A study on effects of Healthcare Costs and Household Income on Premature Age-Adjusted Mortality Rate in US Counties

540455982

Introduction

It has been widely studied that the impact of socioeconomic factors has always been a key area of interest in regards to public health.[1] Among these factors, healthcare costs and household income are important determinants of health. Through this study, I aim to explore the role of these two predictors on the Premature Age-Adjusted Mortality(PAAM) across US counties. The expected outcome of this study is to establish the importance of economic stability, measured via median household income and affordable healthcare(healthcare costs).

The relationship between economic factors and health outcomes is already well documented in health literature. Household income and healthcare costs significantly influence the quality of care and overall health status.[2-5] Part of the reason is that higher household income means access to better quality of healthcare, while lower healthcare costs allow more of the population to have access to health services. Another part is that individuals in higher income brackets have healthier living conditions and, can afford healthier lifestyles.[6] Income also affects mental health and stress, contributing to a persons overall well-being.[7]

Providing evidence from the data, and after controlling for predictors such as college education, prevalance of food insecurity and insufficient sleep which are correlated, there is still a significant correlation between healthcare costs and household income on the PAAM across US counties. The study hypothesis is that higher household income and lower healthcare costs are associated with lower age-adjusted mortality. By investigating this relationship, I aim to provide insights into the importance of these factors in shaping policy decisions regarding reducing mortality rates through timely and impactful economic and healthcare interventions.

In the next section, I will divide the previous literature into 2 subsections, one each for correlation between household income and health, and the other for healthcare costs and health.

Literature and theory

Household income and health

Household income can be considered a key determinant of health, impacting access to health resources, healthcare and living conditions. Research by Cutler,Muney[8] highlights that income is positively correlated to longevity and lower rates of chronic diseases. Several factors contribute to the fact that higher household income is generally associated with better health outcomes. First, individuals with higher incomes can afford healthier living practices, including safer housing, better nutrition and regular physical activity.[9] Additionally, higher household income provides better access to healthcare services, including timely access to medical care and attention.[10]

Healthcare costs and health

High healthcare costs can be a significant obstacle to accessing necessary medical aid. When the cost of medical services is high, families might delay or ignore treatment, leading to worse health outcomes and higher mortality.[11] Past studies have shown that out-of-pocket healthcare expenses substantially impact health-seeking behaviour particularly in uninsured and low income households.[12] Rise of healthcare costs can financially burden and negatively impact mental health, further contributing to poor health outcomes.[13]

The theoretical foundation of this study is based on the social determinants of health, which links economic and social conditions to health outcomes.[14] This study emphasizes the importance of addressing economic disparities and ensuring widespread access to healthcare to improve population health and reduce mortality. The study also highlights a relation between health and belief, suggesting that health behaviours are influenced by barriers to healthcare and a persons perception of health risks.[15]

This research builds on these findings by examining the context on US counties while controlling for a set of variables to account for potential confounding factors. By limiting the focus to healthcare costs and household incomes, this research aims to provide an understanding on how these economic factors influence the PAAM and inform policy interventions aimed at reducing mortality and improving public health.

Data and methodology

Data

The data for this study has been sourced from the County Health Rankings & Roadmaps programs, which is a collaboration between the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute. This dataset provides comprehensive data on health outcomes and determinants for US counties. It includes variables such as PAAM, household income and healthcare costs, along with control variables such as education, food insecurity, mental distress, number of uninsured adults in the county, Insufficient sleep percent. Their detailed information is mentioned below:

Premature age-adjusted mortality (dependent variable): PAAM is sourced from the CDC Wonder mortality data for the years 2014-2016. This variable specifies the PAAM for each US county and includes records for 3142 counties. Since this is the dependent variable, there is no need to recode the data, I am just renaming the column to PAAM in my final dataset for ease of use.

Household income (independent variable): The median household income is sourced from Small Area Income and Poverty Estimates for the year 2016. This variable is one of the two independent variable selected for this study with 3142 records and a mean value of 49522 USD. Again, for ease of use, this has been renamed to “Household_Income” for model development.

Healthcare costs (independent variable): This variable is sourced from Dartmouth Atlas of Health Care conducted during the year 2015 and is a key independent variable in determining the PAAM across US counties. It has been renamed to “Healthcare_Costs” for model development later. It has a mean value of 9630 USD with the minimum being 3896 and maximum as 19803 USD.

Food insecurity: This variable is sourced from Map the Meal Gap conducted during the year 2015 and is used as control variable in this study. It has been renamed to “Food_Insecurity” and shows a correlation with PAAM as well as the Independent variables (refer figure 1).

College Education: This variable is sourced from American Community Survey, 5-year estimates and just describes the Percentage of adults ages 25-44 with some post-secondary education within a county. Again, this is used as a control variable as it shows high negative correlation with PAAM and postive correlation with household income (refer figure 1).

Insufficient sleep: This variable is sourced from the Behavioral Risk Factor Surveillance System and is being controlled for in the model. It shows low correlation for both PAAM and the 2 independent variables. This combined with Food insecurity can be considered indicators of general well-being.

Methodology

To examine the relationship between healthcare costs, household income and PAAM, I will be using a linear regression model. Linear regression allows us to estimate the direct effect of household income and healthcare costs on PAAM while controlling for other variables. The reason for selecting linear regression is that it allows us to quantify the relationship between dependent and independent variables in a simple and easy to understand manner while also being able to control for multiple confounding factors. Since the dependent variable is continuous, linear regression can prove to be a good fit for this study. Despite its advantages, linear regression assumes a linear relation and can be affected by residual confounding due to unmeasured variables.

The dataset will be analysed in depth to identify the relationship between the independent variables and PAAM. This will be done with the help of scatterplots and line of best fit, this allows us to gauge if the relation is a positive or a negative relation. First, as part of the pre-processing, the dataset is cleaned and filtered to select only the required variables and then merged with another dataset containing the "College_Education" variable. This merge is done on the FIPS code which is unique for each county. Next, standardization is performed on median household income and healthcare costs as they show high variability.

A limitation here is that external factors such as county demographics can introduce potential confounding as counties with high population of elderly people can skew results. Another potential limitation here is that of measurement error, since some people can falsely report their household income due to fear of social standing and even tax evasion.

Results

Descriptive Statistics

Table 1 shows some statistics about the variables selected for my model. We can see that household income has a big difference between minimum and maximum value and hence its required to be scaled as part of preprocessing.

Table 1: Variable Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
Food_Insecurity	3142	14	4.2	3	11	16	38
Insufficient_Sleep	3142	33	4.2	23	30	36	47
Healthcare_Costs	3135	9630	1500	3896	8627	10522	19803
Household_Income	3141	49522	12887	22045	41072	55308	134609
Education_College	3142	57	12	16	49	66	94

It is also interesting to note that in the dataset there are a total of 3142 records, one each for each US county. This means that there are also some missing values for both household income and healthcare costs. To deal with this I have imputed the missing data based on the mean value of these variables.

The correlation heatmap shows an initial negative correlation between PAAM and household income while showing a positive correlation between PAAM and healthcare costs.

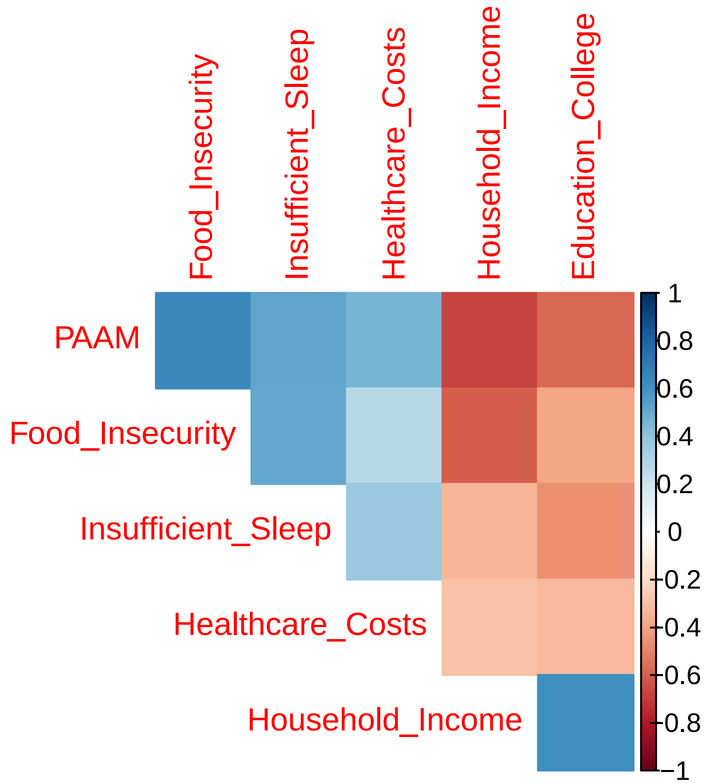

Figure 1: Correlation Heatmap

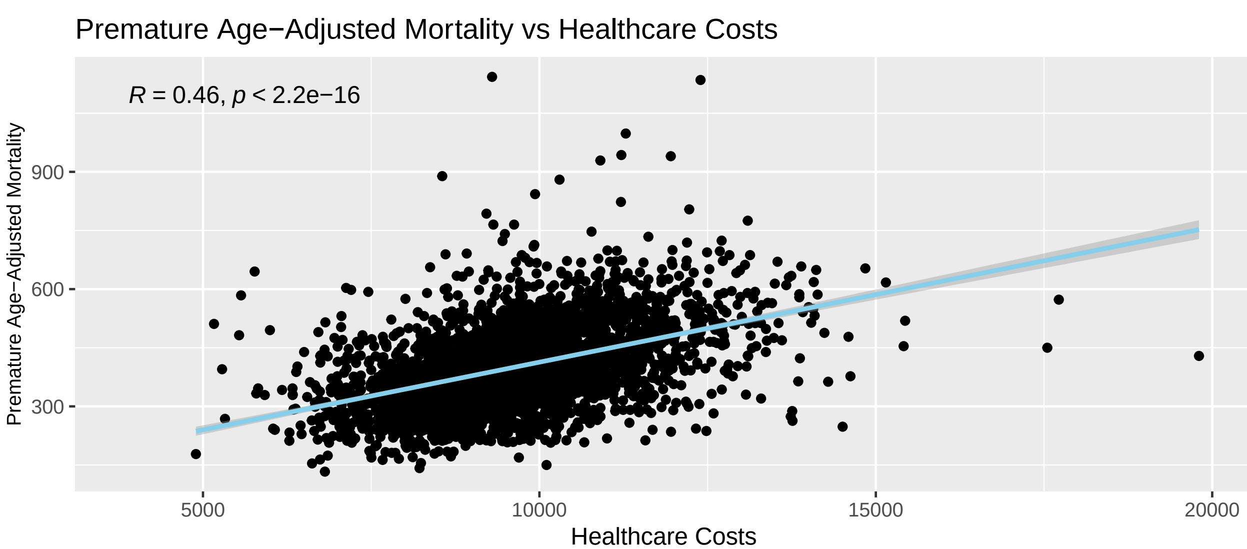

Figure 2: Premature Age Adjusted Mortality vs Healthcare Costs

Heathcare Costs

Figure 2 shows the relationship between rate of PAAM with the healthcare costs. A correlation coefficient of 0.46 indicates that a positive correlation between the two variables. This is highly significant as it shows that as healthcare costs rise so does the PAAM, and vice-versa which is corroborated by past literature. A P-value of $<2.2^{-16}$ tells us that the correlation is statistically significant and gives strong evidence to reject the null hypothesis of no correlation, further indicating that healthcare costs do infact impact the PAAM.

Median Household Income

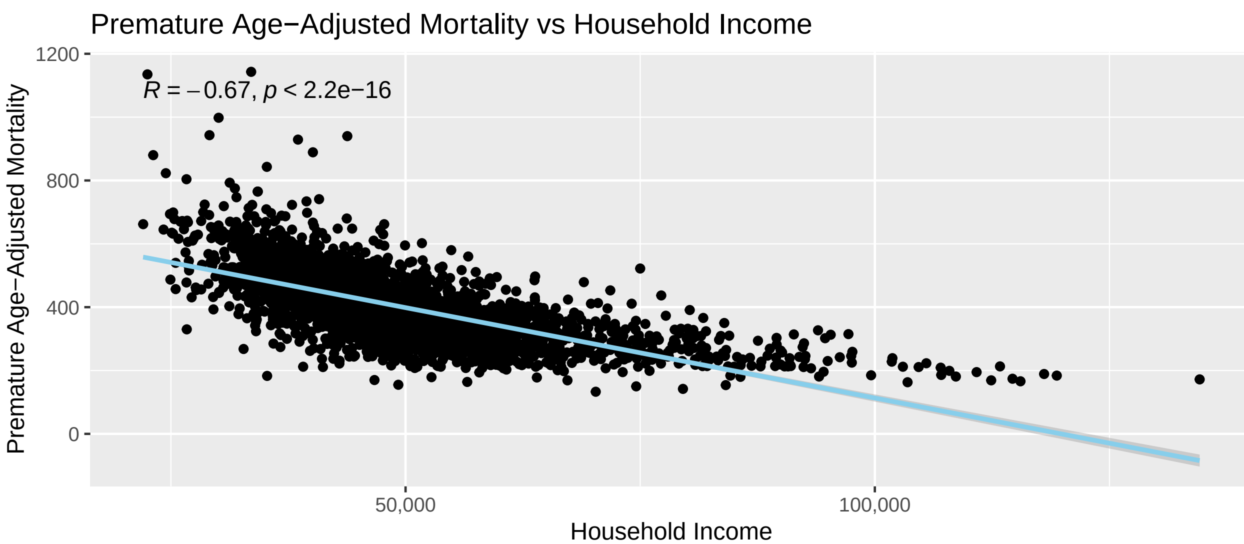

Figure 3: Premature Age Adjusted Mortality vs Household Income

Figure 3 shows a scatterplot between PAAM and median household income. The correlation coefficient in this case is -0.67 which indicates a negative correlation between the two variables. This tells us that household income increases, the mortality rate goes down. This is inline with previous studies and a P-value of $<2.2^{-16}$ tells us that this correlation is statistically significant and gives strong evidence to reject the null hypothesis and consider the alternate hypothesis justifying the fact that household income also impacts PAAM across counties.

These two scatterplots show a general linear relationship between the two selected independent variables(median household income and healthcare costs) justifying the selection of linear regression for the final model development. Linear regression has the added advantage of the results being easy to interpret. After scaling the variables and imputing the missing data, the final model intercepts and coefficients should be easier to understand.

Figure 4 below visualises the Premature Age-Adjusted Mortality as well as median household income across US counties. While this does not help us identify any particular geographic trends as to why PAAM is different in different regions, it does help us to visualise the high variations in both PAAM and household incomes across counties with some counties having above $125000 household income. This suggests that geographic location does not influence the selected variables and can be set aside for this study.

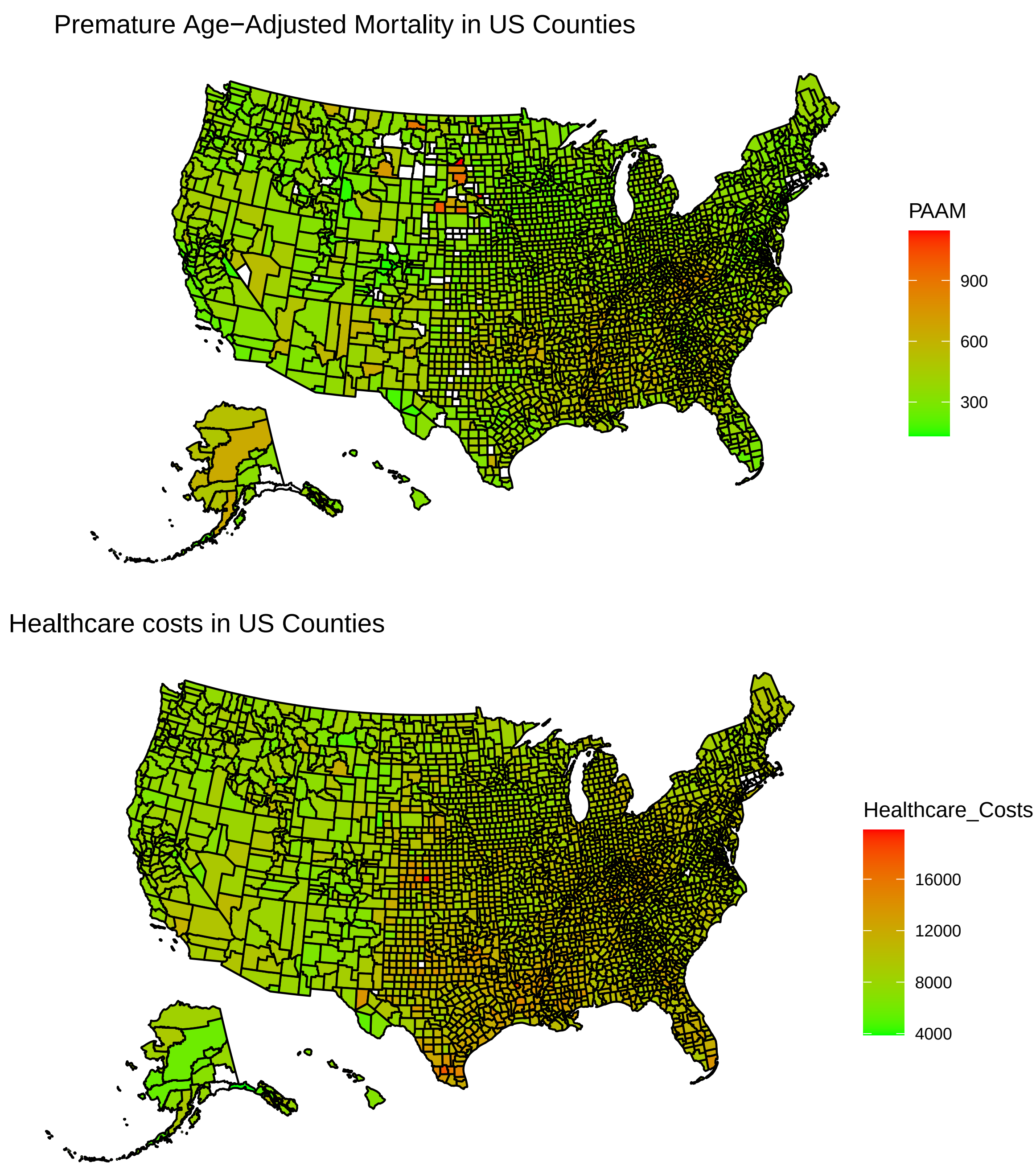

Figure 4: County level map of the United States of America

Table 2: Model Summary

	Coefficient	95%CI
(Intercept)	254.877***	(223.947, 285.806)
Household_Income	-35.294***	(-38.783, -31.804)
Healthcare_Costs	21.804***	(19.147, 24.462)
Education_College	-1.347***	(-1.628, -1.066)
Insufficient_Sleep	3.813***	(3.084, 4.543)
Food_Insecurity	6.776***	(6.006, 7.546)

Note: Significance levels: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 2 highlights the final model coefficients along with their Confidence Intervals(CI). Since the variables Household_Income and Healthcare_Costs have been standardized, the model outputs more interpretable results. The intercept is 254.87 which means that when all the variables are zero then the base PAAM is approximately 254. The coefficient for Household Income is -35.29 which means that unit increase in Household Income results in a 35.29 decrease in PAAM keeping other variables constant indicative of a negative correlation between these two variables. The CI for this variable is (-38.78,-31.80). On the other hand, the coefficient for Healthcare Costs is 21.80 which implies that a unit increase in Healthcare costs results in a 21.80 increase in PAAM with a CI of (19.15,24.46). Looking at the P-values for both the independent variables ($<2e^{-16}$), we see that both are highly significant to our dependent variable(PAAM). Keeping these results in mind, we can confidently reject the null hypothesis and consider the alternate hypothesis stating that Household Income and Healthcare Costs do impact PAAM across counties. This is inline with the literature, further validating my findings with this model.

Conclusion

This study presents strong correlation of household income and healthcare costs to the premature age-adjusted mortality. After controlling for mechanisms such as college education and some factors of general health (Insufficient sleep and food insecurity), the two independent variables still showed strong correlation to the dependent variable. This allows us to reject the null hypothesis and consider the theory that higher healthcare costs and lower median household income lead to increase in a county's Premature age-adjusted mortality.

These findings are assisted by the health literature which also show similar findings, claiming that higher healthcare costs result in individuals hesitant towards healthcare services[16]. Limitations of this study are, firstly, external confounding factors can potentially alter the results such as mental distress. Mental distress is a mechanism which cannot be easily quantified and has the potential for high amounts of measurement error. Secondly, this study uses data sources from reliable organisations but since there was no oversight into the collection of this data, there is a potential for some error here. Lastly, there is potential for a study with comparison of these variables over a period of a few decades with a 10 year gap to reveal more detailed findings.

The results highlight the importance of economic and healthcare policies in public health decision making. Policymakers and stakeholders should consider policies which enhance household incomes and reduce healthcare costs as part of the effort to improve county health.

References

1. Adler, N. E., Boyce, T., Chesney, M. A., Folkman, S., & Syme, S. L. (1994). Socioeconomic Inequalities in Health: No Easy Solution. JAMA, 269(24), 3140-3145.
2. Bartley, M. (1994). Unemployment and Ill Health: Understanding the Relationship. Journal of Epidemiology & Community Health, 48(4), 333-337.

3. Benach, J., et al. (2014). Employment, Work, and Health Inequalities: A Global Perspective. International Journal of Health Services, 44(1), 109-134.
4. McLaughlin, D. K., & Stokes, C. S. (2002). Income inequality and mortality in US counties: does minority racial concentration matter?. American journal of public health, 92(1), 99-104.
5. Deaton, A. (2013). The Great Escape: Health, Wealth, and the Origins of Inequality. Princeton University Press.
6. Grossman, M. (2006). Education and Nonmarket Outcomes. In E. A. Hanushek & F. Welch (Eds.), Handbook of the Economics of Education (Vol. 1, pp. 577-633). Elsevier.
7. Gulliford, M., Figueroa-Munoz, J., Morgan, M., Hughes, D., Gibson, B., Beech, R., & Hudson, M. (2002). What Does 'Access to Health Care' Mean? Journal of Health Services Research & Policy, 7(3), 186-188.
8. Cutler, D. M., & Lleras-Muney, A. (2006). Education and Health: Evaluating Theories and Evidence. NBER Working Paper No. 12352.
9. Kington, R., & Smith, J. P. (1997). Socioeconomic Status and Racial and Ethnic Differences in Functional Status Associated with Chronic Diseases. American Journal of Public Health, 87(5), 805-810.
10. Lynch, J. W., & Kaplan, G. A. (2000). Socioeconomic Position. In L. F. Berkman & I. Kawachi (Eds.), Social Epidemiology (pp. 13-35). Oxford University Press.
11. Lynch, J., et al. (2004). Income Inequality and Mortality: Importance to Health of Individual Income, Psychosocial Environment or Material Conditions. British Medical Journal, 320(7243), 1200-1204.
12. Marmot, M., et al. (2008). The Status Syndrome: How Social Standing Affects Our Health and Longevity. American Journal of Public Health, 98(9), 1651-1658.
13. McWilliams, J. M., Meara, E., Zaslavsky, A. M., & Ayanian, J. Z. (2007). Health of Previously Uninsured Adults after Acquiring Medicare Coverage. JAMA, 298(24), 2886-2894.
14. Rosenstock, I. M. (1974). The Health Belief Model and Preventive Health Behavior. Health Education Monographs, 2, 354-386. Shi, L., & Stevens, G. D. (2005). Vulnerability and Unmet Health Care Needs: The Influence of Multiple Risk Factors. Journal of General Internal Medicine, 20(2), 148-154.
15. Solar, O., & Irwin, A. (2010). A Conceptual Framework for Action on the Social Determinants of Health. Social Determinants of Health Discussion Paper 2 (Policy and Practice). World Health Organization.
16. Wilkinson, R. G., & Marmot, M. G. (2003). Social Determinants of Health: The Solid Facts. World Health Organization.

Table 3: Full Model Summary

```
##
## Call:
## lm(formula = PAAM ~ Household_Income + Healthcare_Costs + Education_College +
##     Insufficient_Sleep + Food_Insecurity, data = data_final)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -259.14  -38.03   -4.65   33.05  665.71
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        254.8769    15.7804  16.152   <2e-16 ***
## Household_Income   -35.2937     1.7804 -19.823   <2e-16 ***
## Healthcare_Costs    21.8044     1.3559  16.081   <2e-16 ***
## Education_College   -1.3468     0.1435  -9.389   <2e-16 ***
## Insufficient_Sleep   3.8132     0.3722  10.246   <2e-16 ***
## Food_Insecurity      6.7760     0.3929  17.248   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 66.52 on 3067 degrees of freedom
##   (69 observations deleted due to missingness)
## Multiple R-squared:  0.6322, Adjusted R-squared:  0.6316
## F-statistic:  1055 on 5 and 3067 DF,  p-value: < 2.2e-16
```