Estimating the Average Treatment Effect of Having Health Insurance on General Health

Charlize Samuels & Jamila Spears

[1] Emory University

Estimating the Average Treatment Effect of Having Health Insurance on General Health

Abstract

The aim of this paper is to investigate the average treatment effect (ATE) of health insurance on general health using data from IPUMS Health Surveys. In addition to our general health healthcare status indicators, we also include the covariates of race, sex, marital status, and poverty status in our linear regression adjustment model. Our findings demonstrate that individuals with health insurance tend to have lower health status in comparison to those without insurance. In addition, we found that being Black or of another race, female, married, or below the poverty line is associated with decreased likelihood of good health.

Our findings underscore the need to address disparities in health outcomes across different demographic groups, especially those who are historically more vulnerable. Access to health insurance remains an important part of promoting better health outcomes.

Introduction

In the United States, access to health insurance has been a topic at the forefront of the public consciousness for years. In the U.S., access to health insurance is likely to determine an individual's ability to afford necessary medical care (U.S. Department of Health and Human Services). This issue is particularly relevant when the country's diverse population and the disparities that exist across different demographic groups are taken into account.

Health insurance is a fundamental part of the healthcare system, since it provides financial protection against the often high costs of medical services and ensures access to essential treatments, including preventive care. Despite efforts to improve healthcare coverage through various governmental programs, disparities in insurance coverage

continue, disproportionately affecting certain demographic groups such as minorities, low-income individuals, and those with pre-existing health conditions (U.S. Department of Health and Human Services).

Because of this, we find it of high importance to study the impact of health insurance on general health. Access to comprehensive healthcare services is essential for maintaining and improving overall health outcomes (U.S. Department of Health and Human Services). Those with health insurance coverage are more likely to seek medical care when needed, leading to earlier diagnosis and treatment of health conditions, which in turn can prevent the progression of diseases and improve individuals' long-term health.

Furthermore, there are proven economic consequences of health insurance coverage, extending beyond just individual health. A healthier population is associated with higher productivity, reduced illness-related absences, and lower healthcare costs in the long run (Huang, Zhou, Xu, & Zheng, 2022). It also contributes to a more efficient healthcare system through reducing the burden of uncompensated care and emergency room visits for conditions that would have been preventable.

The consequences of the COVID-19 pandemic emphasized the importance of access to healthcare and health insurance (Huang, Zhou, Xu, & Zheng, 2022), highlighting existing disparities in healthcare access and outcomes and showing that the health of uninsured and underinsured individuals is more at risk.

So, this information demonstrates the continued importance of understanding the average treatment effect (ATE) of having health insurance on general health. This research aims to demonstrate that health insurance has a significant positive impact on an individual's overall health.

Literature Review

The Institute of Medicine (US) Committee on the Consequences of Uninsurance (2002) finds that access to health insurance has a positive effect on health outcomes. The results of later studies conducted by Fan Chenchen, Li Chunyan, Song Xiaoting (2024) and Michael Anderson, Carlos Dobkin, Tal Gross (2010) continue to support this finding.

"Care Without Coverage: Too Little, Too Late", written by The Institute of Medicine (US) Committee on the Consequences of Uninsurance (2002), is a comprehensive analysis of the impact of health insurance on access to healthcare and health outcomes through employing a systematic review of the results of existing literature on the topic. The literature used in the literature review was carefully vetted with specific selection criteria. The findings from the reviewed studies consistently support the idea that health insurance improves access to care, increases the likelihood of receiving appropriate care, and positively influences health outcomes.

This paper covers a wide range of studies and methodologies, providing a thorough understanding and well-rounded perspective of the topic. The findings from the reviewed studies are also relatively consistent, which helps to validate the positive impact of health insurance on access to care and health outcomes. The study also discusses the methodologies used in the reviewed studies, which is useful in understanding the strengths and limitations of each study.

However, it is important to note that since the reviewed studies are largely observational in nature, there may be inherent biases related to the selection of study participants and the measurement of health insurance status. In addition, observational studies cannot establish causality definitively, so while they show associations, they may not prove that health insurance directly causes improved health outcomes. Despite these limitations, this study still provides a strong theoretical and empirical foundation for the results of this study.

This year Fan Chenchen, Li Chunyan, Song Xiaoting (2024) wrote the paper: "The relationship between health insurance and economic performance: an empirical study based on meta-analysis". This paper aims to deepen our understanding of the relationship between health insurance and economic performance, such as poverty reduction, labor force enhancement, productivity, and healthcare system optimization. Using meta-analysis methodology, this study expanded on the results of previous experiments and used a scientific response in order to precisely determine the magnitude of the correlation between having health insurance and level of economic performance. This study found "a significant positive correlation between health insurance and economic performance" (13).

This study has a limited focus on qualitative sources and non-empirical literature in the meta-analysis, potentially excluding information that could deepen the analysis. However, the utilization of meta-analysis methodology to systematically review existing literature and provide comprehensive conclusions in addition to identifying moderating variables that influence the relationship between health insurance and economic performance are good systems to insure validity of findings. Economic health is an important part of overall health, so this study supports the previous paper's finding that having insurance has positive health effects.

Lastly, the paper "The Effect of Health Insurance Coverage on the Use of Medical Services" by Michael Anderson, Carlos Dobkin, and Tal Gross investigates the impact of health insurance on health care service utilization by taking advantage of quasi-experimental variation in whether or not an individual has insurance resulting from eligibility rules for dependents. Specifically, there are many health insurance contracts from private insurers that expire right after a teenager's 19th birthday, since the insurance only extends to dependents who are 18 and under. Thus, this study employs a regression discontinuity (RD) design to compare health care consumption of teenagers before and after losing insurance coverage. The results of this study indicate that losing health insurance coverage leads to decreased contact with health care providers, as evidenced by

sizable reductions in emergency department (ED) visits and even larger decreases in non-urgent hospital admissions.

Our ability to generalize the results beyond the specific population studied (young adults before and after losing insurance coverage) is limited. But, overall, this paper contributes valuable insights into how health insurance impacts health care utilization, and since we know that using health care should improve overall health, this study also supports the hypothesis that access to health care increased overall health.

In reviewing this literature on the impacts of health insurance on overall health, economic health, and use of medical services, we have found a rich foundation to support the theory that health insurance has positive effects on overall health.

Data & Covariate Choices

The data used for this research was collected from IPUMS Health Surveys. This data is harmonized from a number of health surveys across the country, simplifying the process of analyzing health trends across different populations and demographics in the United States. IPUMS also captures a vast range of variables derived from many different surveys. These variables range from health insurance access to one's Covid vaccination status. IPUMS also has an extensive and comprehensive list of demographic variables such as race, sex, socioeconomic status, and marital status, along with less traditional demographic variables such as number of weeks spent out of the workforce. This makes the IPUMS Health Survey a desirable data source as it allows researchers to analyze particular health disparities across different demographic groups, which is also one of the goals of this particular paper.

In our model, the response variable is HEALTH, the results of a sampling of adults and children whose overall health was rated from "excellent" to "poor" on a 5-point Likert scale. Individuals were also able to select "unknown" which was unrated.

We include a number of covariates that we believe are possibly correlated with our treatment variable: health insurance status. In the end, the four covariates race, sex, marital status, and poverty status are included in the model.

Race is commonly used as a covariate in health research because of the health disparities that typically exist across different racial groups. Underserved racial groups tend to have worse overall health, higher probability of chronic diseases, and less access to healthcare services than other groups.

Sex tends to be another standard covariate to include because we tend to see a difference in the health behaviors as well as the health outcomes between males and females. Due to the number of unique health issues that women face and the tendency for women to be misdiagnosed or diagnosed at a very slow rate, we believe that sex is an important covariate that could possibly confound the treatment assignment.

Marital status is an incredibly important covariate to include because we believe that there is likely a linkage between being married and having health insurance. This is because many individuals in the U.S may only have health insurance through their spouses, making marital status possibly correlated with the treatment assignment.

Lastly, poverty status is another vital covariate to include as it is known to be strongly linked to health care access. Individuals that fall on or below the poverty line tend to be underinsured, or even uninsured, because they cannot afford the price of health insurance. This lead us to include poverty status in the model as well.

In this paper, we assume race, sex, marital status, and poverty status to be an exhaustive list of variables that could possibly confound the treatment assignment. Therefore, with the inclusion of these covariates in our model, we assume unconfoundedness.

The inclusion of these covariates choices is meant to capture important core demographic information, such as race and sex, as well as possible socioeconomic barriers

that influence one's health insurance status. Including these covariates not only ensures that we have unbiased estimates, but it also allows us to see the disparities of these groups using interaction terms. In other words, the results of our estimates should produce an unbiased estimate of the ATE, while also allowing us to dig deeper into the magnitude of the differentials of the ATE between different demographics.

Model & Methodology

The process of estimating the ATE through linear regression adjustment essentially means that separate regressions are run for the treated and untreated groups. The coefficient on the treatment variable for the untreated group is then subtracted from the coefficient on the treatment variable for the untreated group. The difference is the ATE. This approach to estimating the ATE has its advantages relative to simply using the coefficient on the treatment variable from a single regression. For one, regression adjustments help to control for confounding. Fitting regressions for the treated and untreated group allows us to control for confounding through accounting for differences in characteristics between the two groups. This allows for unbiased estimates.

Regression adjustment also addresses interaction effects. The treatment may be different based on other variables (covariates). Running regressions for the treated and untreated groups separately, it is easier to account for interactions, and therefore, allow for a more accurate estimate of the ATE. The robustness of regression adjustments also made it an optimal choice for this analysis. This approach to estimating the ATE tends to be less sensitive to other sources of bias and even other issues such as model misspecification. This approach tends to be more flexible and allows for analyses that are a bit more refined and nuanced. In the last section, we established that the following covariates would be included in the model: Race, Sex, Marital Status, and Poverty Status. In addition, the outcome variable of this model will be General Health, and the treatment variable will be Health Insurance Status. The model is notated as such:

$$\hat{Y}_i = \hat{\alpha} + \hat{\theta} D_i + \hat{\beta}_1 X_{\text{race},i} + \hat{\beta}_2 X_{\text{sex},i} + \hat{\beta}_3 X_{\text{gender},i} + \hat{\beta}_4 X_{\text{marital status},i} + \hat{\epsilon}_i$$

where: - $\hat{Y}_i$ represents the predicted value of the general health outcome for individual i,

- D_i represents the insurance status for individual i,
- $\hat{\theta}$ represents the coefficient on the treatment variable
- $\hat{\beta}_1$ represents the coefficient for the effect of Race on the outcome,
- $\hat{\beta}_2$ represents the coefficient for the effect of Sex on the outcome,
- $\hat{\beta}_3$ represents the coefficient for the effect of Gender on the outcome, and
- $\hat{\beta}_4$ represents the coefficient for the effect of Marital Status on the outcome.

his model represents the overall approach of this research: to see how having health insurance impacts general health through estimating the ATE. However, this model does not represent the regression that will actually be run in our code. The covariates have been recoded as dummy variables due to their discrete, categorical nature.

Race has been separated into 3 dummy variables: White, Black, and Other. Sex has been separated into 2 dummy variables: Female and Male. Marital Status has been separated into 2 dummy variables: Married and Unmarried. Finally, Poverty Status was separated into 2 dummy variables, In Poverty and Not in Poverty. These dummy variables are coded as 1 if the observation identifies with the identification associated with that dummy variable (i.e an observation coded 1 for Female and 1 for In Poverty is a female who is at or below the poverty line).

It is important to note that the creation of these dummy variables means that we need to drop a single dummy variable from each stratum of covariates to represent the comparison group for that stratum. White, Male, Married, and Not in Poverty were

dropped from the Race, Sex, Marital Status, and Poverty Status strata, respectively. Consequently, these dropped dummy variables will be our baseline comparison group.

Therefore, the actual regression that will be ran is the following:

$$\hat{Y}_i = \hat{\alpha} + \hat{\theta} D_i + \hat{\beta}_1 X_{\text{black},i} + \hat{\beta}_2 X_{\text{other},i} + \hat{\beta}_3 X_{\text{female},i} + \hat{\beta}_4 X_{\text{unmarried},i} + \hat{\beta}_5 X_{\text{poverty},i} + \hat{\epsilon}_i$$

where:

- $\hat{Y}_i$ represents the predicted value of the general health outcome for individual i,
- D_i represents the insurance status for individual i,
- $\hat{\theta}$ represents the coefficient on the treatment variable
- $\hat{\beta}_1$ represents the coefficient for the effect of being Black on the outcome,
- $\hat{\beta}_2$ represents the coefficient for the effect of being Other on the outcome,
- $\hat{\beta}_3$ represents the coefficient for the effect of female on the outcome,
- $\hat{\beta}_4$ represents the coefficient for the effect of unmarried on the outcome, and
- $\hat{\beta}_5$ represents the coefficient for the effect of being on or below the poverty line on the outcome.

The model's treatment variable (health insurance status) is binary, coded as 1 if the observation has any type of health insurance coverage and 0 if they do not. The outcome variable, (general health) is also a dummy variable, coded as 1 if the observation has "good" health and 0 if the observation as less than "less than good health."

Because the outcome variable is binary, our model is considered a linear probability model (LPM). This means that in each regression run for the treated and untreated

groups, the coefficient for the treatment variable, which is θ in both regressions, will actually be a probability value. It represents the change in the probability of the outcomes for the treated and untreated groups. Therefore, the ATE can be calculated as:

$$ATE = (\hat{\theta}_{\text{Treated}} - \hat{\theta}_{\text{Untreated}})$$

This means the ATE will represent the difference in the change in the probability of the outcome occurring between the treated and untreated groups. It will essentially be the average treatment of the treatment on the probability of the outcome. For example, if the ATE was a value named P, that means that having insurance increases the probability of having good health by P units compared to the untreated group.

Results

The results of the ATE estimation shows that the ATE is -0.019. This indicates a decrease in probability of having good health for observations with health insurance compared to those that do not have health insurance. The estimate has a small p-value ($p < 0.001$), which is highly statistically significant; this is strong evidence against the null hypothesis of no treatment effect.

$\hat{\beta}_1$ is -0.028, indicating that being Black is associated with a decrease in probability of having good health compared to the reference group, which is White.

$\hat{\beta}_2$ is -0.049, which indicates that being a non-White race other than Black was associated with a decrease in the probability of having good health compared to the reference group.

$\hat{\beta}_3$ is -0.027, which indicates that being female is associated with a decrease in the probability of having good health compared to the reference group, which is male observations.

$\hat{\beta}_4$ is 0.007, which indicates that being unmarried is associated with an increase in the probability of having good health compared to the reference group, which is married observations.

$\hat{\beta}_5$ is -0.080, which indicates that being at or below the poverty line is associated with a decrease in the probability of having good health compared to observations that are above the poverty line.

All of these estimates are statistically significant ($p < 0.05$) indicating that these covariates have a statistically significant association with the probability of having good health.

The interaction term between the treatment and being Black, unmarried, or in poverty are -0.019, -0.011, and -0.095, respectively. This indicates that the effect of having health insurance on the probability of good health is lower for individuals that fall within those demographic groups.

The interaction term between the treatment and being Other or female is 0.035 and 0.024, respectively. This suggests that the effect of health insurance on the probability of having good health increases for individuals that fall within those demographic groups.

The estimates for the interaction effects are also all statistically significant at a 95% confidence level ($p < 0.05$), indicating that there are statistically significant differences in the interaction effect between the treatment and the being a part of different subgroups in the population.

Discussion of Results

The finding that individuals with health insurance may have a lower health status compared to those without insurance raises the necessity of investigating what might have caused this counterintuitive result.

Firstly, this association could be influenced by selection bias, the phenomenon where there is a systematic difference between individuals who choose to have health insurance and those who do not in ways that affect the outcome being studied. Those who purchase health insurance may already be predisposed to poorer health due to pre-existing conditions, chronic illnesses, or higher perceived health risks. This subsequent self-selection into insurance coverage based on health status can create a skewed comparison, making it appear that health insurance is associated with lower health status when, in contrast, underlying health conditions are the true cause of this relationship.

This result could also be caused by reverse causality. In contrast to selection bias, reverse causality directly refers to the direction of causality, meaning that poor health could also lead individuals to prioritize obtaining health insurance coverage. Individuals experiencing health challenges may recognize the importance of access to healthcare services and seek insurance to address their health needs. This reverse causality could contribute to the observed association between health insurance coverage and lower reported health status.

Lower reports of health status could also be caused by disparities in access to quality healthcare services among insured individuals. Health insurance provides access to medical care, but it is not insured that all medical care will be of good quality. Factors such incorrect or late diagnosis, and poor management of health conditions could result in lower reported health status.

Lastly, those who are insured may be more likely to seek care for less severe health concerns or preventive treatments, resulting in a higher reported frequency of health issues. Conversely, uninsured individuals may be deterred by the cost of treatment, leading them to seek treatment less frequently and thus report a higher health status.

Our results also indicate that being a race other than white, being female compared to male, and being at or below the poverty line compared to above the poverty line is

associated with reports of lower health status. This is consistent with what we would expect given that these underserved communities are more likely to face obstacles when it comes to treating health conditions such as discrimination and lack of resources.

In contrast, being unmarried is associated with an increase in the probability of having good health. At first, we expected a negative correlation between these two variables, reasoning that those who are married are likely more economically stable and thus would report a higher health status. Despite this, this result actually does make sense, simply because individuals who are unmarried are more likely to be younger and thus have less health issues.

In addition, the effect of having health insurance on the probability of a higher health status is lower for those who are Black, unmarried, or in poverty in comparison to their reference groups. So, this means that those who are White, married, or not in poverty, less vulnerable groups, are more likely to report good health with health insurance.

However, those who are female or fall into the category of Other race experience health benefits from having insurance. This could be due to the fact that in comparison to their reference groups, these demographics have more to gain from having health insurance, and thus report better health with health insurance.

Limitations

One major limitation of this study is the potential introduction of confounding variables due to the observational nature of the data. Despite our controlling for covariates such as race, sex, marital status, and poverty status, there still exists the potential for unmeasured factors that influence both health insurance status and health outcomes. For example, individuals likely have different ideas of what constitutes good health and bad health, introducing bias into the results. Objective measures of health such as clinical assessments could increase the validity of the findings.

In addition, establishing causality between health insurance and health outcomes is challenging. While regression adjustment helps control for confounding, it cannot eliminate the possibility of reverse causation or other causal pathways that may influence the relationship between health insurance and health status.

Finally, the linear regression adjustment used to estimate the ATE relies on certain assumptions about the functional form of the relationship between covariates and outcomes. Model misspecification or omitted variable bias could influence the accuracy of these estimates.

Conclusion

In conclusion, this study contributes to the wealth of research concerning the relationship between health insurance coverage and health outcomes. Through analysis of health survey data from IPUMS and using linear regression adjustment to estimate the ATE of health insurance on general health, we have been able to further investigate this complex issue.

Our findings suggest that even though individuals with health insurance may report a lower health status, there are several factors that may have contributed to this. These factors include selection bias, potential reverse causality, and disparities in quality healthcare access and utilization.

We also find that more vulnerable groups such as individuals who are Black, unmarried, and at or below the poverty line are less likely to experience the same level of health benefits from healthcare compared to those who are White, married, and not in poverty.

Knowing this, we must interpret the findings of this study while considering its limitations, including the observational nature of the data, the potential for confounding variables, and the challenge of establishing causality with observational data.

Our study highlights the need for policies that address disparities in healthcare access, especially for more vulnerable groups. Increasing the quality, accessibility, and affordability of healthcare services for all individuals, regardless of insurance status, should remain an important consideration for improving overall health outcomes and promoting equity in the healthcare sector.

Future research should continue to explore this complex relationship using longitudinal data, quasi-experimental designs, and robust methodologies in order to strengthen the ability to use causal inference and thus make more accurate and generalizable conclusions about the impact of health insurance coverage on health outcomes.

Tables and Graphs

Figure 1:

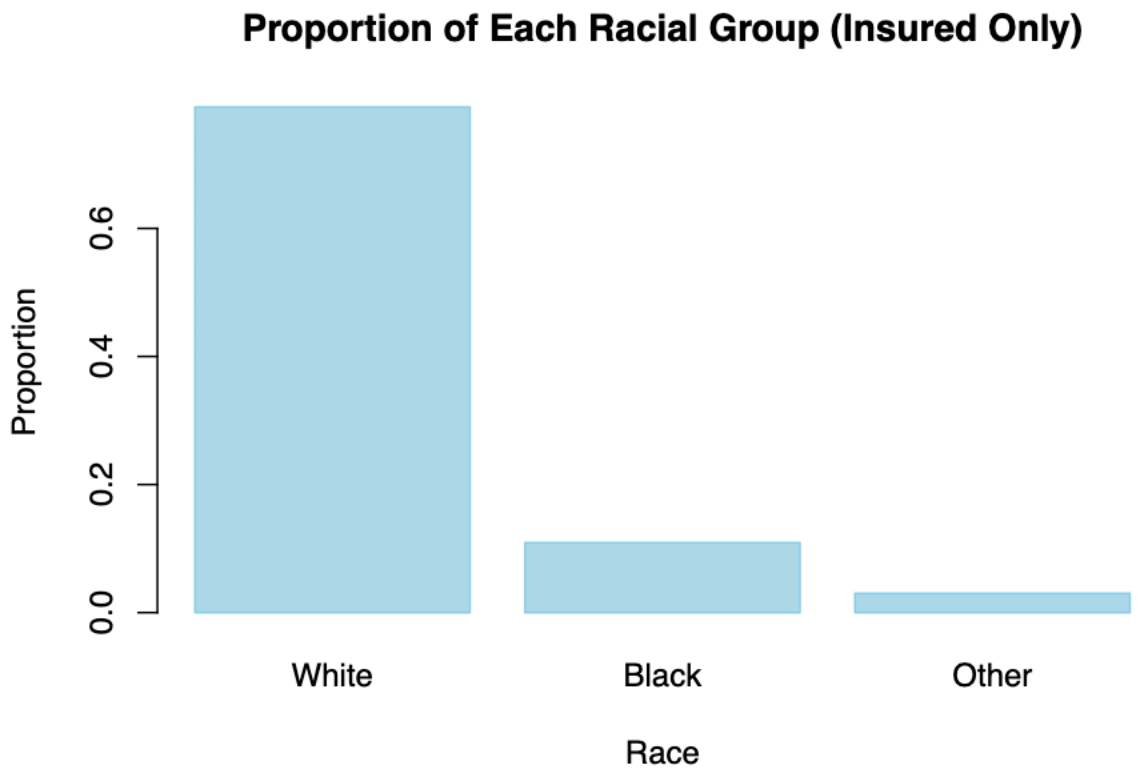


Figure 2:

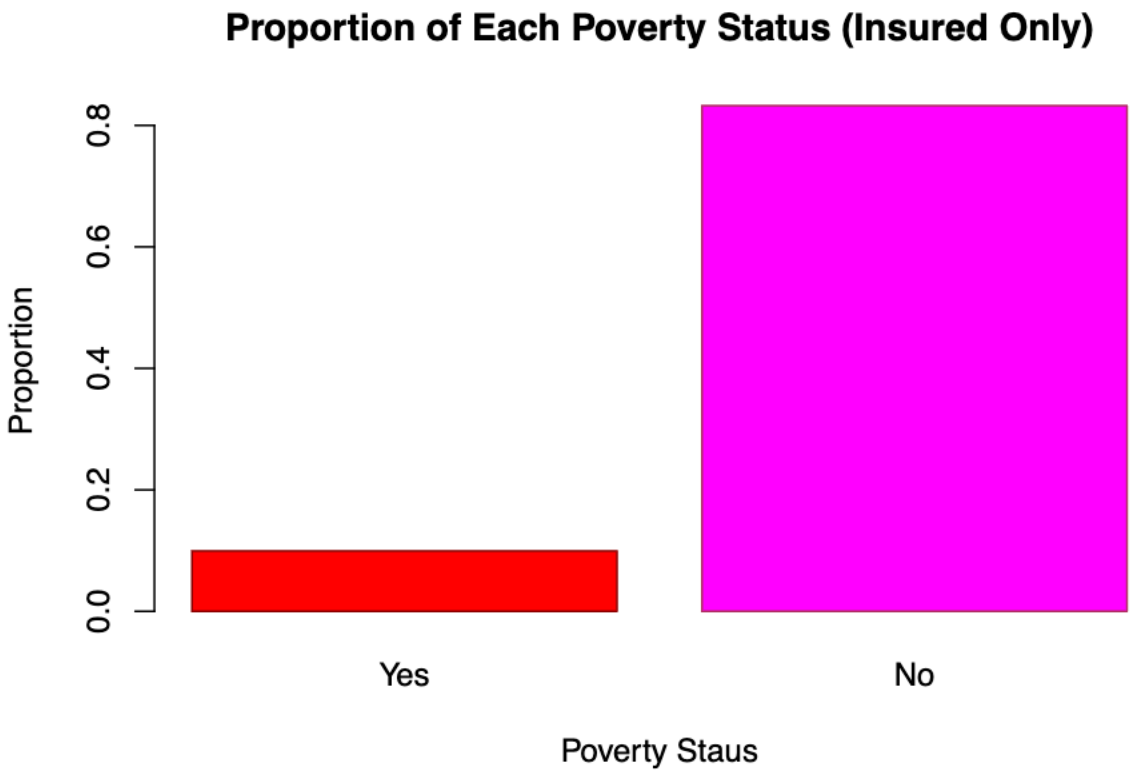
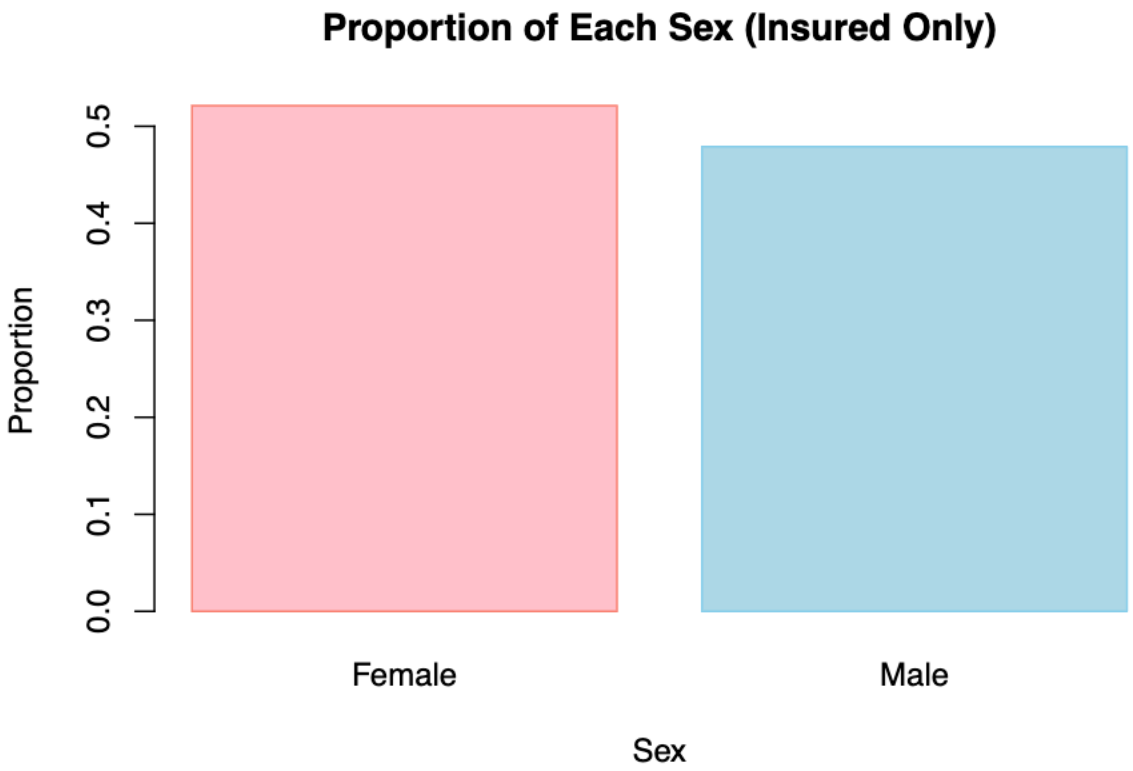
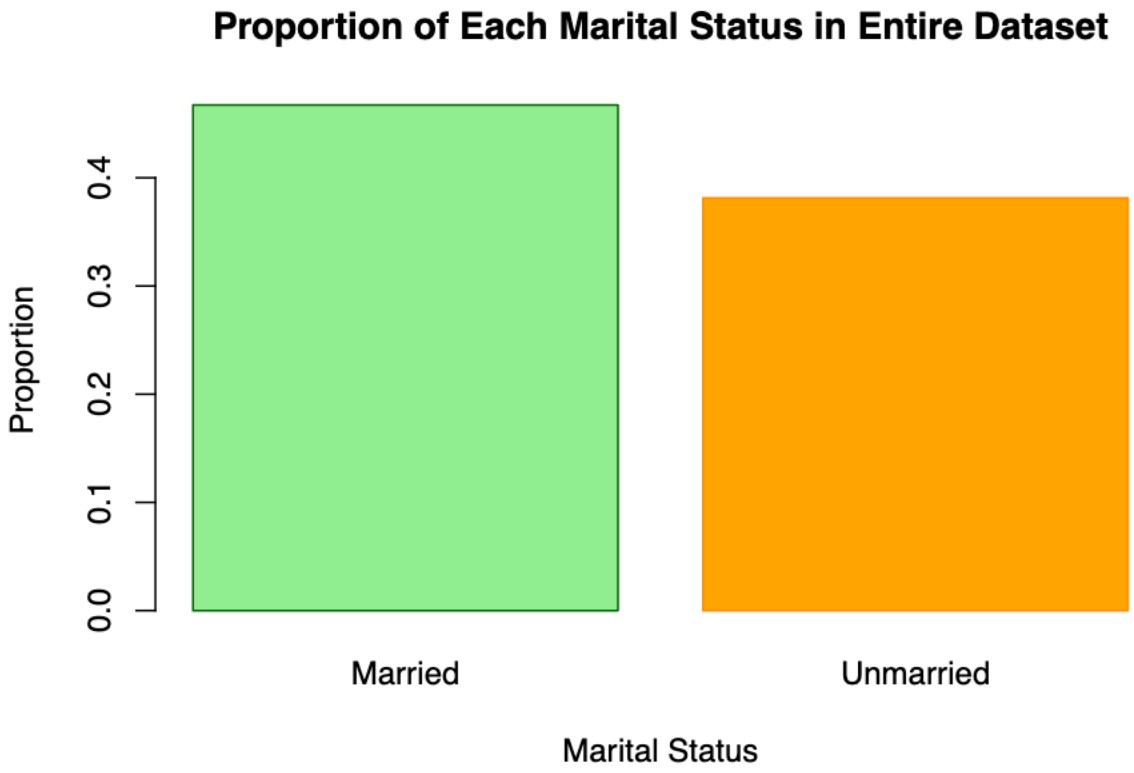

Proportion of Each Poverty Status (Insured Only)
Proportion
0.0
0.2
0.4
0.6
0.8
Yes
No
Poverty Staus

Figure 3:


Proportion of Each Sex (Insured Only)
Proportion
0.0
0.1
0.2
0.3
0.4
0.5
Female
Male
Sex


Proportion of Each Marital Status in Entire Dataset
Proportion
0.0
0.1
0.2
0.3
0.4
Married
Unmarried
Marital Status

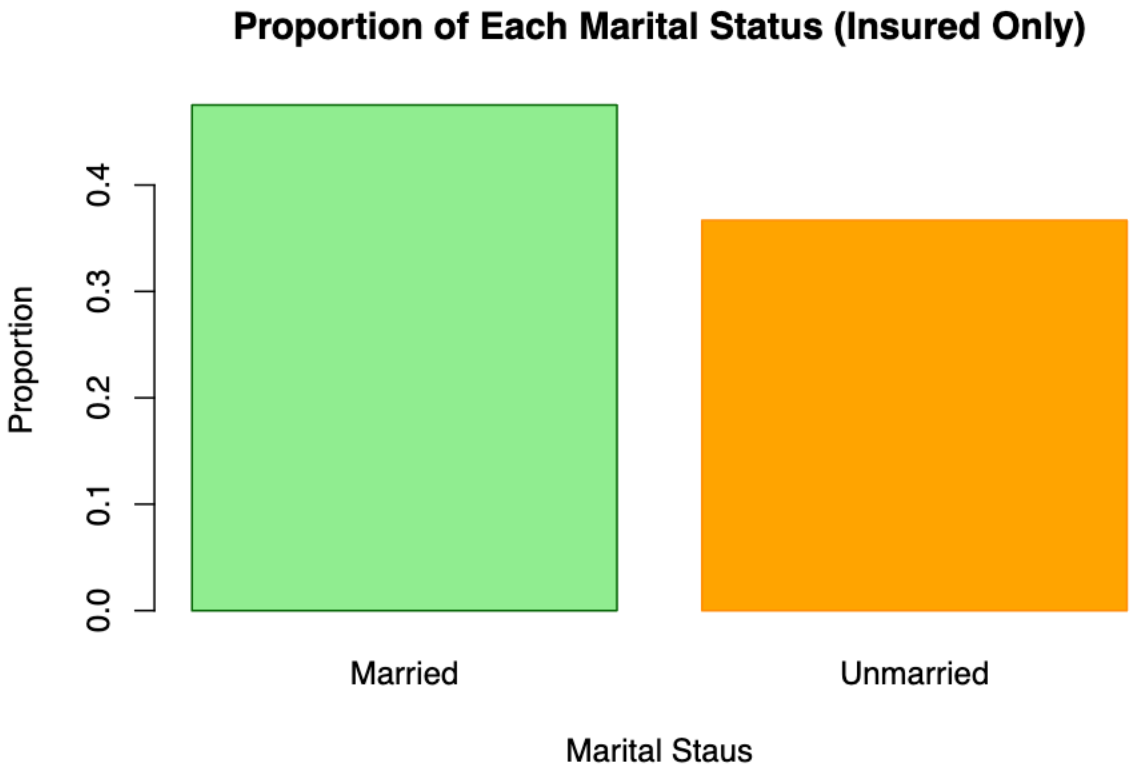


Table 1:

Variable	Correlation
White	0.030424751
Black	-0.027478578
Other	-0.006204035
Female	0.030523977
Male	-0.030523977
Married	0.048772490
Unmarried	-0.091764950
Poverty	-0.102777141
Not in Poverty	0.106523539

References

Anderson, M., Dobkin, C., & Gross, T. (2010). The effect of health insurance coverage on the use of medical services (NBER Working Paper No. 15823). National Bureau of Economic Research. Retrieved from https://www.nber.org/papers/w15823

Fan, C., Li, C., & Song, X. (2024). The relationship between health insurance and economic performance: an empirical study based on meta-analysis. Frontiers in public health, 12, 1365877. https://doi.org/10.3389/fpubh.2024.1365877

Huang, S., Zhou, T., Xu, C., & Zheng, J. (2022). Does Public Health Influence Economic Performance? Investigating the Role of Governance and Greener Energies for the Case of China. Frontiers in Public Health, 10, 864736.

Institute of Medicine (US) Committee on the Consequences of Uninsurance. (2002). Care Without Coverage: Too Little, Too Late. National Academies Press (US).

Office of Disease Prevention and Health Promotion. (n.d.). Access to Health Services. U.S. Department of Health and Human Services.