially as household income rises certain goods, like food, see a decrease in percentage income consumption compared to other goods. The key point being as many consumption items have gotten cheaper households have shifted their consumption onto more preferable goods like healthcare. There are other explanations like technological diffusion as well. It's beyond the scope of this paper to confirm or deny some of the claims about the structural shift of healthcare, but it is an important fact to understanding healthcare cost holistically.

The goal of this paper isn't a political, or moral, evaluation of different healthcare systems. The United States spends nearly double per capita than the next closest country on healthcare. It also spends a significant amount of its GDP on healthcare every year. This paper then, will have two goals. First, I will have a brief analysis of the spending and whether it is inefficient, especially compared to other countries. Second, I want to see what is driving the constant increase in healthcare expenditures in the US. From there I want to see if any of the factors like increasing obesity can predict what future healthcare expenditure might be holding other factors constant.

As I hinted at previously, healthcare is an extremely complicated and intertwined issue. No one factor determines the entire health of a country or its cost, Population dynamics, for example, is not only correlated with healthcare and who predominately needs it, but the economic development of a country as well. This is obviously an issue because economic development is also correlated with the quality of healthcare.

# 2 Literature Review

The wealth of research dedicated to this subject clarifies where some of these costs occur, but with any complex system question, there is no singular answer. Concerning for future price predictions of healthcare is that research papers going back to the early 70s have been studying the increasing cost of medical care. Most papers highlighting the increased care cost, from every decade they were published, mention that healthcare prices are once again rising. This is obviously concerning as it indicates that prices have been consistently rising for decades. This is a figure that is easily checked and can be referred to in Figure 1 (CMS Source). One of the most popular culprits for cost increases in the US is technological innovation in the medical field. As I will detail, even this is subject to debate. Other factors include admin costs, increasing population age, and others.

There are numerous papers that claim that technological cost and the increased capabilities of the healthcare sector are a major, if not the driving, factor to increased healthcare cost. This is the contention of Harvard economist Joseph P. Newhouse who wrote prolifically on the subject in the 1990s. Newhouse proposes that there is a trade off to containing healthcare costs and continuing healthcare innovation [5]. The argument runs along typical economic logical lines: if the opportunity to gain profit is lost in translation while controlling prices, then there will be less incentive to innovate. Some papers produce figures that the increase in prices are anywhere

from half to two-thirds of the total increase [2].

The two areas where the rate of costs have been increasing the fastest are in pharmaceutical drug prices and admin costs in private health insurance which both saw jumps from 11 to 16 percent in the 2000s [1]. Technological abilities increasing costs theoretically works with the increase price of pharmaceutical drugs, but not with the jump in increased admin cost. Moreover, Bodenheimer also argues that it is not technology, but technological diffusion that has some blame in the rising cost. There's two ways Bodeheimer argues that medical technology is diffused into the general populace, through direct to consumer advertisement (like with pharmaceutical drugs after 1988) and through specialists. The US has a greater proportion of medical specialists than most other countries. These specialists tend to advocate for bringing in new technologies at faster rates, and having better technologies can attract better specialists and hospital prestige. Bodenheimer claims this issue is much more prevalent in the United States than it is in Western Europe for example. This medical arms race is consequently raising the prices for medical care in the US. Unfortunately, there doesn't seem to be much introspection for the virtue of this technological expansion. Put another way, are patients coming out ahead of this quality of care to cost trade-off?

It's hard to measure this trade-off in a dynamic enough way to make an empirical decision. One reason is that when measuring "quality" in a hospital, most studies simply use mortality as the measure of quality. Mortality is obviously the most important measure but not the only option. More

research should try to obtain a more holistic reading on quality, especially quality of life after procedures. In the usual quality of care measurements that the data portray, we can see a direct relationship between quality of care and medical cost increasing. One study found that a one percent increase in quality led to a 1.3 percent increase in cost [4]. Interestingly, they state that the cost for deferring someones death in 1983 is 29,000 dollars.

Avoiding death for nearly thirty thousand dollars (in the 1990s) seems cheap for a human life. The obvious economic question is how can we continue to decrease that cost? Due to the multifaceted cost of healthcare it doesn't seem so clear. To be clear, we would have to have a stronger agreement and clearer picture of what's driving the cost. Specifically if healthcare can't be be mitigated at a structural level due to trade-offs of care, admin/pharmaceutical/insurance corruption, or important technological innovation then mitigation might be achieved personally. The cost of medical care still matters, but finding ways of avoiding medical costs might be more convenient for average patients. Meta studies on how different demographics affect personal healthcare spending are well established. Obesity is linked to a 68 percent increase in medication cost and a 36 percent increase in total healthcare cost [3]. This is a staggering price increase on a factor that is controllable for many people without underlying medical issues. Aggregating that cost from a patient to national level is just one goal of this paper. There are, in fact, many well documented papers that outline this aggregation. Their methods span from instrumental variables to other observational

studies. The scope of their work is much beyond this paper but will serve as a helpful comparison to evaluate my results.

# 3 Data

The data in this paper will come from multiple sources. The Centers for Medicare and Medicaid Services, a government agency, has an incredible data set charting national health expenditures since 1960. The data set not only charts total nation expenditure by the year, but subdivides it by department. Although not every department started by 1960, it details the expenditure from categories like "total out of pocket" to departments like Veteran Affairs, Indian Health Services, workers compensation, Maternal/Child Health, and more.

The data for education came from a mixture of the National Center for Educational Statistics (NCES) and the American Community Survey/Census. Much like the health data it doesn't come in annual figures until the 2000s. The age data came from the United Nations Data Portal which summarized US life expectancy from 1960-2021 very well.

Another data set I will be using is from the CDC. This data is the National Health and Nutrition Examination Survey (NHANES). This data is admittedly less neat than expenditure numbers. To start, there is no clean catalog of its numbers from NHANES I (starting in the 70s) to the current yearly survey the CDC conducts. In my final paper I will convert some of

these earlier data sets to have more data, but for this rough draft I will only include the continuous data from the 2000s to now because it is easily available. NHANES is fairly massive in its scope and the data comes from different areas. The NHANES has different sections from demographic, dietary, examination, to lab data. The Demographic and dietary are surveys and estimates on how many adults are smoking, their levels of activity, and even what they eat and obviously ethnicity and age. The examination data is from doctor examinations of patients that go from height, weight, blood pressure and more. For our purposes this is the most important segment of the NHANES data. The lab data is simply urine, blood, and other medical lab related data. The obesity rate came from these sources and the WHO has good data on predicted health outcomes before 2000 when NHANES is not yearly recorded. Exercise was going to be included in this analysis but there is no available data by the CDC before 2000 on this subject that I could find.

The final data set I will be implementing will be yet another CDC data set. This data set, the National Health Interview Survey (NHIS) has some overlap with the NHANES data. For example, the NHIS data also contains different data regarding average levels of activity. I found that the NHIS data is more convenient in measuring the number of adults who are meeting a minimum exercise criteria. The NHIS data also contains yearly data earlier than NHANES started implementing their yearly data. This inherently provides a greater number of figures to work with.

# 4 Methods

In this paper I will use some of the simple techniques we have learned throughout this course. The data has been cleaned and combined using R. All code used to generate these results will be from R. To estimate the effect obesity has on national healthcare expenditure I used the following methodology: create a linear regression model for healthcare expenditure and estimate the betas of many factors that might increase personal healthcare expenditure. I believe the linear regression will be the most efficient way to estimate this as I expect a linear relationship between the rate of obesity and healthcare expenditures and linear relationships for the other variables. As the population has higher rates of obesity, they will need more medical care for obesity-related illness.

$$Expenditure_T = B_0 + B_1 Obes_T + B_2 Age_T + B_3 Educ_T + B_4 Smoke_T + E_i \quad (1)$$

Where Expenditure is equal to the United State's total healthcare spending. "Obes" represents the percentage of US adults who are obese in the US. "Age" is the average life expectancy in the US. I chose life expectancy instead of average age for different reasons. Although the two measures are very closely related they are fundamentally different. People need healthcare throughout their life, but the distribution is somewhat bimodal, with the greatest amount of care coming early and late in ones life. It more or

less captures the full extent of possible dates one will be available to receive healthcare better than the average age would. "Educ" is the average years of education for Americans. "Smoke" is the percentage of smokers in the US. Finally, "Exer" is the percentage of Americans that meet the exercise requirement outlined in the NHIS data from the CDC.

In the data section I mentioned that many of the data sets do not work with yearly data. The NHANES data set, for example, did not become yearly until the 2000s. That's the trend with much of the health data I collected, but my expenditure data goes back into the 60s. Data is still available going to the 60s just not in a convenient yearly format to match my data. To fix this issue, I linearly imputed my data as they all had linear trends in one direction or the other. Some values couldn't be imputed in the early 60s so I cut that data from 1970-2018 for the best data results.

# 5 Findings

My findings come in three different charts. The first is the linear regression of my initial model, the others are the linear regressions of the same coefficients but instead of Total health expenditure, it is hospital hospital expenditures and out of pocket cost. All the figures are in millions of dollars. For example, a one percent increase in obesity correlated with a 144 million dollar increase in healthcare expenditures in the United States. Which is overall smaller than what most figures predict the total cost of obesity is on the healthcare

system. Smoking, obesity, and age all had positive correlations with increased healthcare cost, total hospital cost, and of of pocket cost. Education inversely was seen to decrease all of those associated costs. All my results were also found to be statistically significant except for my education values in all my models. I think that's an interesting result, but the literature also does indicate that education doesn't have a strong effect on healthcare spending. Overall, I believe my findings are relatively consistent with what most of the literature says even if my values are underplayed compared to their results.

# 6 Conclusion

There is no conclusive reason as to why healthcare costs are increasing. This is because it is likely a mixture of many factors, making numerous possible counterfactuals to argue over. From my research and others it does seem clear that certain demographic factors can increase healthcare expenditure and personal healthcare spending. It is therefore plausible that there are many specific practices that we can all do to reduce healthcare spending, even if we prefer more of it. This is especially important for those who cannot afford to go into medical debt or receive routine primary care checkups as other people can. More research should be done on these topics as it becomes a more politically divisive issue so that perhaps more clarity can be reached and more informed policy and societal decisions can be acted upon.

# 7 Bibliography

# References

[1] Thomas Bodenheimer. High and rising health care costs. part 1: Seeking an explanation. *Annals of Internal Medicine*, 142(11):847–854, 2005.

[2] Paul B. Ginsburg. High and rising health care costs. (16), 2008.

[3] Seamus Kent, Francesco Fusco, Alastair M. Gray, Susan A. Jebb, Benjamin J. Cairns, and Borislava Mihaylova. Body mass index and healthcare costs: a systematic literature review of individual participant data studies. *Obesity Reviews*, 18:869 – 879, 2017.

[4] Richard C. Morey, D J Fine, Stephen W. Loree, Donna L. Retzlaff-Roberts, and Shigeru Tsubakitani. The trade-off between hospital cost and quality of care: An exploratory empirical analysis. *Medical Care*, 30:677–698, 1992.

[5] Joseph P. Newhouse. An iconoclastic view of health cost containment. *Health Affairs*, 12(Suppl):152–171, 1993.

# 8 Figures

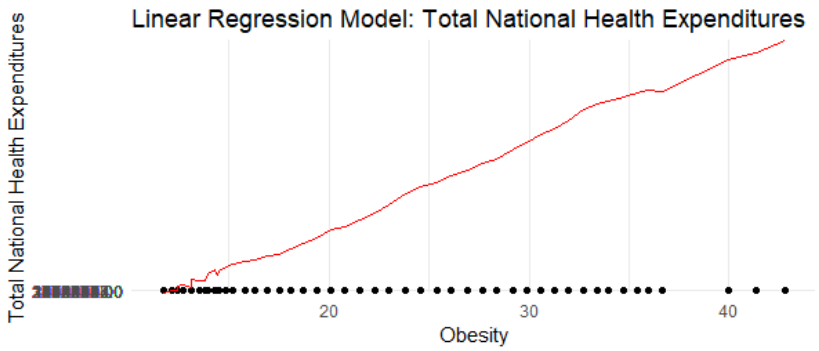


Figure 1: Total National Health Expenditures and Obesity Linear Regression

Table 1: Total Healthcare Expenditure

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -15335012.6611 | 5167252.7324 | -2.97 | 0.0048 |
| Obes | 144936.6986 | 17076.5620 | 8.49 | 0.0000 |
| Age | 177267.5959 | 78256.3761 | 2.27 | 0.0285 |
| Educ | -54665.2347 | 54306.5397 | -1.01 | 0.3196 |
| Smok | 40500.2660 | 18600.6773 | 2.18 | 0.0349 |

Table 2: Total Hospital Cost

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -5055061.5051 | 1910018.0509 | -2.65 | 0.0112 |
| Obes | 44811.4423 | 6312.1630 | 7.10 | 0.0000 |
| Age | 64787.9167 | 28926.6074 | 2.24 | 0.0302 |
| Educ | -28094.2414 | 20073.8142 | -1.40 | 0.1687 |
| Smok | 5651.6767 | 6875.5355 | 0.82 | 0.4155 |

Table 3: Out of Pocket Cost

| | Estimate | Std. Error | t value | Pr(>\|t\|) |
|---|---|---|---|---|
| (Intercept) | -326871.0729 | 95018.3625 | -3.44 | 0.0013 |
| Obes | 1727.7762 | 314.0135 | 5.50 | 0.0000 |
| Age | 4693.2041 | 1439.0225 | 3.26 | 0.0021 |
| Educ | -2459.2316 | 998.6193 | -2.46 | 0.0178 |
| Smok | 37.4562 | 342.0398 | 0.11 | 0.9133 |