Statistical Analysis of Health Care Spending in the United States

Avery Johnson
University of North Carolina at Charlotte

Abstract

The objective of this research is to understand why the health care spending is so much higher in the U.S. as opposed to other high-income countries. Many researchers have considered this problem and discussed the factors that contribute to the high health care costs in the U.S. For example, some recent articles identified the following major factors: high-volume, high-margin procedures; high salary for doctors; high prices for pharmaceuticals; and high administration expenses. Although the existing work evaluated many interesting variables related to health care and reported important findings, their conclusions were reached mainly based on simple descriptive statistics. In this project, we will investigate the problem of U.S. health care spending using more sophisticated statistical tools, in comparison with 10 other high-income countries. We will look into what variables are significantly associated with the health care costs in the U.S. using regression analyses. In particular, we will consider some variables related to health care spending, utilization and outcomes. This research will provide some insights on why the U.S. spends so much on health care and will suggest possible solutions to this problem.

1. Introduction

The U.S. health care system has been a controversial issue for a long time. Health care in the U.S. is exceedingly expensive and many people cannot afford it. As pointed out by Squires and Anderson (2015), although the U.S. does not have a universal health care system, it spends more public dollars on health care than some other high-income countries. The objective of this research is to understand why the health care spending is so much higher in the U.S. as opposed to other high-income countries. In order to get to the bottom of this question, we will utilize statistical tools or methods to look into what variables contribute to the high health care costs in the U.S., in comparison with 10 other high-income countries (Australia, Canada, Denmark, France, Germany, Japan, Netherlands, Sweden, Switzerland, and United Kingdom).

There is an extensive literature on the problem of U.S. health care spending. Among others, an interesting article published on Health Affairs, titled “Access, affordability, and insurance complexity are often worse in the United States compared to ten other countries”, presented some survey data on cost-related access to health care and affordability problems in the U.S. as well as 10 other high-income countries. The authors found that U.S. adults were significantly more likely than their counterparts in other countries to forgo care because of cost, to have difficulty paying for health care even when they are insured, and to encounter time-consuming insurance complexity (Schoen et al. 2013).

More recently, an article published on the Journal of the American Medical Association, titled “Health care spending in the United States and other high-income countries”, has attracted a lot of attention (Papanicolas et al. 2018). This article compared the health care spending and some related variables between the U.S. and 10 other high-income countries, and identified three major factors for the high health care costs in the U.S.: high salary for doctors, high prices for pharmaceuticals, and high administration expenses. Following Papanicolas et al. (2018), Emanuel (2018) further investigated the drivers of high health care costs in the U.S. and pointed out that these four areas - pharmaceuticals; high-volume, high-margin procedures; CT and MRI imaging; and administration - account for about two-thirds of the difference in health care costs between the U.S. and other developed countries.

Although these articles and some others considered many interesting variables regarding health care and provided insights on the problem of U.S. health care spending, they mainly utilized basic summary tables or charts when they presented their data and reached conclusions. The authors did not investigate the health care problem as thoroughly as they could have based on their data. This research is built upon these articles and will consider the same countries as in Papanicolas et al. (2018). We will evaluate the following variables
related to health care: health expenditure, number of MRI scanners/exams, number of CT scanners/exams, public/private/total health insurance coverages, pharmaceutical expenditure, number of physicians, number of doctor consultations, inpatient care average length of stay, tobacco/alcohol consumption, share of population with cancer, flu vaccination rates, and life expectancy. We will obtain the data on these variables from the Organisation for Economic Co-operation and Development (OECD) during the years 2000-2017 for the 11 high-income countries mentioned above. To improve upon the previous work and investigate the health care problem in greater depth, we will utilize more advanced statistical techniques to present and analyze the data.

We will visualize and analyze the data using the statistical software R. To compare the variables for different countries, we will employ the split-apply-combine technique as follows: first, split the data by country; second, for data on each country, draw a plot of the variable of interest (e.g., boxplot or time series plot); third, combine the plots for different countries into one graph for the ease of comparison. To investigate the association between the health expenditure and the other variables, we will draw scatterplots, run regression analyses, and perform hypothesis tests to evaluate the association. Similarly as above, we will employ the split-apply-combine technique to implement these procedures for different countries and summarize or combine the results in one figure.

The remainder is organized as follows. In Section 2, we will provide a more detailed description of our data. In Section 3, we will compare the health expenditures of the 11 high-income countries using boxplots and time series plots. We will also perform shapiro tests and draw QQ plots to check the normality of health expenditures which will be used as the response variable in following regression analyses. In Section 4, we will compare the other variables in the data set for the 11 high-income countries using boxplots. In Section 5, we will draw the scatterplots of health expenditure and the other variables for different countries. We will also run regression analyses by taking the health expenditure as the response variable and the other variables as predictors. In Section 6, we will summarize our key findings and make some remarks.

2. Data Description

Below is a detailed description of the variables that will be used in this project.

- **Health expenditure**: measures the final consumption of health care goods and services including personal health care (curative care, rehabilitative care, long-term care, ancillary services and medical goods) and collective services (prevention and public health services as well as health administration), but excluding spending on investments. This variable is measured as a share of Gross Domestic Product (GDP).

- **Number of MRI scanners**: Magnetic Resonance Imaging (MRI) is an imaging technique designed to visualise internal structures of the body using magnetic and electromagnetic fields which induce a resonance effect of hydrogen atoms. The electromagnetic emission created by these atoms is registered and processed by a dedicated computer to produce the images of the body structures. MRI units help physicians diagnose a range of conditions by producing images of internal organs and structures of the body. Unlike conventional radiography and CT scanning, MRI exams do not expose patients to ionising radiation. This variable is measured as the number of MRI scanners per million population.

- **Number of MRI exams**: Magnetic Resonance Imaging (MRI) exams help physicians diagnose a range of conditions by producing images of internal organs and structures of the body. This variable is measured as the number of MRI exams per 1,000 population.

- **Number of CT scanners**: A Computed Tomography (CT) scanner is an x-ray machine which combines many x-ray images with the aid of a computer to generate cross-sectional views and, if needed, three-dimensional images of the internal organs and structures of the body. They help physicians diagnose a range of conditions by producing images of internal organs and structures of the body. This variable is measured as the number of CT scanners per million population.

- **Number of CT exams**: Computed Tomography (CT) exams help physicians diagnose a range of conditions by producing images of internal organs and structures of the body. CT exams utilize X-rays to form images inside the body but are limited in detail compared to MRI when used for diagnosing soft tissue problems. This variable is measured as the number of CT exams per 1,000 population.
• **Government/social health insurance coverage**: the percentage of population covered with government-based insurance.

• **Private health insurance coverage**: the percentage of population covered with private insurance.

• **Total health insurance coverage**: the percentage of population covered with insurance.

• **Pharmaceutical expenditure**: measured as a share of total health spending.

• **Number of Physicians**: the number of physicians per 1,000 population.

• **Number of Doctor consultations**: the number of consultations patients have with doctors in a given year, measured per capita.

• **Inpatient care average length of stay**: measured in days.

• **Alcohol consumption**: measured in litres per capita for people aged 15+.

• **Tobacco consumption**: the percentage of population aged 15+ who are daily smokers.

• **Population with cancer**: measured as a share of total population with any form of cancer.

• **Flu vaccination rates**: Most people with influenza recover quickly, but elderly people and those with chronic medical conditions are at higher risk of complications and even death. This variable is measured as the percentage of population aged 65+ who have received an annual influenza vaccine.

• **Life expectancy**: measured in years.

As mentioned above, we obtained the data on these variables from the Organisation for Economic Co-operation and Development (OECD) during the years 2000-2017 for the 11 high-income countries (Australia, Canada, Denmark, France, Germany, Japan, Netherlands, Sweden, Switzerland, United Kingdom, and United States). More details on the data or variables can be found on the website of OECD.

### 3. Health Expenditure

In this section, we will compare health expenditures of the 11 high-income countries using boxplots and time series plots. We will also perform shapiro tests and draw QQ plots to check the normality of health expenditures which will be used as the response variable for regression analyses in Section 5.

The boxplots of health expenditures for the 11 high-income countries are shown in Figure 1. As seen from the boxplots, the U.S. has the highest health expenditure compared to the other countries. Figure 2 plots health expenditures over the years 2000-2017 for different countries. It shows that, compared to the other 10 high-income countries, the U.S. consistently spends the most on health care. In addition, the health expenditure as a share of GDP has been increasing over the years for all countries, and the increasing rate for the U.S. seems to be a little higher than the other countries.

![Figure 1: Boxplot for Health Expenditure](image-url)
Figure 3 presents the QQ plots of health expenditures for different countries. Table 1 gives the p-value of the Shapiro test for checking the normality of health expenditures for each country. From these results, we can see that health expenditures are approximately normally distributed for most of the countries considered, including the U.S. Thus, it is reasonable to perform the regression analyses which is to be done in Section 5.

## Table 1: Shapiro Tests for Checking the Normality of Health Expenditure

<table>
<thead>
<tr>
<th>Country</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.2700</td>
</tr>
<tr>
<td>Canada</td>
<td>0.0970</td>
</tr>
<tr>
<td>Denmark</td>
<td>0.0610</td>
</tr>
<tr>
<td>France</td>
<td>0.0260</td>
</tr>
<tr>
<td>Germany</td>
<td>0.0380</td>
</tr>
<tr>
<td>Japan</td>
<td>0.0017</td>
</tr>
<tr>
<td>Netherlands</td>
<td>0.0710</td>
</tr>
<tr>
<td>Sweden</td>
<td>0.0012</td>
</tr>
<tr>
<td>Switzerland</td>
<td>0.3900</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>0.0880</td>
</tr>
<tr>
<td>United States</td>
<td>0.1100</td>
</tr>
</tbody>
</table>

4. Predictors

In this section, we will compare the other variables in the data set for the 11 high-income countries using boxplots. In particular, we will consider the following variables: number of MRI scanners/exams, number of CT scanners/exams, public/private/total health insurance coverages, pharmaceutical expenditure, number of physicians, number of doctor consultations, inpatient care average length of stay, tobacco/alcohol consumption, share of population with cancer, flu vaccination rates, and life expectancy. The boxplots of these variables for different countries are given in Figures 4-19. We have also produced the time series plots of these variables. For the succinctness of presentation, we omit these plots and only make a few remarks about them at the end of this section.

From these boxplots, we can see that (i) the U.S. has relatively higher numbers of MRI scanners/exams and CT scanners/exams than the other high-income countries; (ii) the total health insurance coverage in the U.S. is only about 85% while all the other countries have close to 100% coverage; (iii) the physician density and number of doctors consolations in the U.S. are lower than most of the countries considered; (iv) the alcohol/tobacco consumptions in the U.S. are relatively lower than the other countries; (v) the share of population with cancer in the U.S. is higher than all the other countries except for Australia; (vi) the life expectancy in the U.S. is the shortest among the countries considered.

![Figure 4: Boxplot for the Number of MRI Scanners](image)

We also observe some interesting findings about Japan. The boxplots suggest that (i) Japan has the
largest numbers of MRI scanners/exams and CT scanners/exams among all countries considered; (ii) 100% of people in Japan are covered by health insurance, since Japan has the statutory health insurance system; (iii) the pharmaceutical expenditure as a share of total health spending is largest in Japan; (iv) Japan has the lowest physician density but highest number of doctors consolations; (v) the inpatient care average length of stay is longest in Japan, which is probably due to the population aging; (vi) the share of population with cancer is lowest in Japan; (vii) Japan has the longest life expectancy.

Figure 5: Boxplot for the Number of MRI Exams

Figure 6: Boxplot for the Number of CT Scanners

Figure 7: Boxplot for the Number of CT Exams
Although Japan has the largest numbers of diagnostic machines/exams such as MRI and CT, its overall health spending as a share of GDP is much lower than the U.S. In terms of health outcomes, Japan has much lower rate of cancer and much longer life expectancy than the U.S. It would be interesting to compare the U.S. health care system with that in Japan. There has been some discussion on this topic. For example, the Wall Street Journal published an article, titled “What Japan Can Teach Us About Health-Care Costs”, in 2013. We will make more remarks on the comparison of Japan and the U.S. in Section 6.
Regarding the health insurance, we notice from Figures 8-10 that (i) the U.S. has the lowest health insurance coverage (only about 85%, in contrast to virtually 100% for the other high-income countries); (ii) the main source of basic care coverage in the U.S. is the private employer-based and individual insurance, while the main insurance source for the other high-income countries is the government-based insurance; (iii) it is interesting to note that for some countries where all or most people are covered by government-based insurance (e.g., Australia, Canada, France, and Netherlands), there are still high percentages of people covered
by private insurance and this raises the question on the quality of public insurance in these countries.

Another interesting thing to note from Figure 11 is that the pharmaceutical expenditure as a share of total health spending in the U.S. is comparable to those in the other high-income countries. Papanicolas et al. (2018) pointed out that high prices on pharmaceuticals contribute to the high health care costs in the U.S. Their conclusion was reached by considering both pharmaceutical and health expenditures per capita.
Our analysis provides more insights on the relationship between pharmaceutical and health expenditures by considering the pharmaceutical expenditure as a share of total health spending.

![Boxplot for the Share of Population with Cancer](image1)

![Boxplot for Flu Vaccination Rates](image2)

![Boxplot for Life Expectancy (Years)](image3)

As mentioned above, we have also produced the time series plots of the variables considered. For the succinctness of presentation, we omit these plots and will only remark a few interesting findings. From those time series plots, we found that the numbers of MRI scanners/exams and CT scanners/exams, physician
density and life expectancy have generally been increasing over the years 2000-2017 for all of the 11 high-income countries. On the other hand, alcohol/tobacco consumption and inpatient care average length of stay have a trend of decline over the years for all countries considered. These findings indicate, to some extent, that the health care outcomes or population health for these high-income countries gradually improve.

It is also interesting to note that in the U.S., the numbers of CT and MRI exams began to decrease in 2010s. Driven by this observation, we found an article Rehman et al. (2018) that introduced the Choosing Wisely Campaign (CWC) which came into being in 2012. The objective of CWC is to encourage discussions between health care providers and patients regarding the care based on evidence, free of harm, duplicative or redundant tests/procedures that patients already received, and whether medications/tests/procedures are really necessary. The authors pointed out that although diagnostic tests or procedures are highly valued for decision-making, unnecessary testing creates harmful health services and an economic impact on the healthcare system. The CWC has spread widely throughout the world. One possible explanation for our finding is that CWC or other similar initiatives may affect the usage of diagnostic tests in the U.S.

5. Association of Predictors with Health Expenditure

In this section, we will draw and compare the scatterplots of health expenditure and the other variables for different countries. We will also run univariate linear regression analyses by taking the health expenditure as the response variable and each of the other variables as a predictor. In particular, we will combine the regression results (i.e., estimates of regression coefficients and p-values) for different countries into one figure for the ease of comparison. As shown in Section 2, health expenditures are approximately normally distributed for most of the countries considered, particularly for the U.S. which is our main interest.

For the succinctness of presentation, we will only include the scatterplots and regression results for the variables that have some interesting findings, including the numbers of MRI scanners/exams and CT scanners/exams, private health insurance coverage, pharmaceutical expenditure, physicians density, share of population with cancer and life expectancy. Note that for Japan and Sweden, there is only one data point for some variables and thus the scatterplots or regression results cannot be produced. See Figures 20-37.

From Figures 20-27, we can see that the numbers of MRI scanners/exams and CT scanners/exams are significantly positively associated with the health spending in the U.S. as well as the other high-income countries (except for Japan and Sweden, probably due to missing data). In particular, for the U.S., the p-values for testing the regression coefficients of these variables being 0 are all close to 0.

It is interesting to note from Figures 28-29 that the private health insurance coverage is significantly positively associated with the health spending in Australia, Canada, Denmark, France and Germany; and is significantly negatively associated with the health care spending in the U.S., Netherlands, and United Kingdom. This finding seems to make sense because the countries with positive regression coefficients have universal public insurance systems and the private health insurance is usually duplicative, supplementary or complimentary. For the U.S., the main source of basic care coverage is the private employer-based and individual insurance. It thus seems reasonable that in the U.S., the higher percentage of population covered by the private health insurance, the lower the health expenditure (the less cost out of pocket). For Netherlands, although there is a mandatory health insurance system, about 84% of the population is covered by the private health insurance and this probably explains the finding. For the U.K., the scatterplot looks nonlinear and thus the linear regression is not appropriate.

Based on Figures 30-31, the pharmaceutical expenditure as a share of total health spending is not significantly associated with the health spending in the U.S. (p-value = 0.84). Papanicolas et al. (2018) considered pharmaceutical and health expenditures per capita and concluded that high prices on pharmaceuticals contribute to the high health care costs in the U.S. Our finding adds additional knowledge to the relationship between pharmaceutical expenditure and health spending.

Figures 32-33 suggest that the physicians density is significantly positively associated with the health spending for most of the high-income countries considered (except for France and Netherlands, probably due to missing data). This seems to make sense because it is expected that the more physicians, the more spending on health care. It is interesting to note that the regression coefficient of the physicians density is highest for the U.S. among all countries considered. Probably, this is due to the fact that doctors in the U.S. have much higher salary than the other high-income countries (Papanicolas et al. 2018).

As shown in Figures 34-35, the share of population with cancer is not significantly associated with the
health spending in the U.S., although the association seems significant for some other countries (e.g., France and Japan). In addition, from Figures 36-37, we can see that the life expectancy is significantly positively associated with the health spending for all countries considered. Apparently, this makes sense because the longer people live, the more health care products or services are needed and thus the higher health spending.

6. Key Findings and Remarks
Below is a summary of our key findings in this project as well as some remarks:

• The U.S. spends much more on health care than the other 10 high-income countries, and the increasing rate of health spending for the U.S. seems to be a little higher than that for the other countries.
  – Here are the variables significantly associated with health spending in the U.S.: the numbers of MRI scanners/exams and CT scanners/exams, private health insurance coverage, physicians density, and life expectancy.
• The U.S. has more MRI and CT scanners and performs a higher rate of MRI and CT exams compared to the other high-income countries.
  – Japan has a larger number of MRI and CT scanners than U.S., but Japan spends much less than the U.S. on health care.
  – An MRI scan costs at least $1,500 in the U.S. while only around $100 in Japan due to cost control.
  – The high prices and high volumes of MRI and CT scans cost the U.S. much more than the other countries.
• The U.S. has the lowest coverage of health insurance (only about 85%, in contrast to nearly 100% for all the other countries considered).
  – The main source of basic care coverage in the U.S. is the private employer-based and individual insurance.
  – In contrast, for the other high-income countries, the main source is the government-based insurance.
• The U.S. has the shortest life expectancy among all the high-income countries considered, while Japan has the longest life expectancy.

References
Figure 20: Scatterplot of Health Expenditure vs Number of MRI Scanners

Australia

Canada

Denmark

France

Germany

Japan

Netherlands

Sweden

Switzerland

United Kingdom

United States

Figure 21: Regression Coefficient for the Number of MRI Scanners
Figure 22: Scatterplot of Health Expenditure vs Number of MRI Exams

Figure 23: Regression Coefficient for the Number of MRI Exams

Regression Coefficient

Australia

Canada

Denmark

France

Germany

Netherlands

Switzerland

United Kingdom

United States
Figure 24: Scatterplot of Health Expenditure vs Number of CT Scanners

Figure 25: Regression Coefficient for the Number of CT Scanners

Regression Coefficient

Australia  p = 0.00025
Canada  p = 6.9e−06
Denmark  p = 5.2e−06
France  p = 2.2e−08
Germany  p = 3.1e−06
Japan  p = 3.1e−06
Netherlands  p = 0.07
Sweden  p = 0.00015
Switzerland  p = 0.99
United Kingdom  p = 2.2e−06
United States  p = 7.3e−06
Figure 26: Scatterplot of Health Expenditure vs Number of CT Exams

Figure 27: Regression Coefficient for the Number of CT Exams
Figure 28: Scatterplot of Health Expenditure vs Private Health Insurance Coverage

Figure 29: Regression Coefficient for the Private Health Insurance Coverage
Figure 30: Scatterplot of Health Expenditure vs Pharmaceutical Expenditure

Figure 31: Regression Coefficient for Pharmaceutical Expenditure

p = 0.00017
p = 0.071
p = 8.5e−06
p = 3.5e−07
p = 0.24
p = 0.8
p = 6.4e−06
p = 5.3e−12
p = 0.037
p = 0.41
p = 0.84
p = 0.0017
p = 0.037
p = 0.0017
p = 0.037
p = 0.0017
p = 0.037
Figure 32: Scatterplot of Health Expenditure vs Number of Physicians

Figure 33: Regression Coefficient for the Number of Physicians
Figure 34: Scatterplot of Health Expenditure vs Share of Population with Cancer

Figure 35: Regression Coefficient for Share of Population with Cancer
Figure 36: Scatterplot of Health Expenditure vs Life Expectancy

Figure 37: Regression Coefficient for Life Expectancy