

Healthcare Expenditure versus Outcomes: A Comparative Analysis

Sabina Bimbi, Terence Mpofu, Suphanat Juengsprasertsak, Connor Donahue

*College of Computing
Michigan Technological University
Houghton, MI, USA*

Abstract—There are vast global disparities in health outcomes. Generally, wealthier nations tend to have lower mortality rates and greater life expectancies than poorer, less developed nations. The intuitive hypothesis is that the wealthier nations have better outcomes because they spend more on hospitals, infrastructure, and medical research. This study set out to test that hypothesis using correlation and linear regression models to quantify the strength and magnitude of the relationships between per-capita healthcare expenditure and metrics tracking total mortality and health-related mortality. Expenditure had a strong negative correlation with both metrics, and linear regression resulted in R^2 values of 0.452 and 0.461 for the two mortality metrics evaluated. These results indicate that increased expenditure is associated with better health outcomes, but that most of the variation in mortality is dependent upon other factors. The study also identifies nations with significant outlying values, which can be targeted in case studies aiming to determine other sources of the variation in health outcomes.

Index Terms—healthcare, linear regression, correlation, data science

I. INTRODUCTION

Healthcare systems differ vastly among countries throughout the world. These differences are often manifested in both the costliness of healthcare systems and the medical outcomes of the general populations. This study aims to analyze and compare healthcare expenditure data with key indicators of overall population health outcomes using a combination of World Health Organization (WHO) global health indicators. The study will involve a comparative analysis between healthcare spending and mortality rates among countries across the globe with different healthcare systems.

Understanding how healthcare spending relates to population health metrics is a necessity for policymakers guiding healthcare system reform. It is important for policymakers to understand how the population will be affected by changes in healthcare spending so that resources can be more optimally distributed. The findings of this research will inform national leaders in private and public healthcare systems of this effect so that they may create evidence-driven policies and investment plans which maximize system performance while being cost-effective.

II. RELATED WORKS

This topic has been the subject of prior studies; however, they have often been limited in scope by analyzing only a subset of nations. For example, John Nixon and Philippe

Ulmann performed an econometric comparison of per capita spending with various mortality measures of countries in the European Union, finding that increased spending is strongly related to decreased infant mortality and also has a marginal positive impact on life expectancy [1].

Similarly, Jaison Chireshe and Matthew Ocran analyzed data from 45 countries in Sub-Saharan Africa, finding that increases in health care expenditure as measured by total health care expenditure per capita and public health care expenditure to GDP is related to a decline in childhood mortality rates [2]. They also observed that increases in total health care expenditure per capita is associated with an increase in life expectancy.

The majority of the current literature on this subject focuses on governmental spending, rather than overall health care expenditure: Bokhari et al. performed an econometric study to calculate elasticity values that indicate the strength of relationships between two variables [3]. They found that increased government spending in healthcare was strongly associated with decreased infant and maternal mortality rates in a dataset containing 127 countries.

All of the reviewed literature lends credence to the hypothesis that increased spending results in better health outcomes; however, this study aims to fill the current gap in the literature by evaluating total per capita spending on a global scale, addressing the limited scope and governmental focus of these other studies.

III. METHODS

A. Data Loading and Pre-processing

The data consists of the following three metrics, taken from the World Health Organization (WHO):

- 1) Predictor variable: Total per capita expenditure on healthcare in USD, by country and year [4].
- 2) Outcome variable 1: Adult mortality rate as the number of deaths between 15 and 60 years per 1000 population (grouped by country and year) [5]. This metric will be referred to as "total mortality."
- 3) Outcome variable 2: Probability of dying between the exact ages 30 and 70 years from cardiovascular diseases, cancer, diabetes, or chronic respiratory diseases (grouped by country and year) [6]. This metric will be referred to as "health-related mortality."

The data for each metric was joined into one table and combined on the country name and year. Any countries that were missing data for one or more of the entries were dropped. After merging the three World Health Organization (WHO) datasets, the final sample contained 183 countries out of an initial 196. Thirteen countries (Puerto Rico, Democratic People's Republic of Korea, Saint Kitts and Nevis, Palau, Nauru, Monaco, Andorra, Dominica, Cook Islands, Niue, Marshall Islands, San Marino, and Tuvalu) were dropped due to missing data. The excluded countries would likely be outliers in the analysis because their remote locations (many island nations), small populations, or lack of data transparency (e.g DPRK) present confounding variables. Due to their lack of available data, however, it is not possible to truly classify them as outliers. Therefore, it is judged acceptable to exclude them from the analysis. Furthermore, only data from 2019 was included in the dataset because it was the most recent year for which all metrics were reported before the Covid-19 Pandemic distorted some of the data. There were not missing values after these preprocessing steps, so no imputation or further processing of the data was required.

B. Linear Regression and Correlation Analysis

The main analysis focuses on examining the relation between healthcare per-capita expenditure and both mortality rates, which consists of two groups that will be compared. The first step is to explore the distributions of mortality rates and find statistical values including the mean, standard deviation, minimum and maximum values, and so on. The relationships between per-capita expenditure and both mortality rates is visualized in scatter plots to help explore the types of relationships the variables may have.

After exploring the distributions and relationships, an ordinary least squares (OLS) linear regression model is applied to create two regression models using the natural log of healthcare expenditure as the predictor and each mortality metric as the outcome.

Residual analysis is performed for both linear regression models to evaluate whether the model assumptions are satisfied and to identify outliers which are far away from the fitted line. The residuals are examined through the z-test, with outliers identified where $|z| > 2$.

Following this step, a correlation analysis is conducted to assess the relationship between the expenditure metric and each mortality metric, lending further insights into the strength of the relationships evaluated by the regression models and clarifying how closely related the total mortality and health-related mortality metrics are. Spearman's rank correlation is used instead of Pearson's correlation coefficient because the mortality rate data is not normally distributed.

C. Data Visualization

In this project, multiple visualization tools are used to support data interpretation. These include density plots of the mortality rates and healthcare expenditure per capita, as well as scatter plots illustrating how mortality rates change

as expenditure increases. This helps reveal general patterns in the relationships. After the linear regression models are fitted to the data, the regression lines will be superimposed upon these scatter plots. Residuals will be shown in a scatter plot as well to highlight outliers. Finally, a correlation matrix is also included to summarize the strength and direction of associations among all variables.

IV. RESULTS

A. Exploratory Data Analysis

The investigation of the data set began with an evaluation of the distributions of the healthcare spending, total mortality, and health-related mortality metrics. Figures 1 through 3 contain density plots of the metrics.

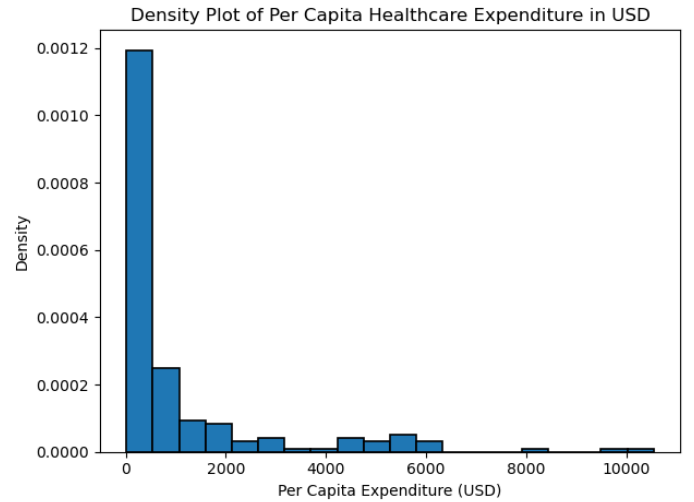


Fig. 1. Density plot of healthcare spending per capita (USD).

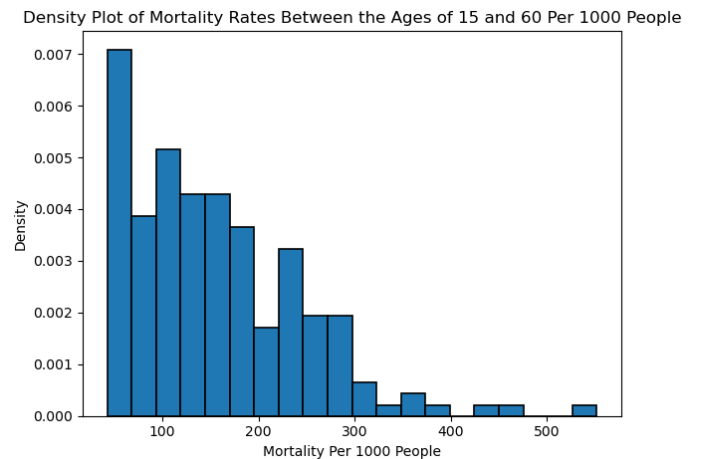


Fig. 2. Density plot of total mortality per 1000 people Age 15-60.

As seen in Figure 1, the distribution of per-capita spending on healthcare is not normal. However, a logarithmic transformation will likely result in a much more normal distribution. This distribution reveals that the vast majority of countries

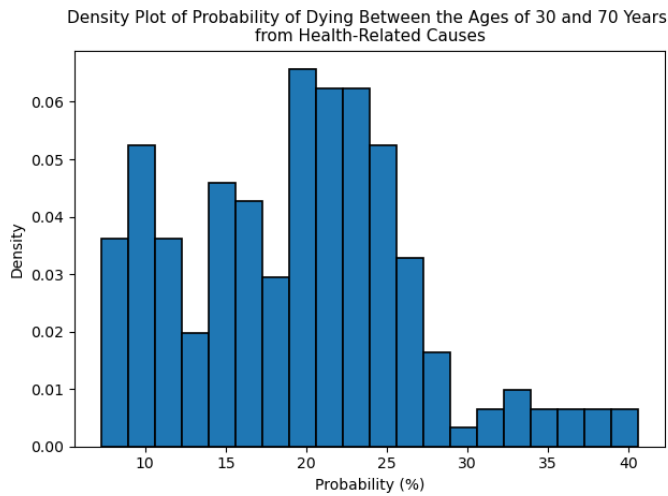


Fig. 3. Density plot of health-related mortality ages 30-70 (probability).

spend between \$0-\$1000 on healthcare per capita, while very few countries spend more than that and a few outlying countries spend as much as \$10000 per capita.

The distribution of total mortality in Figure 2 is strongly right skewed, meaning that most countries have comparatively low to moderate adult mortality, whereas a minority of countries experience extremely high adult mortality, therefore pulling the distribution to the right. Due to the strong skew of this distribution, it is not considered normal.

The distribution of health-related mortality is found In Figure 3. This distribution is more evenly distributed about the mean, but still far from normal and somewhat right skewed as well. Many countries cluster around 18–25% mortality, meaning about 1 in 4 to 1 in 5 adults die between the ages of 30 and 70 from the major non communicable diseases (NCDs) represented by the data. There are some countries with lower risk, 8–12%, and some with higher risk, 30–40%, but the tail is not as extreme as for total adult mortality. This suggests that while NCD burden differs, it's somewhat more “compressed” than overall adult mortality, which also includes injuries, accidents, infectious diseases, and other causes.

Because the main portion of this analysis involves correlation and linear regression, it is important to explore and categorize the relationships between the predictor variable (expenditure) and both outcome variables (total mortality and health-related mortality).

The scatter plots in Figure 4 and Figure 5 display these relationships before transforming the spending data. Figure 4 demonstrates that, as per-capita health expenditure increases, total mortality drops sharply at low ranges of spending, then levels off. This suggests that gains from extra spending have diminishing returns and may have no noticeable impact beyond a certain threshold. A similar pattern is observed in Figure 5, where health-related mortality is plotted against per-capita spending.

The above relationships resemble the negative natural log-

Scatter Plot of Mortality Rates Between the Ages of 15 and 60 Per 1000 People vs. Per Capita Healthcare Expenditure

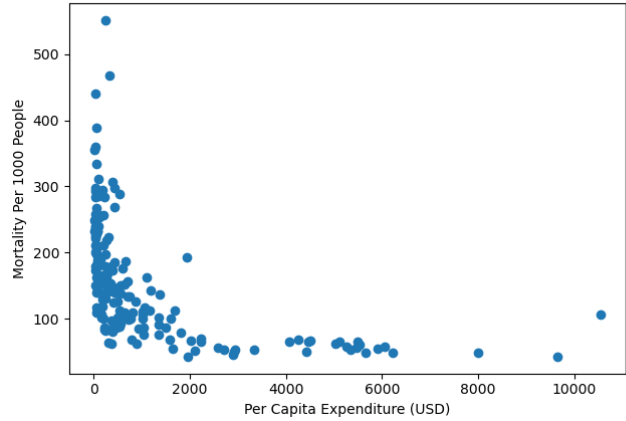


Fig. 4. Scatter plot of total mortality versus spending.

Scatter Plot of Mortality Probability Between the Ages of 30 and 70 from Health-Related Causes vs. Per Capita Healthcare Expenditure

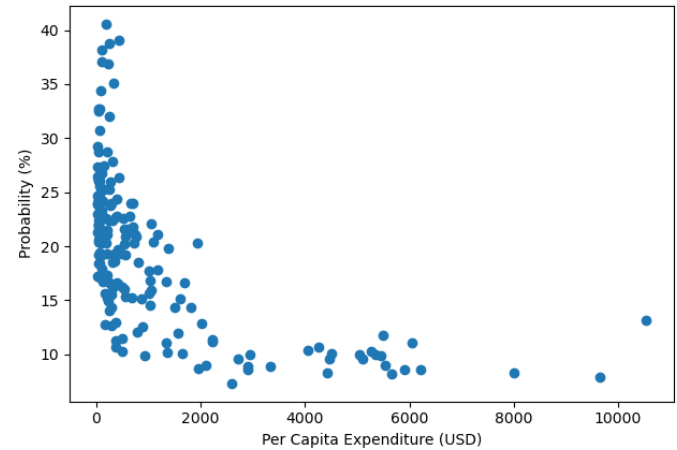


Fig. 5. Scatter plot health-related mortality versus spending.

arithm function, so the relationships of both mortality metrics were also plotted against the log of healthcare expenditure in Figure 6 and Figure 7. The relationship between each mortality metric and the log of healthcare expenditure is inverse and fairly linear, which is a more feasible relationship on which to fit a linear regression model. The models in the following analysis will be fit on this relationship.

B. Linear Regression

Using ordinary least squares linear regression to fit health-related mortality on the natural log of per-capita healthcare expenditure resulted in a slope of -2.96 and intercept of 36.60 (Figure 8). Here, a one-unit increase in log of health expenditure is associated with an average decrease of 2.96% in the probability of dying from cardiovascular disease, cancer, diabetes, or chronic respiratory disease between ages 30 and 70. The slope is highly significant ($p = 2.02e-25$) and the R^2 of 0.452 indicates that 45.2% of the variation in health-related mortality is accounted for by health spending alone.

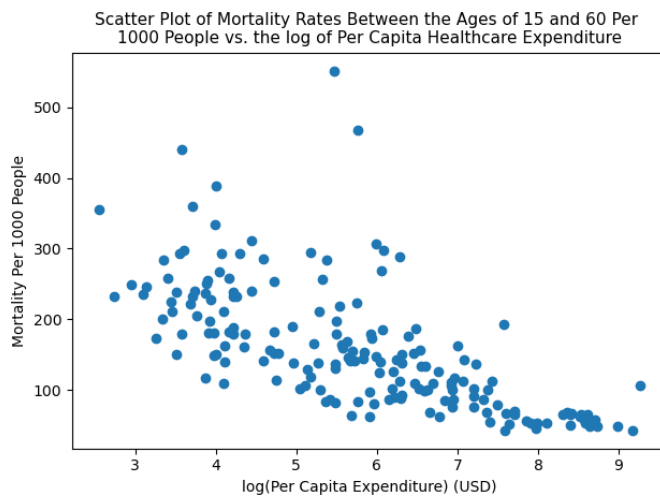


Fig. 6. Scatter plot of total mortality versus log of spending.

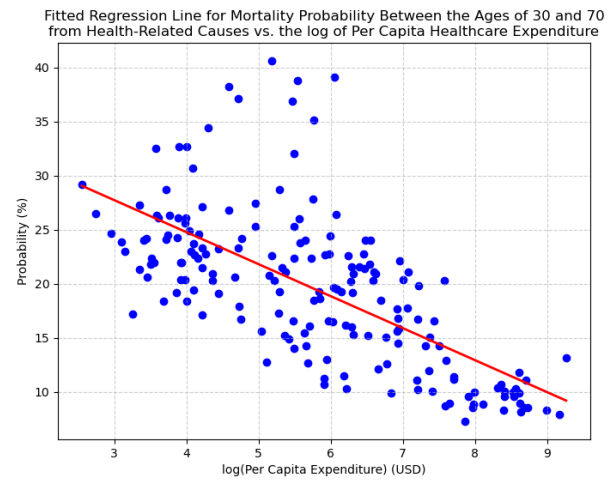


Fig. 8. Scatter plot of health-related mortality versus log of spending with fitted regression line.

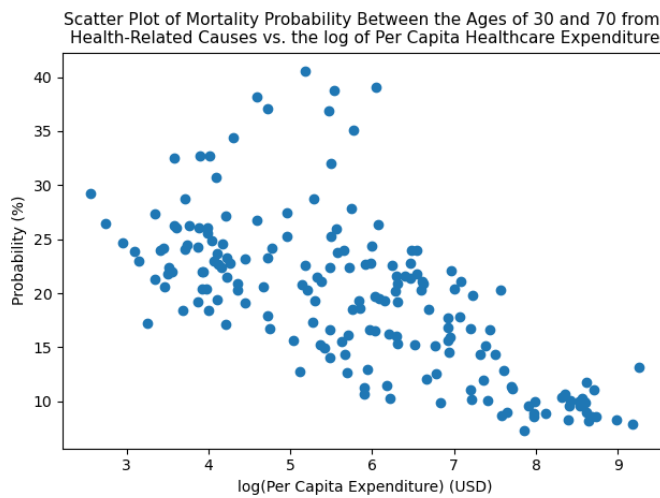


Fig. 7. Scatter plot of health-related mortality versus log of spending.

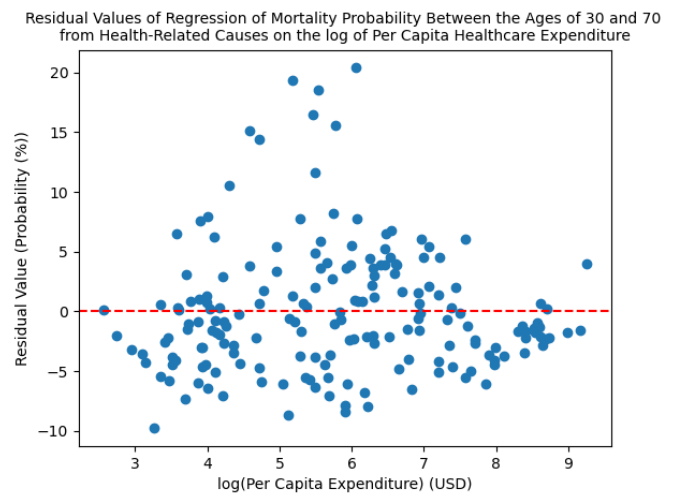


Fig. 9. Scatter plot of residuals from regression of health-related mortality on log of spending.

The residuals of the regression model, plotted in Figure 9, are approximately centered around zero with no obvious trend across log expenditure values, supporting the assumptions of the linear regression model. Z-score analysis revealed a set of outliers, including Eswatini, Lesotho, Micronesia, Fiji, Samoa, Kiribati, Vanuatu, and Solomon Islands – all of which exhibit higher health-related mortality than expected given their reported healthcare expenditure.

Using ordinary least squares linear regression to fit total mortality on the natural log of per-capita healthcare expenditure resulted in a slope of -37.0 and intercept of 371.6 (Figure 10). Here, a one-unit increase in log of health expenditure is associated with an average decrease of 37 deaths per 1000 people between the ages of 15 and 60. The slope is highly significant ($p = 4.20e-26$) and the R^2 of 0.461 indicates that 46.1% of the variation in total mortality is accounted for by health spending alone.

The residuals of the regression model, plotted in Figure 11, are approximately centered around zero with no obvious trend across log expenditure values, supporting the assumptions of the linear regression model. Z-score analysis revealed a set of outliers, including South Africa, Botswana, Namibia, Eswatini, Lesotho, Zimbabwe, and Central African Republic – all of which exhibit higher total mortality than expected given their reported healthcare expenditure.

C. Spearman Correlation

A Spearman correlation was performed on the relationships between healthcare expenditure and each mortality metric (Figure 12). Because this is a rank-based test, there was no need to transform the expenditure data. Expenditure and total mortality had a correlation coefficient of -0.78 ($p \ll 0.01$), while health-related mortality correlated with expenditure had

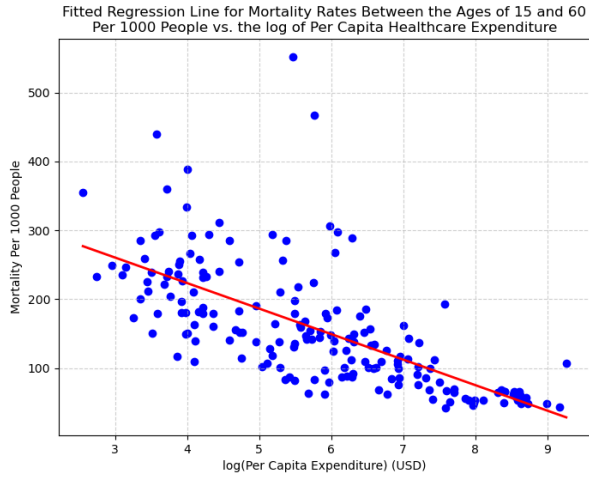


Fig. 10. Scatter plot of total mortality versus log of spending with fitted regression line.

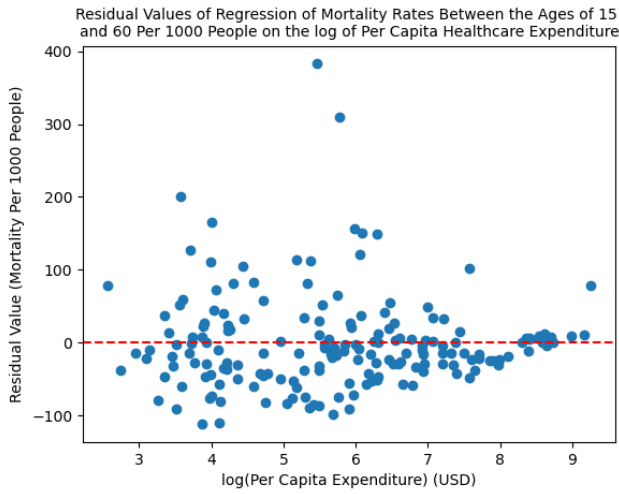


Fig. 11. Scatter plot of residuals from regression of total mortality on log of spending.

a coefficient of -0.71 ($p < 0.01$). Both correlations are strong, negative, and highly significant, confirming that countries that spend more per person on healthcare tend to have lower total mortality and lower health-related mortality. The correlation is slightly stronger for total mortality than for health-related mortality, which is consistent with the slightly higher R^2 obtained in the second regression model. Additionally, the two mortality metrics had a strong positive correlation, with a coefficient of 0.86 ($p < 0.01$).

V. CONCLUSION

This project set out to examine the relationship between healthcare expenditure per capita and two critical health outcome indicators, which are adult mortality for ages 15 to 60 (total mortality) and health specific mortality for ages 30 to 70 from major noncommunicable diseases (health-related

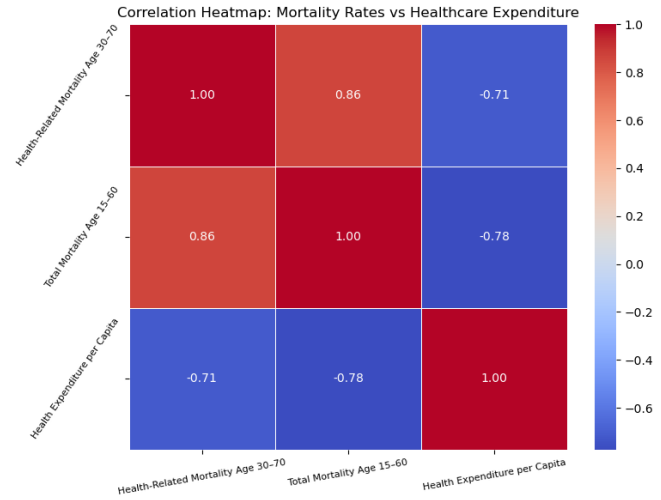


Fig. 12. Correlation heatmap of all three metrics.

mortality). The data tracked 183 countries, and the analysis was performed using data from 2019. By integrating three global health datasets and using linear regression and correlation analysis, it was found that higher national healthcare spending corresponds to better measurable health outcomes.

The global distributions of all three indicators were markedly non normal, with strong right skewness in expenditure and considerable variability in mortality outcomes. This justified the use of non-parametric Spearman correlation rather than parametric approaches. The results demonstrated a strong and statistically significant negative correlation between healthcare expenditure per capita and both total mortality and health-related mortality. These findings support the intuitive expectation that countries investing more financial resources into their healthcare systems tend to achieve lower mortality burdens. However, the strength of these correlations varied, with expenditure showing a stronger association with total mortality than with health-related mortality. While this result is unexpected, it may be due to the differences in the distributions of both mortality metrics because the stronger skew of the total mortality data mirrors the distribution of healthcare expenditure much more closely than health-related mortality. This may result in a stronger correlation for nonparametric tests such as Spearman's correlation.

In the regression analysis, the explained variance of the models were 45.2% and 46.1%, which is high but still less than half of the total variation in the data. Visualizations of the models, including scatterplots with the fitted regression lines, further highlighted the presence of substantial variability across different nations. Even among nations with comparable spending levels, outcome differences were evident, emphasizing that healthcare expenditure alone is not a perfect predictor of health system performance. These patterns point to the influence of additional variables such as healthcare access, governance efficiency, population demographics, and disease specific risk factor prevalence. For example, some countries

with relatively modest per capita spending still demonstrated strong outcomes, possibly reflecting effective public health policy, primary care strength, or environmental and behavioral health determinants.

The linear regression analysis also had unexpected results because the regression with total mortality had greater explained variance (46.1%) than the regression with health-related mortality (45.2%). While these two metrics are closely related (correlation of 0.86), healthcare spending is expected to have a greater impact on deaths caused by noncommunicable disease than total deaths. Further consideration, however, presents reasonable explanations for the observed discrepancy. Perhaps the deaths related to sudden injury and virulent disease, which fall under total mortality but not health-related mortality, actually have a stronger correlation with expenditure because they are highly dependent upon resources such as emergency medical response systems and widespread vaccination programs, whereas noncommunicable diseases are often the result of genetic causes.

The analysis of outlier residuals adds an additional important element to the study as well. For each of the regression models, the outlying residuals were predominantly from Sub-Saharan Africa and Oceania. All of the identified residuals had significantly higher mortality rates than the models predicted. Some potential causes of this inefficiency may be weak preexisting infrastructure or government corruption that results in mismanagement of the money that is reportedly spent on healthcare. Specific to Oceanic nations, one potential cause of high expenditures and poor outcomes is the large overhead costs of shipping medical equipment and outsourcing healthcare professionals from other nations. Future studies could verify these causes and identify new ones by investigating each outlying nation on a case-by-case basis.

The most notable limitation of the study is that healthcare expenditure was treated as a single aggregated metric, though its internal composition can vary dramatically between nations with regards to factors such as administrative costs, preventive care investment, and the relative share of private versus public spending. Furthermore, socioeconomic variables such as GDP per capita, education levels, and inequality were absent from the dataset but are known to interact strongly with health outcomes. Including such variables in future models could provide a more nuanced understanding of health system performance.

In summary, this study concludes that higher healthcare expenditure per capita is associated with better health outcomes globally, particularly lower total mortality and reduced probability of health-related deaths from major noncommunicable diseases. However, expenditure alone does not fully explain global health disparities. The findings highlight the need for more comprehensive multivariable models and deeper exploration of how healthcare systems allocate resources, not just how much they spend. Future research could build on this foundation by incorporating longitudinal data, expanding socioeconomic predictors, and exploring regional clustering to better understand structural health inequalities worldwide.

This analysis contributes to the broader evidence base demonstrating that investment in healthcare matters, but efficiency, policy context, population demographics, and environmental factors play equally essential roles. By recognizing both the value and limitations of expenditure-based assessment, policymakers and researchers can better develop targeted and cost effective strategies to improve population health outcomes on a global scale.

REFERENCES

- [1] J. Nixon and P. Ulmann, "The relationship between health care expenditure and health outcomes: Evidence and caveats for a causal link," *Eur. J. Health Econ.*, vol. 7, no. 1, pp. 7–18, Mar. 2006.
- [2] J. Chireshe and M. K. Ocran, "Health care expenditure and health outcomes in sub-Saharan African countries," *Afr. Dev. Rev.*, vol. 32, no. 3, pp. 349–361, 2020.
- [3] F. A. S. Bokhari, Y. Gai, and P. Gottret, "Government health expenditures and health outcomes," World Bank, Washington, D.C., Working Paper, May 2006.
- [4] World Health Organization, "Current health expenditure (CHE) per capita in US\$," *WHO Global Health Observatory*, 2025. [Online]. Available: [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/current-health-expenditure-\(che\)-per-capita-in-us-dollar](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/current-health-expenditure-(che)-per-capita-in-us-dollar) [Accessed: Oct. 17, 2025].
- [5] World Health Organization, "Adult mortality rate (probability of dying between 15 and 60 years per 1000 population)," *WHO Global Health Observatory*, 2024. [Online]. Available: [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/adult-mortality-rate-\(probability-of-dying-between-15-and-60-years-per-1000-population\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/adult-mortality-rate-(probability-of-dying-between-15-and-60-years-per-1000-population)) [Accessed: Oct. 17, 2025].
- [6] World Health Organization, "Dying between the exact ages 30 and 70 years from cardiovascular diseases, cancer, diabetes, or chronic respiratory diseases, probability (SDG 3.4.1)," *WHO Global Health Observatory*, 2024. [Online]. Available: [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/probability-of-dying-between-exact-ages-30-and-70-from-any-of-cardiovascular-disease-cancer-diabetes-or-chronic-respiratory-\(-\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/probability-of-dying-between-exact-ages-30-and-70-from-any-of-cardiovascular-disease-cancer-diabetes-or-chronic-respiratory-(-)) [Accessed: Oct. 17, 2025].

VI. SELF-DECLARATION OF CONTRIBUTIONS

A. Connor Donahue

Wrote introduction, related works, and abstract sections of the report. Coded data loading and preprocessing section.

B. Terence Mpofu

Wrote results section of the report and coded linear regression analysis.

C. Sabina Bimbi

Wrote conclusion section of the report and coded correlation analysis.

D. Suphanat Juengsprasertsak

Wrote methods section of the report and coded exploratory data analysis.