All-cause Mortality Rates and Primary Care Physician Supply in US Counties 2021

Chaohua Li

2024-02-02

# 1. Part 1

# 2. Summary/Abstract

In this project, I will conduct a cross-sectional study on the county level and examine the potential association between age-adjusted all-cause mortality rate and primary care physician supply for US counties. A multilevel linear model will be used, and multiple potential confounders will be accounted for in the model.

# 3. Introduction

## 3.1 General Background Information

The all-cause mortality rate is an important indicator of general population health. Although the age-adjusted all-cause mortality rate consistently declined in the US from 1935 to 2014, a notable increasing trend after 2015 has been observed (1). Personal level access to primary care is highly associated with mortality, morbidity, and healthcare costs (2). There is a need to study how primary care physician supply on the geographic levels is associated with all-cause mortality rates in the US.

## 3.2 Description of data and data source

Three data sources will be used for this study. (1) Multiple Cause of Death File of 2021 from CDC WONDER (https://wonder.cdc.gov/mcd.html). I will obtain county-level age-adjusted all-cause mortality rates (deaths/100k people) from this source. (2) Area Health Resource File of 2021 from U.S. Department of Health & Human Services (https://data.hrsa.gov/data/download). I will obtain county-level primary care physician supply (physicians/100k people) and rural/urban categorization of the county from this file. (3) estimates of 2021 from 2017-2021 5-year American Community Survey (https://www.nhgis.org/). I will obtain states, regions, and some county-level socioeconomic measures as potential confounders for modeling. These measures may include %Living under poverty line, %Without a high school diploma, %Hispanic, %NH-Black, %Unemployed, %Without health insurance, and so on.

Data from the three sources will be merged by using the FIPS code which serves as the ID for counties.

## 3.3 Questions/Hypotheses to be addressed

My hypothesis is that primary care physician supply is negatively associated with age-adjusted all-cause mortality rate among US counties.Since counties can be considered nested within states or regions, a multilevel linear model will be used. The dependent variable will be mortality rate, main predictor will be primary care physician supply, and the socioeconomic measures will be the covariates in the model.

## 3.4 Reference

1.Woolf SH, Wolf ER, Rivara FP. The New Crisis of Increasing All-Cause Mortality in US Children and Adolescents. JAMA. 2023;329(12):975–976. doi:10.1001/jama.2023.3517 2.Peart A, Lewis V, Brown T, Russell G. Patient navigators facilitating access to primary care: a scoping review. BMJ Open. 2018 Mar 17;8(3):e019252. doi: 10.1136/bmjopen-2017-019252. PMID: 29550777; PMCID: PMC5875656.

To cite other work (important everywhere, but likely happens first in introduction), make sure your references are in the bibtex file specified in the YAML header above (here dataanalysis\_template\_references.bib) and have the right bibtex key. Then you can include like this:

Examples of reproducible research projects can for instance be found in [@mckay2020; @mckay2020a].

# 4. Methods

*Describe your methods. That should describe the data, the cleaning processes, and the analysis approaches. You might want to provide a shorter description here and all the details in the supplement.*

## 4.1 Schematic of workflow

## 4.2 Data aquisition

*As applicable, explain where and how you got the data. If you directly import the data from an online source, you can combine this section with the next.*

## 4.3 Data import and cleaning

*Write code that reads in the file and cleans it so it’s ready for analysis. Since this will be fairly long code for most datasets, it might be a good idea to have it in one or several R scripts. If that is the case, explain here briefly what kind of cleaning/processing you do, and provide more details and well documented code somewhere (e.g. as supplement in a paper). All materials, including files that contain code, should be commented well so everyone can follow along.*

## 4.4 Statistical analysis

*Explain anything related to your statistical analyses.*

# 5. Results

## 5.1 Exploratory/Descriptive analysis

*Use a combination of text/tables/figures to explore and describe your data. Show the most important descriptive results here. Additional ones should go in the supplement. Even more can be in the R and Quarto files that are part of your project.*

## 5.2 Basic statistical analysis

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. simple models with 1 predictor) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

## 5.3 Full analysis

*Use one or several suitable statistical/machine learning methods to analyze your data and to produce meaningful figures, tables, etc. This might again be code that is best placed in one or several separate R scripts that need to be well documented. You want the code to produce figures and data ready for display as tables, and save those. Then you load them here.*

# 6. Discussion

## 6.1 Summary and Interpretation

*Summarize what you did, what you found and what it means.*

## 6.2 Strengths and Limitations

*Discuss what you perceive as strengths and limitations of your analysis.*

## 6.3 Conclusions

*What are the main take-home messages?*

*Include citations in your Rmd file using bibtex, the list of references will automatically be placed at the end*

This paper [@leek2015] discusses types of analyses.

These papers [@mckay2020; @mckay2020a] are good examples of papers published using a fully reproducible setup similar to the one shown in this template.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 7. References