Title

- Examining connections between healthcare spending and the economic development of countries

Authors

- Bharath Jain, Prathik
- Jung, Junsoo
- Litostansky, Bridget
- Mehdi, Mohammad Razeenuddin
- Tatimakula, Akarsh Reddy
- Sarette, Stephanie

Abstract

A 3-5 summary of the paper. It should address the research question, the methods, and the conclusions of your analysis.

TBD: Until conclusion is written

Introduction

The definition of healthcare includes “the efforts to maintain, restore, or promote someone's physical, mental, or emotional well-being especially when performed by trained and licensed professionals (as in medicine, dentistry, clinical psychology, and public health)” (*Definition of HEALTH CARE*). According to the World Health Organization (WHO), health expenditure includes all spending on health services, nutrition activities, family planning, and emergency aid. It excludes any funding for sanitation and healthy drinking water. Total health expenditure as a percentage of Gross Domestic Product (GDP) indicates how much a country spends on its health based on its total wealth. The WHO suggests there is no recommended spending level on health, however, low-income countries will always benefit if more money is used toward health expenditures (*Health Expenditure*).

Global healthcare spending became a topic widely discussed in the media due to the COVID-19 pandemic, however, healthcare spending started in some high-income countries as early as the 1880s (*Ortiz-Ospina, Esteban, and Max Roser*). In 2021, national healthcare spending in the United States grew 2.7% totaling $4.3 trillion. Medicare spending increased by $734.0

billion, private healthcare spending increased by $1,211.4 billion, and out-of-pocket spending increased by $433.2 billion. The largest share of spending was sponsored by the federal government making up 34% of the total healthcare spending, followed by household spending at 27% for the year (*NHE Fact Sheet | CMS*).

Global health expenditures in the world have been increasing steadily at a slower rate over the last few decades. Higher-income countries have been ahead of the curve spending about two times as much as lower-income countries on healthcare spending. In low and middle-income countries, out-of-pocket spending is much higher and contributes to above 50% of the total expenditure in most cases (*Ortiz-Ospina, Esteban, and Max Roser*).

Using statistical modeling, this research paper aims to expand the understanding of the relationship between healthcare and the economy by offering insightful information about the intricate interactions between healthcare spending, quality of insurance, and their impact on economic development. The main goal is to examine the connections between healthcare spending, both at the public and private levels and its effects on a country's economic growth and development, with the assumption that a country's economy will improve with more healthcare spending. Additionally, the relationship between the quality of insurance coverage and the total amount of money spent on medical care will be investigated.

Data

The dataset was collected from WHO global health expenditure databases. Health spending indicators are key guides for monitoring the flow of resources, informing health policy development, and promoting the transparency and accountability of health systems.

Sources of Bias: WHO works collaboratively with member states (190 WHO member states since 2000) to update the database annually, using available information such as health account data, government expenditure records, and official statistics. Where necessary, modifications and estimates are made to ensure the comprehensiveness and consistency of the data across countries and years.

Bias: There is no data collected from the non-member states. Moreover, we are also assuming that reported data from each of the member nations is accurate.

Variable Code	Variable Name
country	Country name
code	Country code (ISO-3)

region	Region (WHO)
income	Country income group 2020 (World Bank)
year	Year
che_gdp	Current Health Expenditure (CHE) as % Gross Domestic Product (GDP)
che_pc_usd	Current Health Expenditure (CHE) per Capita in US$
che	Current Health Expenditure (CHE)
gghed	Domestic General Government Health Expenditure (GGHE-D)
pvtd	Domestic Private Health Expenditure (PVT-D)
dom_che	Domestic Health Expenditure (DOM) as % of Current Health Expenditure (CHE)
gghed_che	Domestic General Government Health Expenditure (GGHE-D) as % Current Health Expenditure (CHE)
gdp_usd2020_pc	Gross Domestic Product in constant (2020) US$ per capita.
oops_che	Out-of-pocket (OOPS) as % of Current Health Expenditure (CHE)
vpp_che	Voluntary Prepayments as % of Current Health Expenditure (CHE)
gghed_gdp	Domestic General Government Health Expenditure (GGHE-D) as % Gross Domestic Product (GDP)
gghed_gge	Domestic General Government Health Expenditure (GGHE-D) as % General Government Expenditure (GGE)

Methods

Data Cleaning Methods:

It was discovered that the year 2021 had many null and missing values during the exploratory data analysis. Any column with more than 30% of null values populated was dropped. If a column had missing values that pertained to an average then the average of the whole column was used to fill the null value. The column names of the dataset were already in a state that could be used for visualizations, so no editing was done to the column names. The cleaned dataset can be found using the following link: https://github.com/prathikbafna/STAT-5000-project-fall2023

Testing Data Authenticity:

The relative frequency distribution for a number's leading digit in a dataset is described by Benford's law. Smaller values for the first digits appear more frequently than bigger values. According to this rule, less than 5% of numbers begin with a 9, while about 30% of numbers begin with a 1. If the leading digits don't align with Benford's distribution, the data set may be considered unauthentic. Benford's Law was applied to the data set in R to validate the data. The observed frequency of each digit for the leading number in the Current Health Expenditure (CHE) column was calculated. The expected frequency for each digit using Benford's Law was

also calculated. A bar plot was created comparing the expected frequencies and observed frequencies for each digit (figure 1) to determine the authenticity of the data.

Hypothesis Testing:

Three hypotheses were tested to determine the relationship between a country's economy and healthcare spending.

Hypothesis 1

The first hypothesis tested determines if the final consumption expenditure of households and profit institutions serving households measured in constant 2020 international dollars (PPP) per capita is positively correlated with the economic development and overall well-being of a country. This suggests that as the income and prosperity of individuals and households increase, there is a corresponding rise in their spending on goods and services, ultimately contributing to higher final consumption expenditure.

> **Null Hypothesis (H0)**: There is no significant difference in the mean Final Consumption Expenditure of Households and Profit Institutions Serving Households (in constant 2020 international $ PPP per capita) across different countries.
> **Alternative Hypothesis (H1)**: There is a significant difference in the mean Final Consumption Expenditure of Households and Profit Institutions Serving Households (in constant 2020 international $ PPP per capita) across different countries.

An ANOVA test was implemented in R on the final consumption expenditure of households and profit institutions serving households measured in constant 2020 international dollars (PPP) per capita. The test calculated the p-value to determine the outcome of the hypothesis test. A Linear Regression model was used for further analysis. The calculated R-squared value explains the proportion of variance of the final consumption expenditure of households and profit institutions serving households across different countries. The F-statistic was also calculated in R to explain how well the overall model fit the data.

Hypothesis 2

The second hypothesis tested determines if better insurance is inversely related to healthcare expenditure by an individual.

> **Null Hypothesis (H0)**: There is no relationship between the amount spent on insurance and the amount of money spent out of pocket.
> **Alternate Hypothesis (H1)**: The higher the amount spent on insurance, the more the amount of money spent out of pocket.

Hypothesis 3

The third hypothesis tested was that countries with higher income spend a lot of money on insurance compared to other countries.

Null Hypothesis: There is no difference in the amount of money spent on insurance between high and low income countries
Alternative Hypothesis: Higher income countries spend more on insurance than low income countries

A sample of the data from 2017, 2018, and 2020 was taken from the original dataset to test this hypothesis. From this sample, the mean income for countries with high income was calculated and the mean income for countries with low income was calculated. The Welch Two-Sample T-test was run in R to compare the two means and determine if there was a significant difference between them.

Results

Figure 1

A data.frame: 9 × 4

Digit	Observed.first_digits	Observed.Freq	Expected
<int>	<fct>	<dbl>	<dbl>
1	1	0.30085384	0.30103000
2	2	0.19914616	0.17609126
3	3	0.12757408	0.12493874
4	4	0.09618282	0.09691001
5	5	0.07207433	0.07918125
6	6	0.06353591	0.06694679
7	7	0.04871924	0.05799195
8	8	0.04168759	0.05115252
9	9	0.04721246	0.04575749

First Digit Analysis vs. Benford's Law

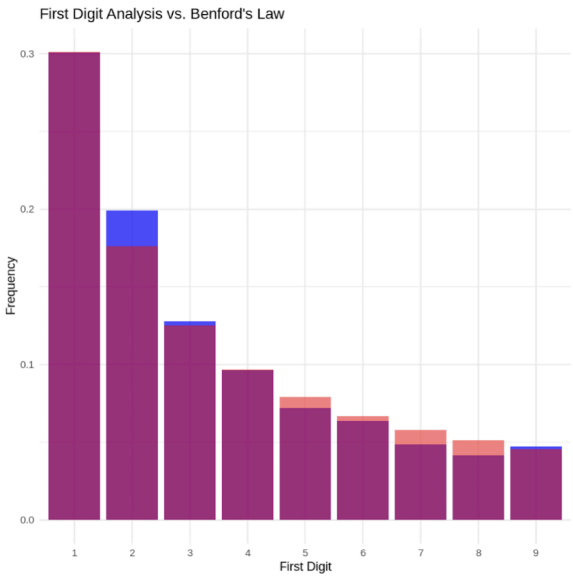


Figure 1. shows the testing results for the Benford Law test. The observed frequencies in the dataset are very similar to the expected frequencies by Benford's Law.

Figure 2

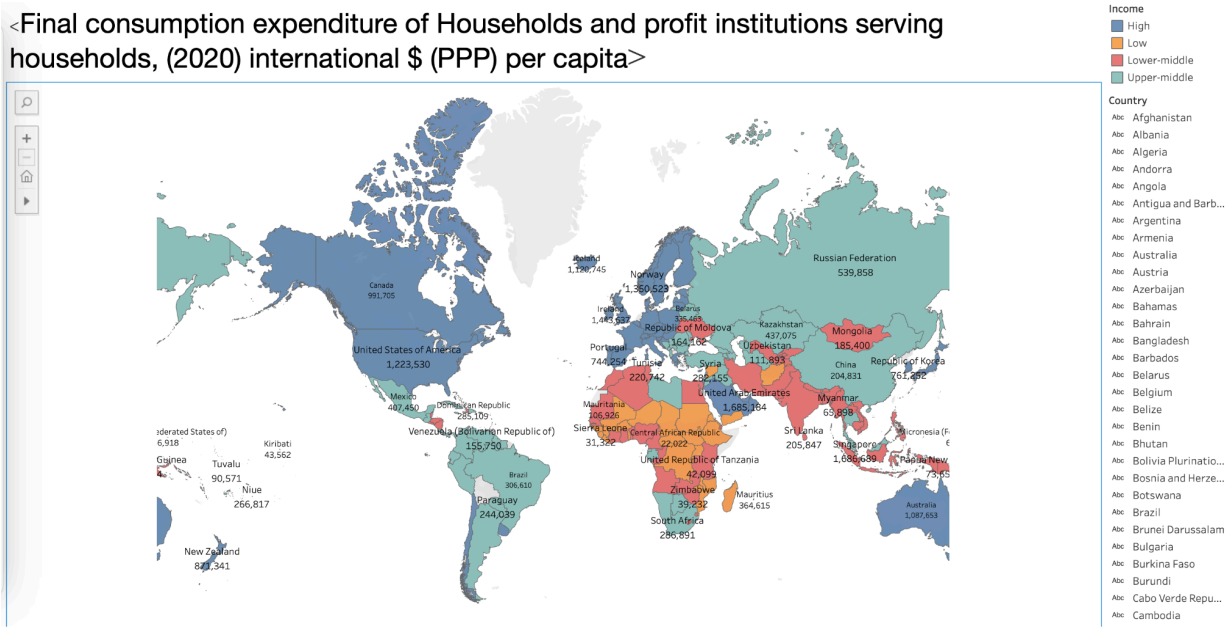


Figure 2 shows the income level for each country. It can be seen that North America, Australia, Western Europe, the United Arab Emirates, South Korea, Japan, and parts of South America are high income countries. Most of South America, Russia, China, Kazakhstan, Eastern Europe, and parts of Africa are upper-middle countries. Most of Africa falls in the lower-middle or low income category. India, Mongolia, Pakistan, Iran, and Indonesia fall in the lower-middle income categories. Afghanistan, Syria, Yemen, and Madagascar are all low income countries.

Figure 3

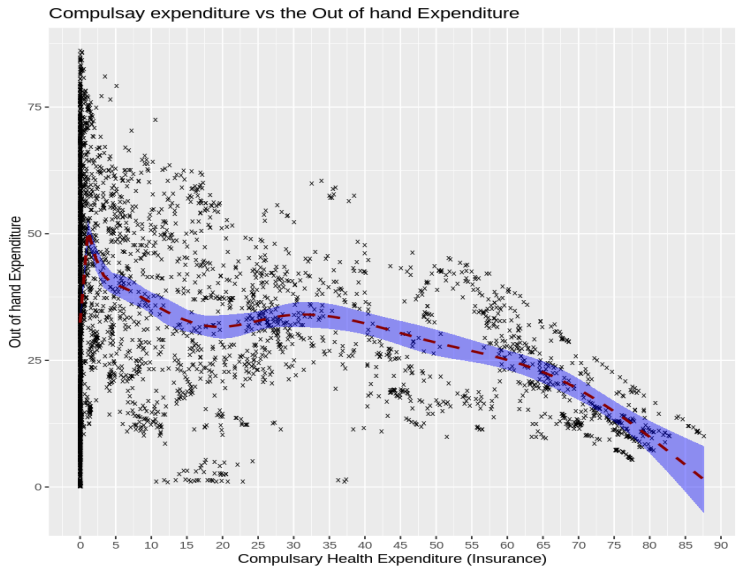


Figure 3 shows the relationship between compulsory health expenditure (insurance) and out of hand expenditure. There is a slight negative correlation.

Figure 4

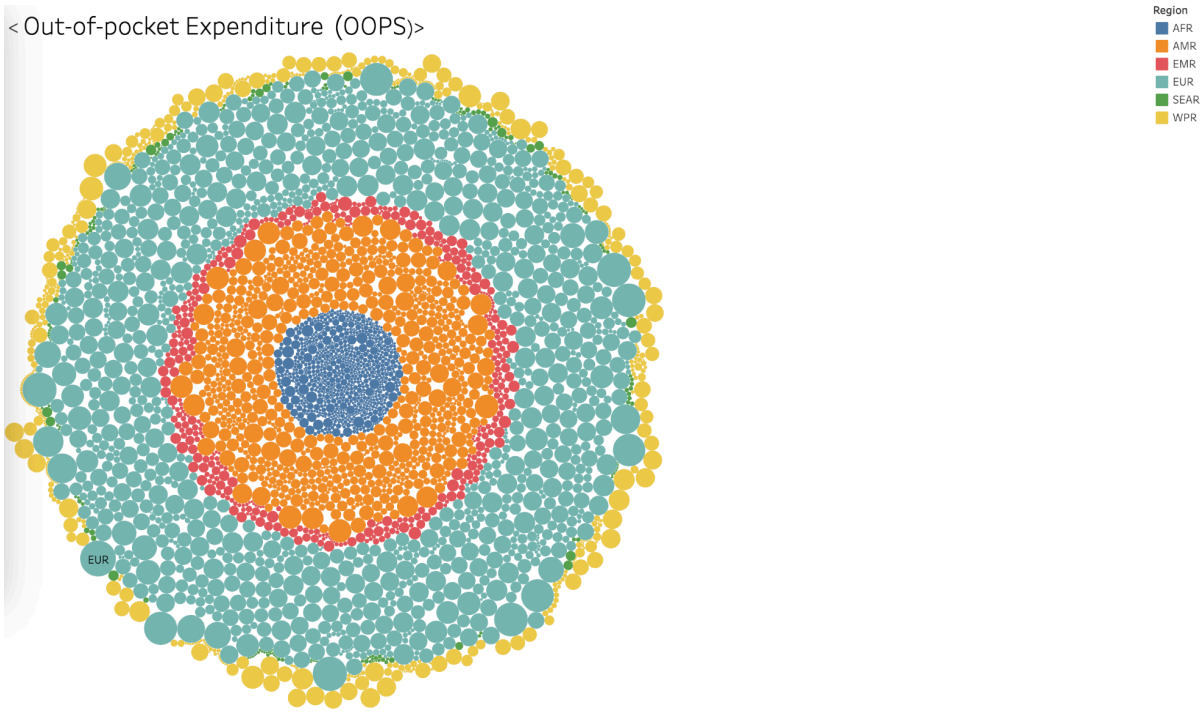
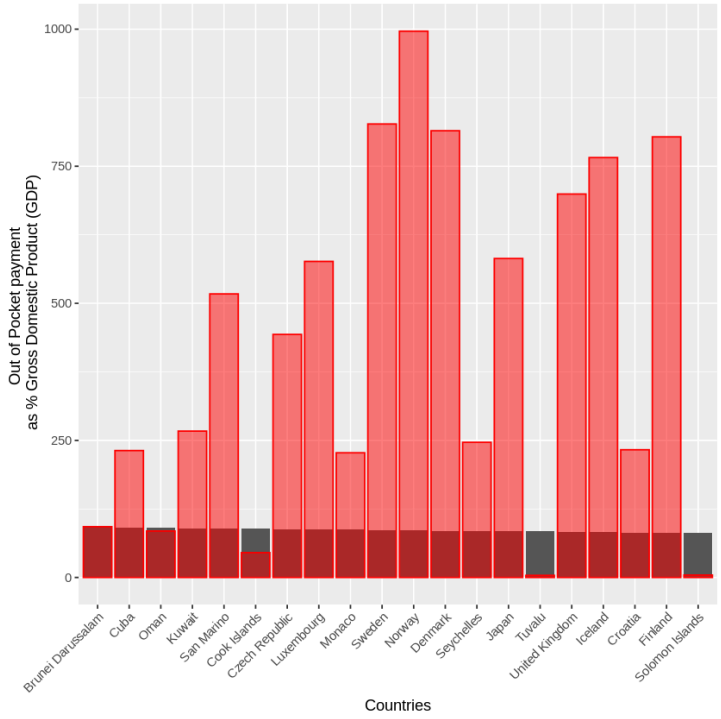


Figure 4 shows the out of pocket expenditure needed for each region. The figure reveals that Europe has the largest out of pocket expenditure, while Africa has the smallest.

Figure 5

Figure 5 shows how the gross domestic product and the final consumption expenditure of households and profit institutions serving households for the top 5 healthiest countries has changed from 2000 to 2020.

Figure 6



(Figure 9. gghed_che vs oops(red) of top 20 gghed countries)

Figure 6 shows the domestic general government health expenditure (GGHE-D) as percent of the current health expenditure (gray) compared to the out of pocket spending (red) of the top 20 GGHE-D countries. The figure reveals that out of pocket spending varies greatly compared to general government health expenditure.

Figure 7

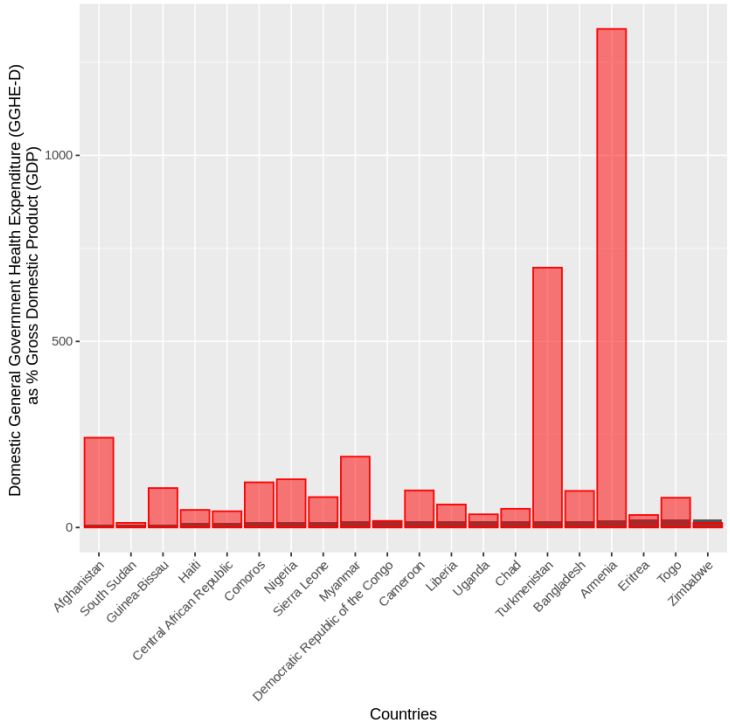


Figure 7 shows the domestic general government health expenditure (GGHE-D) as percent of the current health expenditure (gray) compared to the out of pocket spending (red) of the bottom 20 GGHE-D countries. The figure reveals that out of pocket spending is almost always more than the domestic general government health expenditure.

```
Residual standard error: 2698 on 3814 degrees of freedom
  (218 observations deleted due to missingness)
Multiple R-squared:  0.9462,    Adjusted R-squared:  0.9435
F-statistic: 351.2 on 191 and 3814 DF,  p-value: < 2.2e-16
```

The coefficient table shows the estimated coefficients for each level of the "country" variable. The coefficient for "countryAlbania" is 7070.30. This suggests that, on average, the Final Consumption Expenditure in constant (2020) international $ (PPP) per capita for Albania is estimated to be 7070.30 higher than the intercept value.

Conclusions

- Summarize what the paper has done, and discuss the implications of your Results.
- Explicitly connect the results to the research question.
- Discuss how you would you extend this research

https://colab.research.google.com/drive/193thKjY5uv-6uSXrZdq8233UwfS8mCIK?usp=sharing

Citations

.bib file has been created

#final merged code?

https://colab.research.google.com/drive/1PBbMGmHgcNJpmdmlu-ELDCZRQVxKzTAY?usp=sharing