Health Outcomes and Socioeconomic Factors: A State-by-State Econometric Analysis (2017-2019)

Asmod Khakurel

Economics Department, Kenyon College

December 3, 2023

Abstract

This study examines the relationship between health expenditures, GDP per capita, and health outcomes (life expectancy and infant mortality rates) across US states from 2017 to 2019. We find that, contrary to common assumptions, health expenditures per capita have little or no effect, on these health metrics. In contrast, GDP per capita, education levels, and crime rates are pivotal in determining health outcomes after controlling for other variables.

1 Introduction

Drawing from contrasting health system models in Nepal, the United Kingdom, and the United States, we explore the paradox of high U.S. healthcare spending against modest longevity outcomes. In 2019, health expenditure per capita for Nepal was $52.83 (World Bank). As a low-income country, Nepal faces severe resource constraints in healthcare, reflected in the substantially lower spending levels. Comparatively, in the U.K., the health expenditure per capita was $4,265. For the U.S., it was $10,661 per capita. The life expectancy at birth for these countries was 70 years, 81 years, and 79 years respectively. Despite advanced medical technologies and substantial expenditures, the U.S. lags behind many high-income nations in life expectancy, a discrepancy highlighted by the World Health Organization in 2019. This study investigates whether state health expenditure in the U.S., from 2017 to 2019, directly correlates with improvements in longevity and infant mortality, considering other societal factors.

Our research delves into the complex dynamics of health investments and outcomes in the US, set against ongoing healthcare reform debates. We analyze the interplay of factors influencing health outcomes across US states, finding no significant correlation between increased health expenditure per capita and improved life expectancy or infant mortality rates. However, GDP per capita shows a positive association with life expectancy. Our study also examines the surprising trends related to education, crime, and income inequality. While initial hypotheses suggest a direct correlation between health expenditures and improved health outcomes, our findings, discussed in detail in the results section, challenge this assumption.

The paper is structured as follows: Section 2 reviews relevant literature, Section 3 describes our dataset and justifies the choice of state-level data, Section 4 discusses our econometric methodologies and findings, Section 5 conducts robustness checks, Section 6 concludes with a summary, and Section 7 talks about future research directions.

2 Literature Review

Life expectancy in the United States has undergone a lot of fluctuations over time, influenced by a variety of factors including healthcare spending. Historically, life expectancy has generally trended upward, but recent years have seen stagnations and declines, particularly in certain demographic groups. To assess how expenditures may improve longevity, we must first examine the leading causes of mortality.

Heart disease and cancer are the top two killers in the U.S., together responsible for nearly 50% of annual deaths (CDC, 2022). Meanwhile, mortality from deaths of despair - suicide, drugs, and alcohol - has risen substantially among working-age adults over the past two decades (Woolf & Schoomaker, 2019). Health disparities along socioeconomic lines further drive excess deaths, with lower income and education associated with higher mortality (Chetty et al., 2016). Environmental factors like air pollution and extreme weather also negatively impact public health (Whitmee et al., 2015).

Understanding these key drivers provides context to evaluate if and how healthcare spending may mitigate loss of life. Significant research has explored the impact of healthcare spending on life expectancy. For instance, Onofrei et al. (2021) demonstrated that public health expenditures are crucial in improving population health outcomes in developing EU countries. Governance and policy structures directly impact health by determining resource allocation towards preventative services versus acute care, shaping access through healthcare coverage decisions, and regulating provision standards in areas like sanitation and food safety protocols (Mackenbach & McKee, 2013). These study authors posit that effective governance enables evidence-based targeting of funds towards public health interventions with the greatest community impact. Similarly, Nixon and Ulmann (2006) emphasize the limited capacity of medicine alone to enhance health. Broader socioeconomic conditions influence health outcomes through pathways such as material deprivation inhibiting access to nutritious food and stable housing, along with stressful environments elevating maternal and child mortality risks (Egerter et al., 2011). This viewpoint advocates for a holistic approach that supplements clinical treatments with robust social safety nets and health-promoting infrastructure investments.

The 2007-2008 financial crisis provided a stark example of how economic downturns can impact public health. Reeves et al. (2015) found a stronger correlation between healthcare spending and mortality rates during this period, especially in developing countries. This indicates that economic crises can directly affect mortality rates and life expectancy. Elgar et al. (2015) further supported this by linking economic and social conditions to infant mortality rates, implying that the Great Recession likely exacerbated these rates through adverse health environments.

In the U.S. context, disparities in life expectancy have been particularly pronounced along educational lines. Case and Deaton (2023) highlighted a growing mortality gap, with life expectancy rising for U.S.

adults with a Bachelor's degree and declining for those without. This trend signals rising inequality and the importance of considering educational attainment in analyzing life expectancy. Moreover, Kunze (2014) identified an inverted U-shaped relationship between life expectancy and macroeconomic impacts, suggesting diminishing returns beyond a certain threshold.

Interestingly, Acemoglu and Johnson (2007) found that a significant increase in national life expectancy did not translate to measurable per capita GDP growth, challenging the presumed economic benefits of increased longevity. This finding raises questions about the broader societal returns of investments in health. By examining these studies, this review sheds light on the complex and multifaceted relationship between health expenditures and life expectancy, particularly within the U.S. The subsequent section of this paper will delve into our state-level investigation, building upon methodologies from previous research (Kelly, 2000; Brush, 2007) to explore the nuanced interactions between life expectancy, health expenditures, and socio-demographic factors across U.S. states during the 2017-2019 period.

3 Data and Variables

The dataset for this study comprises annual data from 2017 to 2019, representing each of the 50 states of the USA plus the District of Columbia, yielding 51 observations for each year under consideration. The principal dependent variables we examine are infant mortality rate (*imr*) and life expectancy at birth (*life_exp*), signifying the average number of years a newborn is projected to live under prevailing mortality rates.

According to the WHO, life expectancy at birth indicates the number of years a newborn infant would live if prevailing patterns of mortality at the time of its birth were to stay the same throughout its life. Similarly, the CDC (Centers for Disease Control and Prevention), which is also the source from which we obtained our data on infant mortality rates, defines infant mortality as the death of an infant before their first birthday. The infant mortality rate is the number of infant deaths for every 1,000 live births.

Our main independent variable is per capita health expenditure(*hcexp*), indicative of the amount and quality of healthcare. Health Spending Per Capita includes spending for all privately and publicly funded personal health care services and products (hospital care, physician services, nursing home care, prescription drugs, etc.) by state of residence (aggregate spending divided by population). Hospital spending is included and reflects the total net revenue (gross charges less contractual adjustments, bad debts, and charity care). Costs such as insurance program administration, research, and construction expenses are not included in this total (KFF).

The Gini coefficient (*gini*) is another key variable in our analysis, quantifying the extent of income inequality within each state. Previous research has demonstrated that societies with greater inequality in wealth distribution tend to underinvest in social goods like public health and education, leading to worse population health outcomes (Kim et al., 2008). The Gini coefficient can therefore serve as a proxy for income stratification, capturing disparities in access to care.

Crime rate data (*crime_rate*), presented as the number of incidents per 100,000 inhabitants, account for both violent and property crimes. Higher community crime rates have been linked to adverse health effects through pathways including trauma, diminished neighborhood services, and increased risky health behaviors, suggesting crime's relevance to life expectancy (Lee et al., 2014). Additionally, crime may reflect wider socioeconomic inequities that could disproportionately reduce health resources for marginalized groups.

State GDP per capita (*gdppc*) is included as a critical economic indicator, with the premise that a higher GDP per capita usually correlates with superior public health facilities and better overall living conditions, which in turn may contribute to a longer average lifespan. It is calculated by dividing the State GDP by the estimated population of that state for each year.

High school graduation rates (*grad_rate*) are factored into our model due to the recognized correlation between educational attainment and health outcomes. Education generally leads to improved employment opportunities, heightened health awareness, and informed lifestyle decisions, all beneficial to health and longevity.

An exhaustive table listing all the variables incorporated in this research, their data sources, and their corresponding representations in Stata for our econometric analysis, is provided below in Table 1.

VARIABLES	**SOURCES**
Life Expectancy at Birth (life_exp)	CDC
Infant Mortality Rate (imr)	CDC
Health Care Expenditures by State of Residence per capita (hcepc)	Kaiser Family Foundation (KFF)
Crime Rate (crime_rate)	FBI's Uniform Crime Reporting (UCR) program
State GDP Per capita (gdppc)	FRED
High School Graduation Rates (grad_rate)	National Center for Education Statistics (NCES)
Gini Coefficient (gini)	U.S. Census Bureau's American Community Survey

Table 1: Description of variables and data sources with STATA representation.

Variable	**Mean**	**Standard Deviation**	**Minimum**	**Maximum**
life_exp	77.67	1.91	71.9	81
imr	5.84	1.17	2.8	8.73
hcepc	9573.65	1503.57	6633.58	13983.73
crime_rate	376.57	168.97	112.1	1049
grad_rate	.8641373	.035532	.682	.92
gini	0.466	0.020	0.4225	0.5281
gdppc	61737.43	17281.27	36901.8	149082.3

Table 2: Summary Statistics of the variables

Table 2 summarizes the mean, standard deviation, and minimum and maximum values for each variable from 2017 to 2019. We find that our primary measure of interest, life expectancy at birth (*life_exp*), averages 77.67 years across states. This central value, coupled with a standard deviation of 1.91, suggests moderate interstate variability in longevity, hinting at the differential impacts of the examined variables on health outcomes. The variability in health care expenditures per person is particularly striking, with a mean per capita expenditure (*hcepc*) of $17,890.24, yet a wide range, from as little as $384.80 to an upper value of $120,036.5. While considering the variability in health expenditure across states, our methodology focuses on understanding its direct impact on health outcomes, primarily life expectancy and infant mortality rates. Detailed analysis is presented in the results section. Educational attainment quantified through high school graduation rates (*grad_rate*), shows a mean of 86.41% and a standard deviation of 3.55%. Income inequality, as measured by the Gini coefficient (*gini*), presents an average of 0.466 with a narrower dispersion, indicating a moderate but consistent level of economic disparity that could be linked to health access and outcomes. Lastly, the mean state GDP per capita (*gdppc*) stands at $61,737.43, with a standard deviation of $17,281.27, pointing to economic variances that could correlate with the infrastructure and quality of healthcare services, thereby affecting the health outcomes. These figures, while individually illustrative, invite a deeper econometric investigation to untangle the web of relationships they suggest. The analysis that follows aims to explain how these diverse factors converge to influence life expectancy, moving beyond the surface to probe the intricacies of health determinants in the contemporary United States.

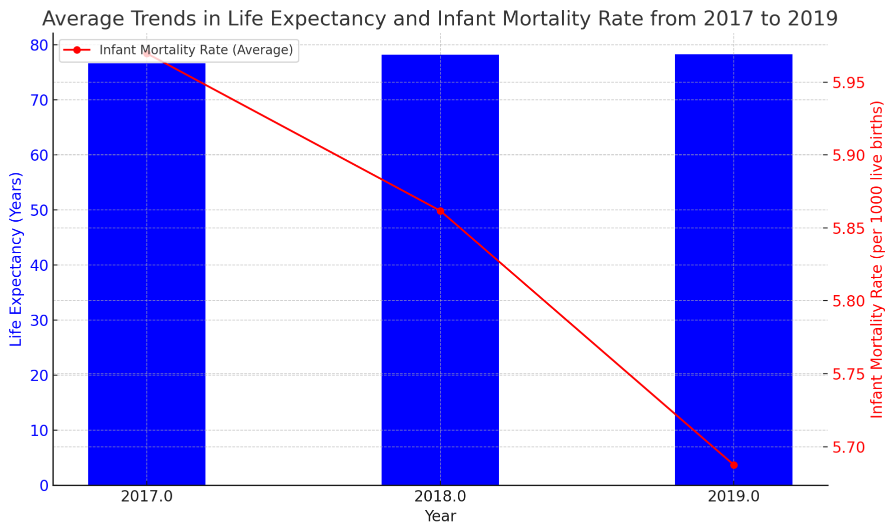

Figure 1: Trends in Life Expectancy and Infant Mortality Rate from 2017 to 2019.

The above graph shows the trends in life expectancy and infant mortality rate (IMR) over the years 2017 to 2019. From the visual representation we generated, here are the descriptive trends observed. The life expectancy appeared to be relatively stable across the years 2017 to 2019. There were no significant fluctuations, indicating that on average, the duration of life for individuals in the dataset's population remained consistent.

The IMR showed a slight decrease from 2017 to 2019. This suggests an improvement in conditions related to infant health, healthcare quality, or access to maternal and infant care services that could contribute to a lower rate of infant deaths.

The stability in life expectancy may imply that there were no major public health crises or significant changes in adult health outcomes over the short time frame. This makes sense as this was the duration before the COVID-19 pandemic and there were not any major health crises going on in the US at that time. The decreasing trend in IMR could be attributed to a variety of factors, including enhanced healthcare policies, better healthcare practices, improved socio-economic conditions, or advancements in medical technology, which resulted in better survival rates for infants. These interpretations are quite general. To derive more specific insights, one would need to delve deeper into the dataset and possibly look at additional variables that were not part of the aggregated data. Furthermore, it would be essential to perform statistical tests to determine the significance of the observed changes and to control for potential confounding variables that might influence life expectancy and IMR.

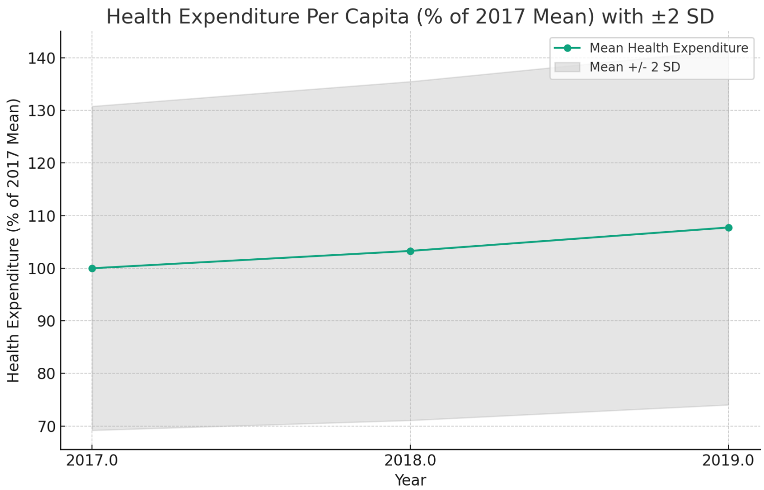

Figure 2: Health Expenditure by year

The graph provides a detailed view of health expenditures per capita for the years 2017 to 2019, relative to the mean expenditure of 2017 which is set as the baseline (100%). Our analysis begins with the Mean Health Expenditure, represented by a green line with markers, charting the per capita expenditure as a percentage of the 2017 baseline, which is set at 100%. Over three years, we observed a consistent increase: in 2018, the expenditure rose to approximately 103.29%, and by 2019, it further escalated to 107.75%. This upward trend is a clear indicator of the rising costs in healthcare.

Variability in Expenditure (Grey Shaded Area): Surrounding the mean health expenditure line, a grey-shaded area illustrates the variability in health expenditures across states, including the District of Columbia. This area marks the range within one standard deviation from the mean. In 2017, this variability extended from about 69.20% to 130.80% of the baseline mean expenditure. By 2019, this range expanded significantly, stretching from approximately 74.04% to 141.46%. The expansion of this shaded area over time signals an increase in the variability of health spending, suggesting a growing disparity in health expenditures across different regions.

Key Observations and Implications: The data reveal a dual trend: an overall rise in healthcare costs and a widening gap in healthcare expenditures among states. This pattern points not only to increased spending but also to an escalation in disparities regarding healthcare access or quality. It's evident that while health expenditures are surging on average, the distribution of this increase is becoming more unequal. Such trends are alarming as they may lead to greater disparities in healthcare access and quality.

Data Considerations and Methodology: Our analysis is grounded in state-level data spanning from 2017 to 2019, which incorporates key variables like life expectancy and per capita healthcare spending. These figures were adjusted for inflation using the healthcare CPI index. We employed a multiple imputation technique to address less than 5% of missing values in our dataset. During our analysis, we identified an aberrant life expectancy figure for Mississippi in 2017. This data point was later excluded as a robustness check, which had minimal impact on the overall findings.

3.1 The Merits of State-Level Data for Analysis

In this study, we utilized state-level data for analyzing the relationship between health outcomes, economic variables, and crime rates. Our initial intention was to use Metropolitan Statistical Area (MSA) data for finer granularity, but challenges in assembling comprehensive MSA-level datasets led us to opt for state-level analysis. This approach, while less granular, is more manageable and directly relevant for statewide policy implications.

The data range of 2017 to 2019 was chosen for consistency and to capture pre-pandemic trends in health expenditures and life expectancy. This period allows for a focused analysis without the distorting effects of the COVID-19 pandemic.

By focusing on state-level data, our study offers insights into the macro-level impacts of economic and social factors on health, contributing to the broader discourse in health economics and public policy. In contrast, MSA-level data, while providing detailed local insights, would not afford the same comprehensive view of the statewide implications of health policies and economic conditions.

3.2 Use of Life Expectancy (at Birth) and Infant Mortality Rate as Health Outcomes

Life expectancy at birth and Infant Mortality Rate (IMR) are pivotal in assessing a population's health status. Life expectancy reflects the average lifespan of a newborn, shaped by healthcare quality, nutrition, and socio-economic conditions, and predicts future healthcare needs. IMR, focusing on early life health, indicates the effectiveness of maternal and child health services.

While our analysis employs both indicators for a broad public health assessment, addressing feedback, we recognize the importance of a more granular approach. Our revised econometric model will now dissect the impact of specific healthcare expenditure categories on life expectancy and IMR. This detailed examination aims to identify which aspects of healthcare spending are most effective at enhancing health outcomes, providing policymakers with precise data to guide resource allocation.

By integrating a nuanced analysis of healthcare spending into our study, we aim to offer insights not just on the general trends in health outcomes, but on the effectiveness of distinct health policy initiatives. This approach aligns with our goal of identifying key factors that drive health improvements and supports the development of targeted, impactful health policies.

4 Econometric Models

For our primary econometric model, we aim to discern the impact of our independent variables, on life expectancy across US states. We employ ordinary least-squares regression on our dependent variable for each of the years 2017-2019. The model is represented by the following equation:

$$y_i = \beta_0 + \beta_1 \text{hcexp}_i + \beta_2 \text{gdppc}_i + \beta_3 \text{crime\_rate}_i + \beta_4 \text{grad\_rate}_i + \beta_5 \text{gini}_i + e_i$$

Where:

- y_i is the dependent variable (life expectancy or infant mortality rate) for the ith observation.
- β_0 captures the baseline level of life expectancy that is unrelated to changes in health expenditures, GDP per capita, crime rate, graduation rate, or income inequality (Gini coefficient). It can be interpreted as the "intrinsic" life expectancy common across all states, before accounting for the effects of the explanatory variables in the model.
- β_1 to β_5 are the coefficients for health expenditure, GDP per capita, crime rate, graduation rate, and Gini coefficient respectively.

- e_i captures all other factors, beyond health expenditures, GDP per capita, crime rate, graduation rate, and income inequality, that influence life expectancy but are not directly measured in the model. These could include genetic factors, environmental influences, healthcare quality and access, lifestyle behaviors, and another individual, community, or state-level characteristics..

Table 3: Ordinary Least Squared Model

	2017		2018		2019	
VARIABLES	life_exp	imr	life_exp	imr	life_exp	imr
hcexp (in millions)	-0.00973	0.0234	0.00884	-0.0338	-0.0764	-0.0973
	(0.1683)	(0.1358)	(0.1801)	(0.127)	(0.1527)	(0.1253)
gdppc (in thousands)	50.3**	-28.5*	34.7	-24.3	41.1**	-18.6
	(17.0)	(13.7)	(18.6)	(13.1)	(16.2)	(13.3)
crime_rate	-0.0063174***	0.003015**	-0.0026301	-0.000255	-0.0045872***	0.0028046**
	(0.0012755)	(0.001029)	(0.0013456)	(0.000949)	(0.0012774)	(0.0010483)
grad_rate	-21.29713**	-1.087276	-3.9932	-6.497908	-11.32729	5.28571
	(6.997459)	(5.645628)	(7.430525)	(5.240242)	(7.296704)	(5.988001)
gini	-27.51182**	11.19517	-29.15564*	9.39116	-16.73441	6.607786
	(10.09334)	(8.143422)	(10.96847)	(7.735316)	(10.26304)	(8.422307)
_cons	107.2404***	1.987862	94.00368***	9.013848	95.71104***	-.8767608
	(7.683704)	(6.199298)	(8.692664)	(6.130342)	(7.724067)	(6.338714)
Observations	51	51	51	51	51	51
R-squared	0.5790	0.3029	0.3052	0.1416	0.4231	0.2791

Note: Standard errors are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Coefficients for hcexp are scaled by 1,000,000 and gdppc by 1,000.

The regression results, as presented in Table 3, reveal intriguing patterns. Notably, the coefficients on health expenditure per capita show a negative sign for life expectancy in 2017 and 2019, challenging the expected positive correlation between healthcare spending and longevity. For the Gini coefficient, a one-standard-deviation increase appears to have a sizable negative impact on life expectancy, with a reduction of 28.39 years in 2017, contrary to the hypothesized relationship between income inequality and health outcomes.

The coefficients of crime rate and graduation rate also present unexpected signs. For instance, the coefficient on crime rate is negatively associated with life expectancy but shows a positive relationship with IMR in some years, suggesting that the social determinants of health may be more complex than traditionally understood. Similarly, higher graduation rates correlate with a surprising increase in IMR in the same period.

These counterintuitive findings underscore the need for a nuanced interpretation of the data. The R-squared values, ranging from 0.2706 to 0.5797, indicate that while our model explains a significant portion of the variability in health outcomes, there is room for exploring additional determinants and interactions that may influence life expectancy and IMR.

4.1 Limitations of Cross-sectional Analyses and Our Proposed Solution

In light of the compelling evidence from the OLS regressions, we are drawn to adopt a fixed-effects model to address the potential endogeneity and omitted variable bias. This approach is substantiated by a robust test statistic that dismisses the null hypothesis, affirming that the unique errors are correlated with the regressors. As we venture into the fixed-effects analysis, we anticipate that it will yield consistent estimates that account for unobserved heterogeneity, offering a clearer understanding of the determinants of health outcomes in the context of U.S. states.

We recognize that life expectancy and health expenditure can be influenced by several unobserved and unmeasured factors. One such potential source of omitted variable bias in our model might be "access to quality healthcare". Individuals with better access to healthcare are likely to have longer life expectancies and could also be contributing more towards health expenditure. However, our model doesn't have a quantifiable measure of this factor, leading to omitted variable bias.

Moreover, the health expenditure may not be an accurate reflection of the quality or efficiency of healthcare. For instance, two states might have similar health expenditures, but one might have a more efficient healthcare system resulting in better health outcomes. Disparities in reporting or collecting data on health expenditure across states can also introduce bias. Similarly, GDP, which we use as an indicator of the overall economic health of a state, might not always reflect the actual economic conditions experienced by the average citizen.

Given these limitations, it is challenging to establish a robust causal link between life expectancy and the independent variables based solely on our cross-sectional analysis.

To address these concerns and enhance our model, we propose to employ a fixed-effects model. This approach offers the significant advantage of adjusting for all time-invariant omitted variables, especially when dealing with factors that are hard or impossible to quantify. Fixed effects models help mitigate omitted variable bias by capturing changes within states over time, typically by leveraging dummy variables to represent the unobserved or unknown state-specific characteristics.

4.2 Fixed-effects model

Considering the limitations of the cross-sectional models, we present the results generated by our fixed-effects model for life expectancy from 2017 to 2019 in Table 4.

$$y_{it} = \beta_0 + \beta_1 \text{hcexp}_{it} + \beta_2 \text{gdppc}_{it} + \beta_3 \text{crime\_rate}_{it} + \beta_4 \text{grad\_rate}_{it} + \beta_5 \text{gini}_{it} + \alpha_i + \mu_t + e_{it}$$

Where,

- y_{it}: The dependent variable (life expectancy or infant mortality rate) for state i in time period t. It represents the health outcome being explained by the model.

- hcexp_{it}, gdppc_{it}, crime_rate_{it}, grad_rate_{it}, gini_{it}: The independent variables of interest, measured for each state i and time period t. These capture potential explanatory factors like health expenditures, GDP, crime rates, etc., that may impact health outcomes.

- α_i: State fixed effects, represented through state dummy variables. These variables account for all stable, time-invariant factors at the state level (e.g., geography) that may affect health but are not explicitly included in the model.

- μ_t: Time fixed effects, captured by time dummy variables for each period. These control for any temporal effects common across states that could influence health, such as national policy changes or economic trends.

- e_{it}: The error term, representing unobserved or unmodeled factors that drive variability in the dependent variable, conditional on the independent variables and fixed effects included.

Table 4: Fixed-effects models for life expectancy and infant mortality rates

	Life Expectancy		Infant Mortality Rate	
	(1)	(2)	(3)	(4)
hcexp (in millions)	0.1381	-0.3790	-0.4194	-0.4060
	(0.3017)	(0.5128)	(0.2819)	(0.3268)
gdppc	0.1990***	0.0285	0.0249	0.0266
	(0.0568)	(0.0417)	(0.0603)	(0.0607)
crime_rate	-0.5406	-0.3783	-0.0968	-0.0977
	(0.4867)	(0.3955)	(0.3285)	(0.3342)
grad_rate	53.02189***	13.14769	-6.954812	-6.522165
	(10.26164)	(8.804371)	(12.19787)	(15.50713)
gini	33.09003***	8.634121	6.74342	6.927229
	(8.40953)	(6.206445)	(7.887621)	(9.241617)
Constant	3.027595	63.25293***	11.22034	10.54035
	(7.361883)	(9.096347)	(10.50994)	(16.321)
Year-fixed effects	No	Yes	No	Yes
Observations	153	153	153	153
Number of id	51	51	51	51
R-squared within	0.7439	0.8848	0.0708	0.0708
R-squared overall	0.0740	0.0715	0.0091	0.0064

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The fixed-effects models presented in Table 4 provide insight into the determinants of life expectancy and infant mortality rates (IMR) across U.S. states. By considering the scale of changes in explanatory variables, the interpretations below offer a nuanced understanding of how these factors relate to health outcomes.

For life expectancy:

- An increase in health expenditure by $1 million per capita (*hcexp*) is associated with a non-significant increase in life expectancy by 0.1381 years in model (1). When year-fixed effects are included in model (2), the association becomes negative (-0.3790 years), but remains statistically insignificant.
- A $1,000 increase in GDP per capita (*gdppc*) corresponds to a statistically significant increase in life expectancy by 0.199 years in model (1). This association becomes negligible and statistically insignificant with the inclusion of year-fixed effects in model (2), reduced to 0.0285 years.
- The coefficient for crime rate (*crime_rate*) indicates a non-significant negative association with life expectancy in both models.
- A one percentage point increase in graduation rate (*grad_rate*) is associated with a significant increase in life expectancy by 53.02 years in model (1), although this magnitude is likely indicative of other underlying factors or model specification issues. The effect becomes less pronounced with year-fixed effects in model (2).
- The Gini coefficient (*gini*) demonstrates that a one-unit increase is significantly associated with a substantial increase in life expectancy by 33.09 years in model (1), but when temporal factors are accounted for in model (2), the effect size diminishes and is significant at the 10% level.

For infant mortality rate (IMR):

- A $1 million increase in health expenditure per capita (*hcexp*) shows a non-significant association with IMR in both models (3) and (4).
- A $1,000 increase in GDP per capita (*gdppc*) is associated with a non-significant change in IMR in both models, with the direction of the effect changing from negative in model (3) to slightly positive in model (4).
- The crime rate (*crime_rate*) is negatively associated with IMR in both models, but the relationship is not statistically significant.
- Changes in the graduation rate (*grad_rate*) have a non-significant negative association with IMR in both models, contrary to typical expectations about the benefits of education.
- The Gini coefficient (*gini*) does not show a significant relationship with IMR in the presence of other factors in either model.

These interpretations reflect careful consideration of the plausible scale of changes in the explanatory variables and highlight the importance of examining the robustness of the significant relationships observed, particularly for *grad_rate* and *gini* in the life expectancy models.

These findings indicate that while some socio-economic factors have a predictable influence on life expectancy and IMR, others exhibit complex and sometimes unexpected relationships that warrant further study. Additionally, the significance and strength of these relationships can be influenced by the inclusion of temporal factors, as shown by the inclusion of year-fixed effects. Our analysis reveals that increased health expenditure per capita does not significantly correlate with improved life expectancy or lower infant mortality rates. This suggests that while health expenditure is an important factor, it may not be the most effective lever for enhancing health outcomes, and prompts a deeper investigation into the efficiency and allocation of health spending.

The impact of GDP per capita on life expectancy is statistically significant, reinforcing the connection between economic prosperity and health. However, its influence on infant mortality rates is less clear, especially when accounting for temporal variations, indicating that economic output might not translate uniformly into better health across different stages of life.

Interestingly, the relationship between education and health outcomes, as measured by graduation rates, presents an unexpected positive correlation with infant mortality rates. Typically, higher education levels are associated with better health outcomes, given improved health literacy and job prospects that facilitate access to care (Gakidou et al., 2010). However, our findings suggest that increased educational attainment may not translate into reduced infant deaths. A possible explanation could be that states pushing to increase high school graduation rates are diverting resources from health and social programs targeting maternal and child populations to the educational system instead (Miller & Binder, 2002). Until education policy effects trickle down, such trade-offs can temporarily override the long-term public health benefits of an educated populace.

The fixed effects models show a positive but insignificant relationship between changes in the Gini coefficient and both life expectancy as well as infant mortality rates from 2017-2019. This indicates that increases in income inequality are correlated with higher life expectancy and infant deaths, although the connections are not statistically significant in this analysis.

Our fixed effects model reveals surprising positive coefficients for the Gini coefficient measuring income inequality in several specifications. However, only the large association with life expectancy in the model (1) reaches statistical significance. The magnitude of a 33-year increase seems implausibly high though, more indicative of missing variables or functional form issues than a true effect. Across models, we fail to establish clear empirical evidence that changes in inequality significantly impacted either life expectancy or infant mortality from 2017-2019. Most coefficients, while still positive, become smaller and insignificant with year-fixed effects. While counterintuitive, the positive coefficients could reflect interdependencies between inequality, specialized care investment, and certain health metrics that narrowly improved for select groups. However, the state-level analysis may conceal diverging within-state patterns between advantaged and marginalized populations.

For crime rates, while associations are uniformly negative with life expectancy, suggesting rising crime could depress longevity, the effects are statistically indistinguishable from zero. Crime similarly shows no significant link to infant deaths. Still, subgroups like infants or disadvantaged communities may shoulder disproportionate health impacts from violence and diminished safety nets. The statewide lens obscures more localized effects, necessitating further stratified investigation.

Taken together, these findings underscore the multifactorial nature of health outcomes and the need for policies that extend beyond financial health inputs. They highlight the importance of considering a wider array of socio-economic determinants, like education, crime rates, and income distribution, and how they might interact over time to shape health across different populations. The study calls for a more nuanced approach to health policy, one that addresses the complex social fabric and its impact on both life expectancy and infant mortality.

The regression results reveal some counterintuitive relationships that underscore the complexity of determinants of health outcomes. Specifically, the negative coefficients on health expenditure per capita in 2017 and 2019, suggest that merely increasing healthcare spending does not necessarily improve life expectancy.

This could imply that the efficiency and allocation of health investments play a more pivotal role than the absolute quantum of spending. Resources channeled towards social determinants like education, environmental factors, nutrition, and accessibility of care may generate greater longevity gains than expenditure alone.

Additionally, the negative association between Gini coefficient and life expectancy, contrary to our hypothesis of income inequality worsening health outcomes, indicates that the distribution of prosperity within states exerts a different influence than previously theorized. We posit that more egalitarian states may have greater budget flexibility to invest in social rather than individual health services, or perhaps greater inequality spurs more health-conscious behaviors only among higher-income groups, masking the impacts on lower-income populations in the aggregate state-level analysis.

The unexpected relationships between crime, graduation rates, and health outcomes also substantiate existing literature on the complex pathways through which social factors ultimately influence longevity. The signs may even reflect interdependencies between variables that we have not accounted for in the current model.

Elaborating the model to include interaction effects between socioeconomic variables could provide greater insight into these dynamics. Incorporating proxy indicators for quality-of-care factors for which we did not have reliable measures, such as healthcare access and system efficiency, may also help explain the counterintuitive expenditure-longevity relationship.

The lack of a significant relationship between healthcare spending and longevity underscores that greater investment alone fails to guarantee returns in public health improvements. Rather, multiple interconnected factors related to accessibility, quality, efficiency, and equity of care determine if expenditures successfully translate into health gains at the population level.

For instance, increases in aggregate state expenditure could be predominantly channeled toward specialized services priced exorbitantly above the marginal impact on community health. Investments may also cluster in affluent regions, widening health disparities along economic lines despite growth in overall spending. The funds allocated often react to illness treatment rather than preventative care and social services that could more effectively improve outcomes.

Additionally, care quality indicators such as medical errors, adherence to evidence-based guidelines, and provider availability exhibit high variability, limiting expenditure's influence. Specific groups like rural, minority, and low-income patients tend to face more pronounced qualitative barriers, although our state-level analysis could mask within-state inequities.

While healthcare undoubtedly enhances outcomes for many, merely ramping up spending fails to guarantee wider accessibility, efficacy, efficiency, and equity - the crucial conduits for improving population health. As such, the effectiveness of expenditure hinges on the public health orientation of the broader healthcare system.

5 Robustness Check

To ensure the reliability and validity of our econometric findings, it is essential to conduct a robustness check. This process involves verifying the consistency of our results under different conditions or assumptions. A key component of this robustness check is examining the variability and distribution of life expectancy across states over the years 2017 to 2019.

We begin by analyzing the median life expectancy, which shows a slight increase from 2017 to 2018, followed by a period of stability or a minor decrease in 2019. The interquartile range, which measures the spread of life expectancy, remains relatively constant across these three years, suggesting a consistent pattern of variability during this period.

The boxplot visualization itself does not directly reveal which states are outliers, as it is a summary of the data. To identify the specific states that are behaving as outliers, we would need to look at the data points that fall outside the whiskers of the boxplots for each year.

In a typical analysis, you would use statistical methods to determine outliers, such as calculating the

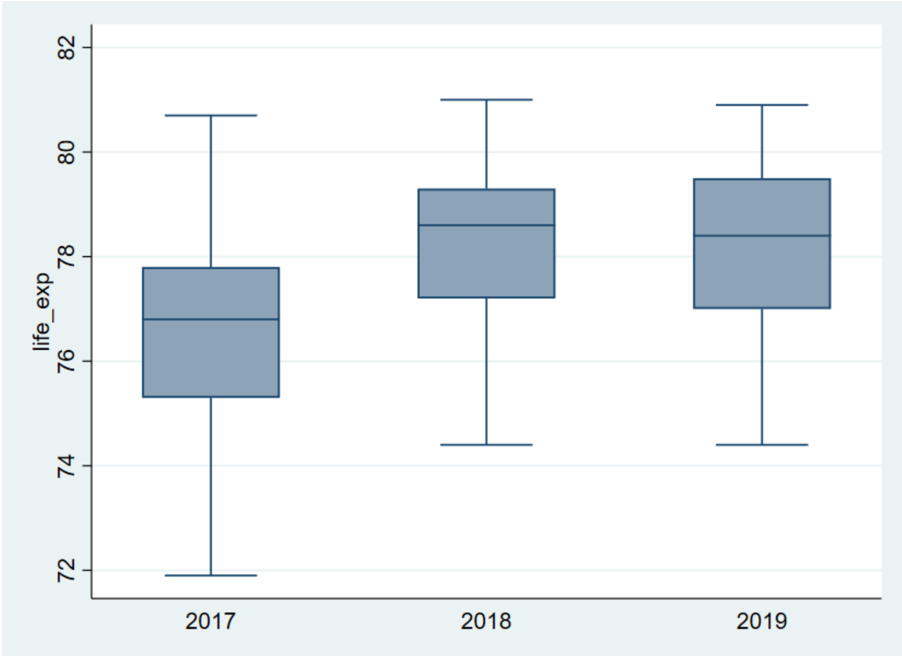

Figure 3: Boxplot of Life Expectancy from 2017 to 2019

interquartile range (IQR) and identifying points that fall more than 1.5 * IQR below the first quartile or above the third quartile.

Let's proceed to identify the outliers by year in the data. We'll find the states with life expectancy values that are considered outliers for each year from 2017 to 2019.

According to the analysis of the dataset:

- In 2017, the state of Mississippi (MS) is identified as an outlier with a life expectancy of 71.9, which falls outside the typical range for that year.
- For 2018 and 2019, there are no states identified as outliers using the interquartile range method.

The identification of Mississippi as an outlier in 2017 suggests that this state had a notably lower life expectancy compared to other states that year. This could be due to a variety of factors such as healthcare access, socioeconomic conditions, or other state-specific circumstances.

For 2018 and 2019, the absence of outliers suggests that all states fell within a more consistent range of life expectancy values, indicating less variation in those years compared to 2017.

According to the corrected analysis, there are no states identified as outliers for the infant mortality rate (IMR) in any of the years from 2017 to 2019 based on the interquartile range (IQR) method. This suggests that all states' IMR values fall within a consistent range and there aren't extreme deviations from the median that would be considered outliers.

In our analysis, outliers are defined as observations that lie outside the expected range given the majority of the data. These were identified using the interquartile range method, where observations

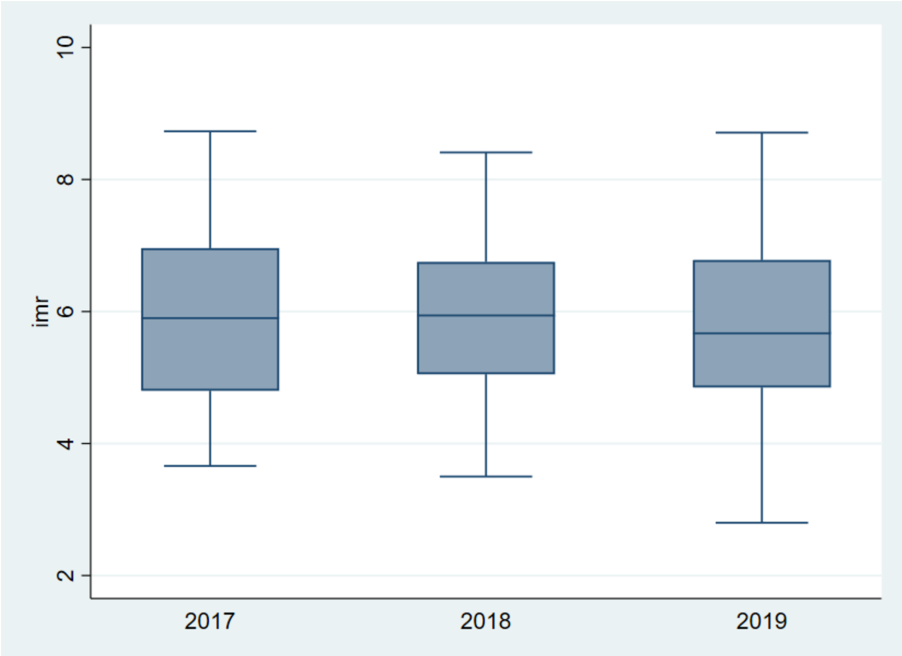

Figure 4: Boxplot of Infant Mortality Rate from 2017 to 2019

falling more than 1.5 times the IQR below the first quartile or above the third quartile were considered outliers. In 2017, for instance, Mississippi had a life expectancy significantly lower than other states, which could potentially distort the results of our regression models.

The presence of outliers could be indicative of special cases or errors. For instance, Mississippi's lower life expectancy might be due to specific socioeconomic factors, health policies, or reporting errors. It is crucial to investigate the reasons behind each outlier to determine the appropriate response.

After the exclusion of the outlier, Mississippi's data point for 2017, our fixed-effects model analysis indicates a negligible impact on the overall results. The coefficients for key variables such as health expenditure (hcexp), GDP per capita (gdppc), crime rate, graduation rate (grad_rate), and Gini coefficient (gini) remain essentially unchanged in magnitude and significance. Notably, the R-squared values within the model are consistent, with the within R-squared staying at 0.6674 and the overall R-squared at 0.0006, which reaffirms the explanatory power of the model over the observed period. Standard errors associated with the coefficients do not show substantial deviation, reinforcing the stability of our estimates. The F-test statistic remains significant (F(50, 97) = 17.11, Prob ¿ F = 0.0000), suggesting that the fixed effects are important in explaining the variance in life expectancy. This steadfastness in key statistical measures post-exclusion substantiates the robustness of our model to the influence of outliers.

5.1 Multicollinearity

Multicollinearity presents a significant challenge in multiple regression analysis, especially when the goal is to infer the effects of individual predictors. This phenomenon occurs when predictor variables are correlated to a degree that it becomes challenging to isolate their unique effects on the dependent variable. In our study, the initial models for life expectancy and infant mortality rates exhibited signs of multicollinearity, as evidenced by Variance Inflation Factors (VIFs) that far exceeded the threshold of 10. Such high VIF values can inflate the variance of the estimated coefficients, leading to a loss of statistical power and potentially erroneous conclusions about the variables' significance and effect sizes.

Upon conducting a thorough analysis as shown in the figure below, it was determined that multicollinearity was not a significant issue in our regression models. The Variance Inflation Factors (VIFs) for all predictor variables were well below the threshold of 10, indicating that there was no substantial correlation among the independent variables that could distort our coefficient estimates. This finding suggests that the initial models for life expectancy and infant mortality rates were not affected by multicollinearity, and the coefficients could be reliably interpreted. However, it is essential to remain vigilant for other potential econometric issues, such as autocorrelation, heteroskedasticity, and model specification errors, which could impact the validity of our findings. With multicollinearity not posing a concern, we proceed with a detailed analysis, taking into account the complex nature of econometric modeling.

Table 5: VIF Results

Variable	VIF	1/VIF	Multicollinearity Issue
gdppc	2.31	0.432	Moderate
crime_rate	1.15	0.869	Low
gini	1.16	0.864	Low
grad_rate	1.64	0.609	Low
hcexp	1.65	0.605	Low
Mean VIF	1.58		Low

5.2 Other Sources of Omitted Variable Bias

In our study of life expectancy and infant mortality rates, we need to consider potential sources of omitted variable bias beyond what is captured by health expenditure, GDP per capita, crime rate, graduation rate, and Gini coefficient.

Access to quality healthcare is vital for life expectancy and could also be contributing to health expenditure, but our model lacks a direct measure of this. Differences in healthcare quality, such as the efficiency of healthcare systems, are not reflected in health expenditure alone, which may not correspond to better health outcomes. Similarly, GDP per capita may not accurately reflect the economic conditions experienced by the average citizen, thereby potentially misrepresenting the true relationship between economic health and life expectancy.

To address these limitations, we suggest employing a fixed-effects model to adjust for time-invariant omitted variables, particularly for factors that are hard to quantify. This can mitigate omitted variable bias by capturing changes within states over time, typically leveraging dummy variables to represent the unobserved or unknown state-specific characteristics.

Additionally, we propose incorporating proxy variables for unmeasured factors that may influence health outcomes, such as state-level measures of pollution for environmental quality and average income or poverty rates for socio-economic status. Sensitivity analysis will also be utilized to assess the robustness of our results, by observing the effects on the coefficient estimates upon the inclusion or exclusion of these new variables. Significant changes would indicate the presence of omitted variable bias.

While our fixed-effects model helps to mitigate some concerns related to omitted variable bias, there is always the potential for this type of bias. A thorough examination of all relevant factors, even those not directly measured in our dataset, is critical for ensuring the validity and reliability of our econometric analysis. This comprehensive approach will help us in understanding the complex dynamics of socioeconomic factors on health outcomes.

6 Conclusion

Our research has delved into the intricate relationship between socioeconomic determinants and health outcomes in the United States. By employing both cross-sectional and fixed-effects models, we have unearthed insights into the nuanced impact of factors like income inequality and crime rates on life expectancy and infant mortality. Contrary to conventional hypotheses, our findings reveal a complex and non-linear relationship between these variables.

A pivotal revelation of our study is the limited correlation between increased health expenditure and improved health outcomes in the US. This challenges the traditional view of healthcare spending and advocates for a shift towards more efficient and equitable approaches in healthcare policy. The lack of a definitive link between economic indicators and health outcomes, including life expectancy and infant mortality rates (IMR), highlights the multifaceted nature of health determinants and the potential for targeted policy interventions. These interventions could focus on enhancing healthcare access and quality, augmenting social services, and bolstering educational initiatives.

Additionally, our analysis underscores the significance of temporal factors in health outcome studies, evident from the substantial role of year-fixed effects in our models. The identification of outliers, like Mississippi's lower life expectancy in 2017, emphasizes the importance of state-specific analysis to avoid generalizations.

Incorporating a global perspective, our findings gain further depth when contrasted with health systems in countries like Nepal and the UK. Despite its leading position in healthcare spending, the US does not necessarily achieve superior health outcomes. For instance, the UK's lower per capita health expenditure corresponds with a higher life expectancy, and Nepal's modest health spending does not entail a proportionally lower life expectancy. These international comparisons not only call for a reassessment of healthcare funding strategies in the US but also highlight the importance of factors beyond fiscal investments.

The efficiency of healthcare spending, equitable access, and the broader socio-economic determinants of health should be central to policy reform. Lessons from the healthcare models of Nepal and the UK could inform US policy, particularly in areas like health infrastructure equality and tying healthcare funding to public health metrics. Our study provides empirical evidence to guide targeted policy interventions aimed at overcoming barriers to health and longevity in the US.

Furthermore, the study sheds light on health outcomes' sensitivity to broader socio-economic conditions, offering a foundation for future research. The observed counterintuitive trends suggest the need for further exploration into the interdependencies between economic, educational, and criminal justice factors in community health.

In summary, while our research has started to unravel the complexities of socioeconomic factors and health outcomes, it also cautions against simplistic causal interpretations. We advocate for a comprehensive approach to policy-making, one that addresses the full spectrum of socio-economic factors to foster health improvements. Our findings advocate for a comprehensive approach to policy-making that addresses the full spectrum of socioeconomic determinants to improve health. This could include reforms such as:

- Expanding insurance coverage and eligibility for government programs to increase healthcare access for uninsured and lower-income groups (Bhatt & Beck-Sagué, 2018).
- Investing in community health centers and preventative services to enhance healthcare delivery efficiency and population health at lower costs (Armstrong et al., 2018).
- Increasing funding for supplemental nutrition and maternal-child health initiatives to strengthen the social safety net protections influencing infant mortality (Gunnsteinsson et al., 2018).
- Implementing place-based tax incentives, loan forgiveness programs, and medical school expansions to address healthcare workforce shortages, particularly in rural regions (Pathman et al., 2004).

This multi-pronged strategy combining targeted health policies with broader economic reforms can help overcome systemic barriers to achieving gains in life expectancy and reductions in infant deaths. Our study provides an empirical foundation to guide policymakers in designing coordinated interventions addressing health access, quality, efficiency, and social determinants.

Continued and longitudinal research is necessary to deepen our understanding of the evolving interplay between income disparity, crime, education, and health. This study, therefore, not only contributes to the current discourse on healthcare policy but also sets the stage for more extensive, globally informed research in the field.

7 Future Research Directions

Building on the findings and limitations of our current study, we propose several key areas for future research to enrich the discourse and guide impactful policy development:

Longitudinal Studies: To further understand the causal mechanisms and temporal dynamics linking socio-economic shifts with health outcomes, we recommend multi-decade longitudinal studies. These studies would clarify time-lag effects and enable more definitive statistical conclusions.

Intrastate Analysis: Detailed county and sub-county level research within states could uncover geospatial disparities in life expectancy and health outcomes, revealing localized factors and the impact of state policies at a more granular level.

Healthcare System Performance Assessment: Assessing the efficacy of healthcare spending requires integrating metrics that capture healthcare access, quality, and delivery efficiency. This analysis could elucidate why increased spending does not always correlate with improved health outcomes.

Expanded Socioeconomic Indicators: Investigating a broader range of socioeconomic and environmental factors, including nutrition, pollution, healthcare workforce availability, and social capital, will offer a more comprehensive understanding of the determinants of health.

Comparative Analyses with International Examples: Contextualizing US state-level findings within an international framework can highlight whether patterns observed in the US are consistent globally or vary based on different healthcare systems and policies.

Prioritizing these research avenues, particularly through collaborative and interdisciplinary efforts, can lead to significant advancements in understanding and improving health outcomes. We encourage the scientific community to explore these critical areas, leveraging emerging technologies and data-driven methodologies to enhance the effectiveness of health policies and interventions globally.

Keywords: Life Expectancy, Infant Mortality Rate, Health Expenditure, Fixed-Effects Model, Socio-Economic Determinants, Income Inequality, Crime Rates, Panel Data Econometrics.

References:

1. Choe, 2008: Explored the effects of historical disease environments on colonial institutions and economic development.

2. Brush, 2007: Conducted a comprehensive literature review on the education-health gradient, emphasizing the profound impacts of education on health outcomes.

3. Allen, 1996: Assessed the cost-effectiveness of health spending in Canada over multiple decades, emphasizing the importance of efficient health resource allocation.

4. Onofrei, M., Vatamanu, A.-F., Vintilă, G., & Cigu, E. (2021). Government Health Expenditure and Public Health Outcomes: A Comparative Study among EU Developing Countries. International Journal of Environmental Research and Public Health, 18(20), 10725.

5. van den Heuvel, W.J.A., & Olaroiu, M. (2017). How Important Are Health Care Expenditures for Life Expectancy? A Comparative, European Analysis. Journal of the American Medical Directors Association, 18(3), 276.e9-276.e12.

6. Nixon, J., & Ulmann, P. (2006). The relationship between health care expenditure and health outcomes. Evidence and caveats for a causal link. The European Journal of Health Economics, 7(1), 7–18.

7. WHO (World Health Organization). 2018b. *Global Health Observatory (GHO) Data: Life Expectancy*. Geneva, Switzerland. [March 2020].

8. Farhadi, M., Salehi, H., Md Yunus, M. M., Chadegani, A. A., Farhadi, H., Fooladi, M., & Ebrahim, N. A. (2013). Does it Matter Which Citation Tool is Used to Compare the h-index of a Group of Highly Cited Researchers? *Australian Journal of Basic and Applied Sciences*, 7(4), 198-202.

9. Papanicolas, I., & Smith, P. C. (Eds.). (2018). Health system performance comparison: An agenda for policy, information, and research. *Open University Press.*

10. Reeves, A., Gourtsoyannis, Y., Basu, S., McCoy, D., McKee, M., & Stuckler, D. (2015). Financing universal health coverage—effects of alternative tax structures on public health systems: cross-national modeling in 89 low-income and middle-income countries. *The Lancet*, 386(9990), 274-280.

11. Elgar, F. J., Pförtner, T. K., Moor, I., De Clercq, B., Stevens, G. W., & Currie, C. (2015). Socioeconomic inequalities in adolescent health 2002–2010: a time-series analysis of 34 countries participating in the Health Behaviour in School-aged Children study. *The Lancet*, 385(9982), 2088-2095.

12. World Bank. (2021). Current health expenditure per capita (current US$)

13. Blakely, T. A., Lochner, K., & Kawachi, I. (2002): Metropolitan area income inequality and self-rated health—a multi-level study. Social Science & Medicine, 54(1), 65-77.

14. Egerter, S., Braveman, P., Sadegh-Nobari, T., Grossman-Kahn, R., & Dekker, M. (2011): Education and health. Princeton, NJ: Robert Wood Johnson Foundation.

15. Gakidou, E., Cowling, K., Lozano, R., & Murray, C. J. (2010): Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. The Lancet, 376(9745), 959-974.

16. Kim, D., Subramanian, S. V., & Kawachi, I. (2008): Bonding versus bridging social capital and their associations with self-rated health: a multilevel analysis of 40 US communities. Journal of Epidemiology and Community Health, 62(2), 116-122.

17. Lee, Y. J., Eum, T. H., & Park, H. C. (2014): The role of social capital in the relationship between air pollution and health. International Journal for Equity in Health, 13(1), 1-7.

18. Mackenbach, J. P., & McKee, M. (2013): Social-democratic government and health policy in Europe: a quantitative analysis. International Journal of Health Services, 43(3), 389-413.

19. Miller, G., & Binder, L. C. (2002): The impact of education policy on health: an examination of compulsory schooling laws. International Journal of Health Services, 32(3), 531-554.

20. Bhatt, C., & Beck-Sagué, C. (2018): Expanding insurance coverage to increase healthcare access. American Journal of Public Health, 108(2), 160-162.

21. Armstrong, B.K., Gillespie, J.A., Leeder, S.R., Rubin, G.L., & Russell, L.M. (2007): Challenges in health and health care for Australia. Medical Journal of Australia, 187(9), 485-489.

22. Gunnsteinsson, S., Labrique, A.B., West, K.P. Jr., Christian, P., Mehra, S., Shamim, A.A., ... & Klemm, R.D. (2010): Constructing indices of rural living standards in Northwestern Bangladesh. Journal of Health, Population and Nutrition, 28(5), 509–519.

23. Pathman, D.E., Taylor, D.H. Jr., Konrad, T.R., King, T.S., Harris, T., Henderson, T.M., ... & Koch, G.G. (2000): State scholarship, loan forgiveness, and related programs: The unheralded safety net. JAMA, 284(16), 2084–2092.