1. **Introduction**

An accurate state-wise assessment of excess mortality during the COVID-19 pandemic is necessary to understand which health policies were more effective at reducing mortality. However, the numbers of reported COVID-19 deaths represent only a partial count of the total death toll from the COVID-19 pandemic. The number of excess COVID-19 pandemic deaths is defined as the difference between the number of deaths during the pandemic and the number of deaths that would have been expected if the pandemic had not happened. The expected deaths without COVID-19 are forecasted using historical data (i.e., death counts before the pandemic). Gaps exist between the reported and excess deaths related to the COVID-19 pandemic and we will observe how the gap decreases through time when the new tests and policies to detect COVID-19 were implemented.

Three different types of models were employed to forecast the expected number of deaths without COVID-19: (1) logarithmic trend with monthly dummy variables and autoregressive terms, (2) exponential smoothing, and (3) ARIMA models. The residuals from each of these models were utilized to fit another model using an autoregressive tensor for other causes of death. Additional information regarding changes in other causes of death could influence and enhance the forecast for a specific cause of death. For instance, if there is an increase in diabetes-related deaths in the last three months, this may affect heart disease deaths, potentially leading to either a slight decrease in deaths due to heart disease (competing risk between diseases) or a slight increase in deaths from heart disease (due to the same reason diabetes deaths have increased). The models will be based on only a few parameters because the amount of data available is limited; there are only 50 months of data for training the models. The models were run on the training set (the data from January 2015 till February 2019) and evaluated on a hold-out period before COVID-19 (March 2019 to February 2020). Once the best model was selected, that model will be rerun using data from January 2015 to February 2020, including the previous hold-out period so as to include the latest data in the model before the forecast.

1. **Methods** **and Results**
   1. **Data**

All deaths data by month and state were downloaded from the CDC1-2 mortality dataset from January 2015 to September 2023. Note that, since there is a 6 months lag for the assignment of non-natural causes of death (unintentional accidents and self-harm), the data for the last 6 months is not considered reliable for this cause of death. There are 50 observations for each state (i.e., for the 50 states and the District of Columbia) and 165 causes of death. This data can be used to build models to forecast the number of deaths for the 14 main causes of death (except COVID-19) and the combination of the other 151 causes of death for the pre-Covid period from January 2015 to February 2019.

Population data by state and year were collected from the Census3 for the period from 2015 to 2022; the 2022 population was used in 2023 because the 2023 data was not available.

* 1. **Excess mortality definition and metric to forecast.**

Excess mortality for the Covid-19 period from March 2019 to September 2023 is defined as the difference between the number of actual deaths and the number of forecasted deaths calculated as if the pandemic had not occurred:

Excess deaths = Actual deaths – Forecasted deaths without COVID (1)

In order to be able to compare across different states which have different population numbers, the crude death rate could be used instead:

Crude death rate by month = 100,000\* (Monthly deaths)/Population (2)

In this paper we shall use the daily crude rate instead of the monthly crude rate since it is more stable and smoother than the monthly crude rate.

Daily crude rate = Monthly crude rate / Number of days in the month (3)

Plot 1 compares the monthly death counts in the US from January 2015 to May 2023 to the monthly crude rates. The seasonality from both metrics is the same, but the crude rate does not increase as much as the number of deaths because it is divided by the population and thus takes into account population increases over time.

Plot 2 compares the daily crude rate in the US from January 2015 to May 2023 to the monthly crude rate. The trend is the same with both metrics and the seasonality is smoother using the daily crude rate than the monthly crude rate because the daily crude rate considers the number of days in each month.

The number of excess deaths will be estimated using the dependency between excess deaths and daily crude rate as expressed in equation (4).

Excess deaths = Observed deaths – Forecasted deaths without COVID

= Observed deaths - Forecasted daily crude rate \* Number of days in a month \*

Population/100,000