Supplement to Manuscript: All-cause Mortality Rates and Primary Care Physician Supply in US Counties 2021

Chaohua Li

2024-04-04

# 1. Exploratory analysis

‘Rate2’: age-adjusted all-cause mortality rate

‘pcp\_100k’: primary care physician supply per 100,000 population

‘metro’: metropolitan status of the county

‘hispanic\_pct’: proportion of county population that are Hispanic

‘NHB\_pct’: proportion of county population that are Non-Hispanic Black

‘noHS\_pct’: proportion of county population without a high school diploma

‘poverty\_pct’: proportion of county population living under federal poverty line

‘unemployed\_pct’: proportion of county population that are unemployed

‘uninsured\_pct’: proportion of county population without health insurance

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Data summary table.   | **Characteristic** | **N = 3,079** | | --- | --- | | Mortality rate | 1,376/368 | | PCP supply | 51/37 | | Metropolitan status | NA | | Metro | 1,162 / 3,079 (38%) | | Non-metro | 1,917 / 3,079 (62%) | | % Hispanic | 10/14 | | % NH-Black | 9/14 | | % No high school diploma | 12.1/5.8 | | % Poverty | 14.5/6.1 | | % Unemployed | 5.27/2.59 | | % No health insurance | 9.6/5.0 | |

|  |
| --- |
| Figure 1: Distributions in the nurmeric variables. |

From the histograms, we can tell the distribution for the outcome (mortality rate) is fairly normal. Distributions for PCP supply, %Hispanic and %NH-Black are very skewed. We probably have to log-transform PCP supply.

|  |
| --- |
| Figure 2: Scatterplot of Mortality Rate and Log(PCP Supply). |

There seems to be a linear relationship between Mortality Rate and log(PCP supply). Linear regression might be suitable for analysis.

|  |
| --- |
| Figure 3: Scatterplot of Mortality Rate and Log(PCP Supply) by Metro Status. |

The relationship looks similar for the metro and non-metro counties. Maybe there is no interaction by metro status.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Correlations among covariates.   |  | hispanic\_pct | NHB\_pct | noHS\_pct | poverty\_pct | unemployed\_pct | uninsured\_pct | | --- | --- | --- | --- | --- | --- | --- | | hispanic\_pct | 1.0000000 | -0.1152436 | 0.4163789 | 0.0744451 | 0.0437684 | 0.3863902 | | NHB\_pct | -0.1152436 | 1.0000000 | 0.3066258 | 0.4267227 | 0.3521652 | 0.1481195 | | noHS\_pct | 0.4163789 | 0.3066258 | 1.0000000 | 0.6063866 | 0.3739226 | 0.5454885 | | poverty\_pct | 0.0744451 | 0.4267227 | 0.6063866 | 1.0000000 | 0.5641002 | 0.3975236 | | unemployed\_pct | 0.0437684 | 0.3521652 | 0.3739226 | 0.5641002 | 1.0000000 | 0.2160019 | | uninsured\_pct | 0.3863902 | 0.1481195 | 0.5454885 | 0.3975236 | 0.2160019 | 1.0000000 | |

The correlations among the covariates are relatively low. We can keep all the covariates for modeling.

# 2. Model Performance

Model 1: Mortality rate as outcome, log\_PCP as the only predictor - RMSE=358

Model 2: Mortality rate as outcome, log\_PCP and all covariates - RMSE=292

|  |
| --- |
| Figure 4: Predicted vs Observed Values by Models with and without Covariates. |

|  |
| --- |
| Figure 5: Residuals vs Predicted Values by Models with and without Covariates. |

The full model with covariates has smaller RMSE, and predicted values closer to the observed values. Thus, Model 2 with all covariates is selected as the final model to describe the relationship between mortality rates and PCP supply.