Analyzing Chronic Disease Indicators and Their Impact on Mortality in the United States

Annick Faal

January 2025

1 Introduction

In this project, I explored how chronic diseases such as diabetes and chronic obstructive pulmonary disease (COPD) contribute to mortality rates across different states in the US. COPD, which includes emphysema and chronic bronchitis, causes breathing difficulties, while diabetes is a metabolic disorder characterized by high blood sugar levels. This analysis aims to uncover how these diseases impact public health. By examining the relationship between these diseases and mortality, I seek to identify patterns that could help target prevention and intervention efforts.

2 Used Data

2.1 Chronic Disease Indicators Dataset:

Source: Chronic Dataset

This dataset contains information about chronic disease indicators across U.S. states. Key columns include:

- **Year**: The year the data was recorded.
- **State**: The state in which the data was collected.
- **Topic**: This represents the name of the chronic disease, e.g., diabetes.
- **ChronicDiseaseValue**: This value represents the age-adjusted death rate for the chronic disease.

2.2 NCHS Leading Causes of Death Dataset:

Source: NCHS Dataset

This dataset provides data on the leading causes of death in the U.S. Key columns include:

- **Year**: The year the data was recorded.
- **State**: The state in which the data was collected.
- **Cause Name**: This represents the name of the disease, e.g., diabetes.
- **Deaths**: The total number of deaths recorded for a specific cause in that year and state.
- **Age_Adjusted_Death_Rate**: The age-adjusted death rate for a specific cause of death.

2.3 Data Structure:

After merging the datasets, the following structure was obtained:

- **Year**, **State**, **Topic**, **ChronicDiseaseValue**, **Deaths**, **Age_Adjusted_Death_Rate**

Age-Adjusted Death Rate:

The age-adjusted death rate is a measure that helps compare death rates between different populations by removing the effect of differences in age. Since older people generally have higher death rates, this rate adjusts for those differences by using a standard age distribution. This makes it easier to compare death rates across populations, time periods, or regions, even if they have different age structures. This is why I use it more in my analysis.

3 Analysis

Data Focus and Scope: After data preparation, I was only able to get data focused only on Chronic Obstructive Pulmonary Disease (COPD) and Diabetes as mentioned in my data report.

3.1 1. Trend Analysis of Chronic Disease Mortality

Objective: The goal of this analysis was to examine the trends in age-adjusted death rates for Chronic Obstructive Pulmonary Disease (COPD) and Diabetes over the years, specifically from 2010 to 2017.

Method: I used the Year, Age_Adjusted_Death_Rate, and Deaths columns for both diseases from 2010 to 2017 to track how mortality rates have changed over time.

I created two separate plots to better understand these trends:

- **Age-Adjusted Death Rate Over Time**: This plot showed how the age-adjusted death rates for both COPD and Diabetes changed from 2010 to 2017.
- **Death Rates Over Time**: This plot displayed the raw number of deaths for each disease, giving a clearer picture of the actual mortality trends over the same period.

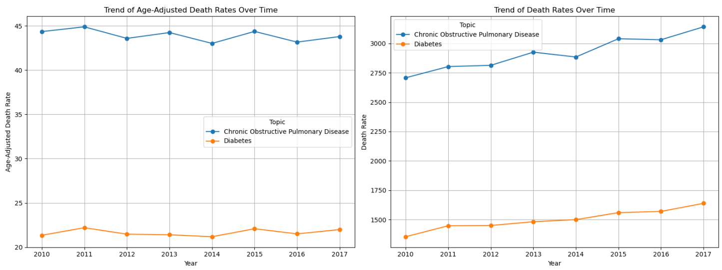

Figure 1: Age-Adjusted and Death Rate Over Time

Result & Interpretation:

- **COPD**: The age-adjusted death rate for COPD showed an overall increase from 2010 to 2017, with some fluctuations. The number of deaths also increased during this period, highlighting the growing burden of COPD as a public health issue.
- **Diabetes**: The age-adjusted death rate for Diabetes fluctuated but remained relatively stable. Deaths were lower than COPD but still significant and steadily increasing, suggesting that while awareness and control may have improved, Diabetes remains a major health concern.

COPD's rising mortality indicates a pressing public health challenge, while the stable trend in Diabetes death rates points to the need for ongoing management and intervention efforts for both diseases.

3.2 2. State-by-State Comparison of Chronic Disease Mortality

Objective: The goal of this analysis was to compare the mortality rates for COPD and Diabetes across different U.S. states from 2010 to 2017, identifying the states with the highest burden of these diseases.

Method: I aggregated the data by state and disease, calculating the average age-adjusted death rates for both COPD and Diabetes across the years 2010 to 2017. This allowed me to identify states with the highest mortality rates for each disease.

Result & Interpretation:

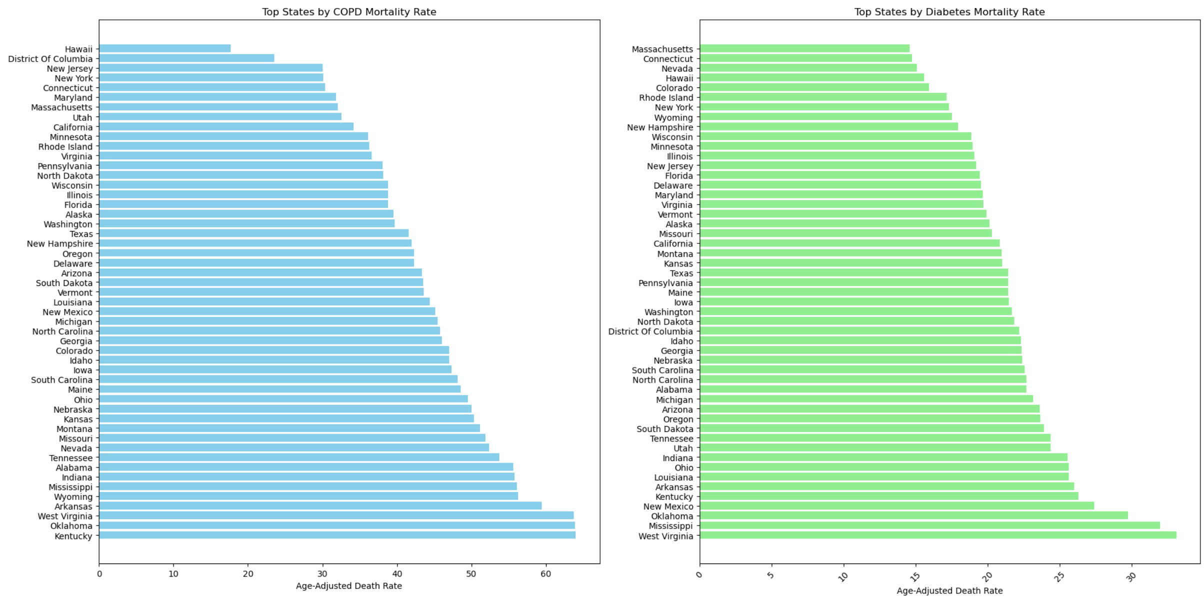

Figure 2: State Comparision

- **COPD**: States with the highest COPD mortality rates included Kentucky, Oklahoma, and West Virginia, with Kentucky leading at an average death rate of 63.98. These states show a significant burden from COPD, which may be linked to factors such as smoking rates and environmental conditions.

- **Diabetes**: West Virginia, Mississippi, and Oklahoma had the highest average diabetes mortality rates, with West Virginia topping the list at 33.08. These states are facing considerable challenges in managing and preventing diabetes, possibly due to socio-economic factors and limited healthcare access.

This comparison highlights the varying regional impact of both diseases and suggests the need for targeted public health interventions in states with the highest mortality rates.

3.3 3. Correlation Analysis

Objective: The goal of this analysis was to compare the correlation between COPD and Diabetes mortality metrics to understand their relationship over time.

Method: I aggregated the yearly deaths and age-adjusted death rates for COPD and Diabetes. Using the pandas library, I computed the correlation matrix to quantify the relationships between these variables. To enhance the interpretation, I used a heat map from the Seaborn library to visualize the correlation matrix, providing an intuitive view of the strength and direction of the relationships, where darker colors indicate a stronger correlation (positive or negative) and lighter colors represent weaker or no correlation.

Result & Interpretation:

- **Deaths_COPD and Deaths_Diabetes**: These have a very strong positive correlation (0.98), indicating that the number of deaths from both diseases tends to rise together.

- **Age_Adjusted_Death_Rate_COPD and Age_Adjusted_Death_Rate_Diabetes**: These have a moderate positive correlation (0.61), suggesting a less strong but still significant relationship in age-adjusted death rates between the two diseases.

The strong positive correlation between the deaths from COPD and Diabetes highlights the potential interrelated risk factors or healthcare challenges. The moderate correlation in age-adjusted death rates suggests that both diseases may share similar public health dynamics, warranting combined intervention strategies.

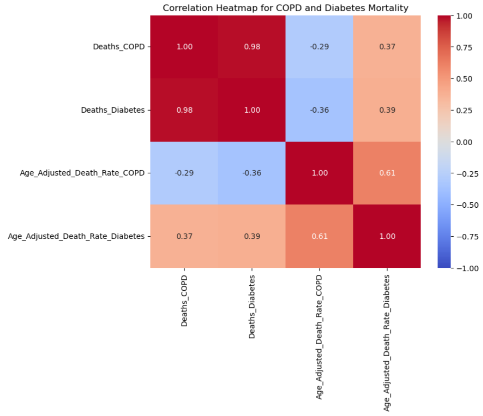

Figure 3: Correlation analysis

3.4 4. Policy Implication Analysis

Objective: To identify high-risk states with elevated Age-Adjusted Death Rates for chronic diseases and provide a visual representation to inform targeted public health policies and interventions.

Method: I identified high-risk states by filtering for those with Age-Adjusted Death Rates above the 75th percentile threshold to focus on the top quartile of states with the highest Age-Adjusted Death Rates. Using geospatial data, I merged this information with geographic boundaries and plotted it on a map. States with Age-Adjusted Death Rates above this threshold were highlighted in red, indicating areas that could benefit most from targeted public health interventions. This visualization provides a clear geographic focus for policy recommendations.

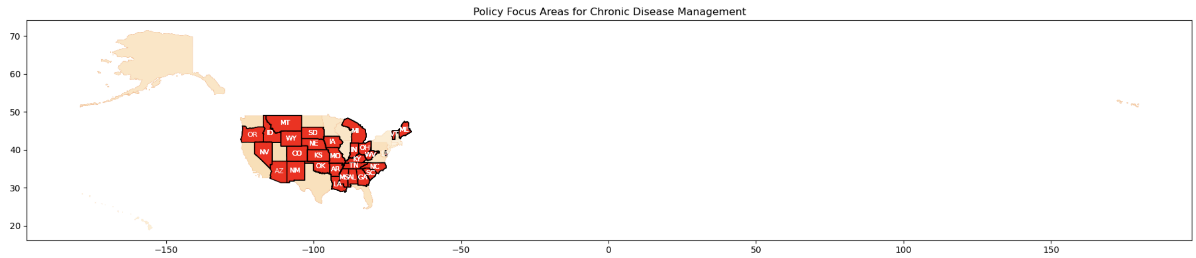

Figure 4: High Risk States

Result & Interpretation: The geospatial analysis highlighted high-risk states with age-adjusted death rates above the 75th percentile for chronic diseases. These states, shown in red on the map, are areas where public health interventions could have the most significant impact.

4 Conclusion

From my analysis, I gained insights into how COPD and Diabetes contribute to mortality rates. I found a strong correlation between deaths from these diseases, suggesting shared risk factors. I also identified high-risk states with elevated age-adjusted death rates, which could help direct public health resources more effectively.

However, there were limitations. The data revealed correlations but not causality, leaving out factors like socioeconomic status and healthcare access. Additionally, the analysis focused only on COPD and Diabetes, excluding other chronic diseases that might affect mortality. Expanding the scope to include more conditions and factors could provide a more comprehensive understanding. Despite these limitations, this project offers a valuable starting point for exploring the relationship between chronic diseases and mortality rates in the U.S.