# Introduction

Smoke inhalation, whether it be from wildfire sources or not, can be extremely harmful leading to respiratory and cardiovascular issues in individuals. The Environmental Protection Agency (EPA) cites research that has provided evidence for increased risk in both respiratory and cardiovascular-related events due to wildfire smoke exposures and intensity of wildfire smoke [1].

Examining the impact of wildfire and smoke impact on a community may help inform local and federal government for new policies and guidelines for communities when wildfires and smoke become an issue. Examining this impact may also provide further justification for additional resources specifically for individuals who are susceptible to the negative health outcomes attributed to wildfire smoke exposure.

# Background

The EPA along with some of the medical community has investigated the impact of wildfire smoke exposures on health outcomes. The EPA specifically states that negative health impacts such as difficulty breathing, increased risk of asthma exacerbation and aggravation of other lung diseases, stroke, etc. are all attributed to exposure to wildfire smoke [2][3]. The EPA has also described the number of wildfires in the United States to be relatively stagnant, with consistent numbers close to 80,000 fires per year [4].

Our group has examined has previously developed a simple smoke estimation model based on United States Geological Survey (USGS) wildfire data. For our analysis, we sought to verify these statements by the EPA using our smoke estimation. Specifically, we ask the questions: (1) Are there any statistically significant associations between our smoke estimation and the prevalence of asthma and high blood pressure for the communities in Rochester, New York amongst all age groups? (2) What is the projected prevalence of asthma and high blood pressure for the next 10 years based on the smoke estimation for Rochester, New York communities amongst all age groups? To answer these questions, we acquired publicly available data from the Behavioral Risk Factor Surveillance System (BRFSS) from the Center of Disease Control (CDC). This data only provides data down to the state level in terms of granularity. Because of this, we use data associated with New York for our analysis for Rochester, New York.

Our group hypothesizes, based on the previous findings by the EPA and related parties, that there will be statistically significant positive associations with the smoke estimation and our selected health outcomes. We also hypothesize that prevalence of asthma and high blood pressure, based on the smoke estimate alone, should remain relatively the same because the number of fires has not significantly changed over time.

# Methods

Our analysis was performed in Python 3.10 with the following packages – polars, pandas, matplotlib, shapely, statsmodels, scipy, pyproj.

## Smoke Estimate

Smoke estimations for Rochester, New York were calculated based on USGS wildfire data over the period of 1961-2021. The USGS wildfire provides ring coordinates to describe an individual wildfire. We transformed these coordinates into WGS84 latitude and longitudes. Using the ring of coordinates, we constructed a polygon representation of the wildfire using the shapely package and calculated the distance to Rochester, New York using the centroid of the polygon of latitude and longitude coordinates.

Using both the distance of the fire and the total acres burned of the fire, we define the following as our smoke estimate:

$$Smoke\ Estimate = 100 \times \frac{Total\ Acres\ Burned}{Distance\ to\ City\ in\ Miles}$$

With the conditions that

$$Smoke\ Estimate = 1\ iff\ 100 \times \frac{Total\ Acres\ Burned}{Distance\ to\ City\ in\ Miles} < 1$$

$$Smoke\ Estimate = 500\ iff\ 100 \times \frac{Total\ Acres\ Burned}{Distance\ to\ City\ in\ Miles} > 500$$

This simplistic model was chosen as we believed that a larger fire would have a greater smoke output and that if it was closer to the city then the impact of smoke would be greater. We multiply this ratio by 100 and add upper and lower bounds of 1 and 500 to achieve a range of values comparable to the Air Quality Index (AQI).

Given our analysis was to be performed on an annual basis, we created a separate dataset which represented the average smoke estimate for each year. This was simply created by averaging all the smoke estimates we had from the USGS data by year.

## Forecasting Smoke Estimate

For predicting the prevalence of asthma and high blood pressure based on our smoke estimate, we needed to predict the smoke estimate of Rochester, New York. We implemented an Autoregressive Integrated Moving Average (ARIMA) model using the *Python* package *statsmodels* to predict the average smoke impact for the next 25 years (2022-2051).

To establish the optimal parameters for the ARIMA model, we performed an Augmented Dickey-Fuller test (ADF) to test whether our data of average yearly smoke impact was stationary or not. The reported ADF statistic was -5.393798563447438 with a p-value of 3.4902853946216193e-06 suggesting that yearly average smoke estimate data was stationary and did not require differencing.

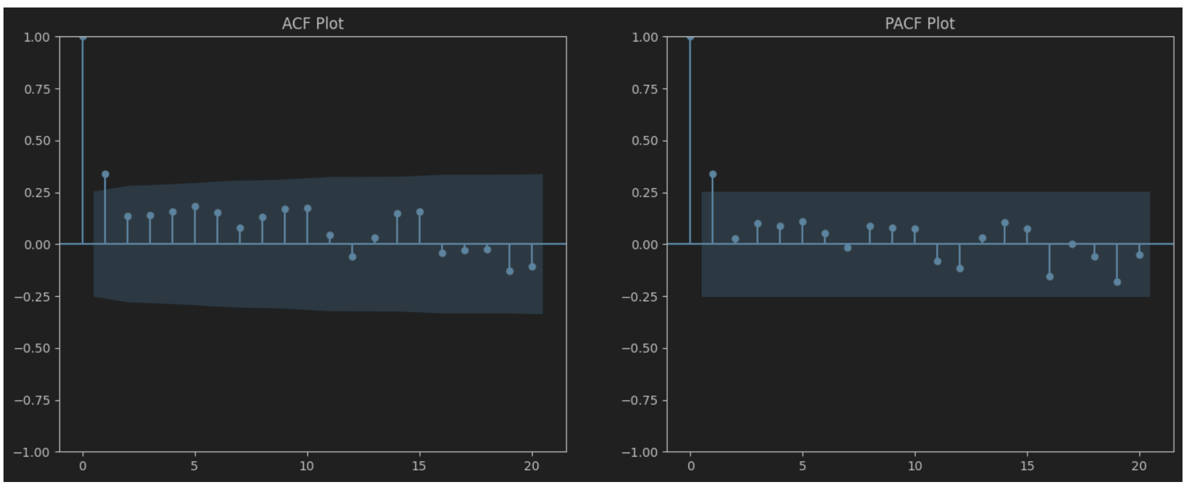



*Figure 1: ACF and PACF Plots*

We also plotted the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to determine the remaining parameters for the ARIMA model (Figure 1). We observed that for both the ACF and PACF plot there was a large spike at lag 1 and perhaps at lag 2. Through iteration we found that an ARIMA model with parameters *q=1, d=0, p=1* had the best-fit using AIC / BIC as the criteria.

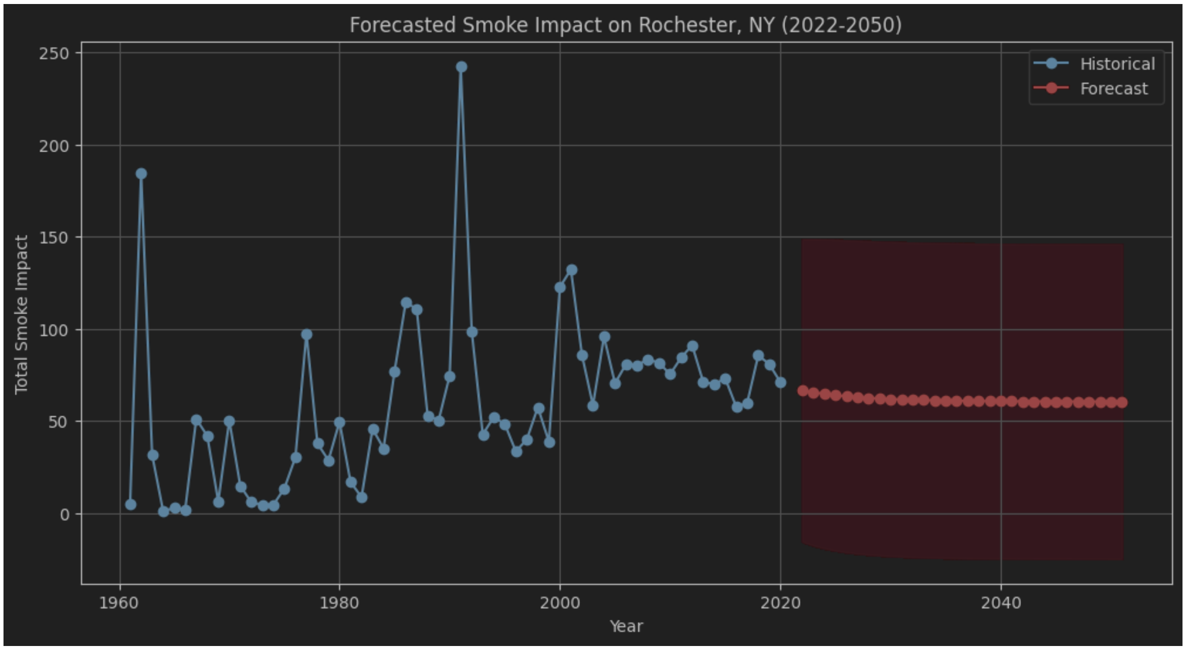



*Figure 2: Forecasted Smoke Estimate*

Using the ARIMA model, we then forecasted the smoke estimate for the next 25 years in Rochester, New York (Figure 2).

## Clinical Definitions

We defined Asthma prevalence as the percentage of individuals who answered "Yes" to the question of "Adults who have been told they currently have asthma". We defined High Blood Pressure prevalence as the percentage of individuals who answered "Yes" in the Topic of "High Blood Pressure". This was done because Asthma had two questions, one regarding past diagnosis and one of present diagnosis, while High Blood Pressure only had one question.

## Analyzing Prevalence Associations and Forecasting

We used simple linear regression in *statsmodels* to analyze the association between the smoke estimation and the prevalence of our chosen health outcomes in New York. We then used the established linear regression model to predict the prevalence of the health outcomes based on forecasted smoke estimates from the previous ARIMA model for the next 10 years (2020 - 2030)

# Results/Findings

## Forecasted Increase in Asthma Overall

We performed simple linear regression in order to predict the prevalence of Asthma in New York. This regression analysis identified that the overall population would see an increase in Asthma prevalence in the next decade based on our smoke estimate (Figure 3).

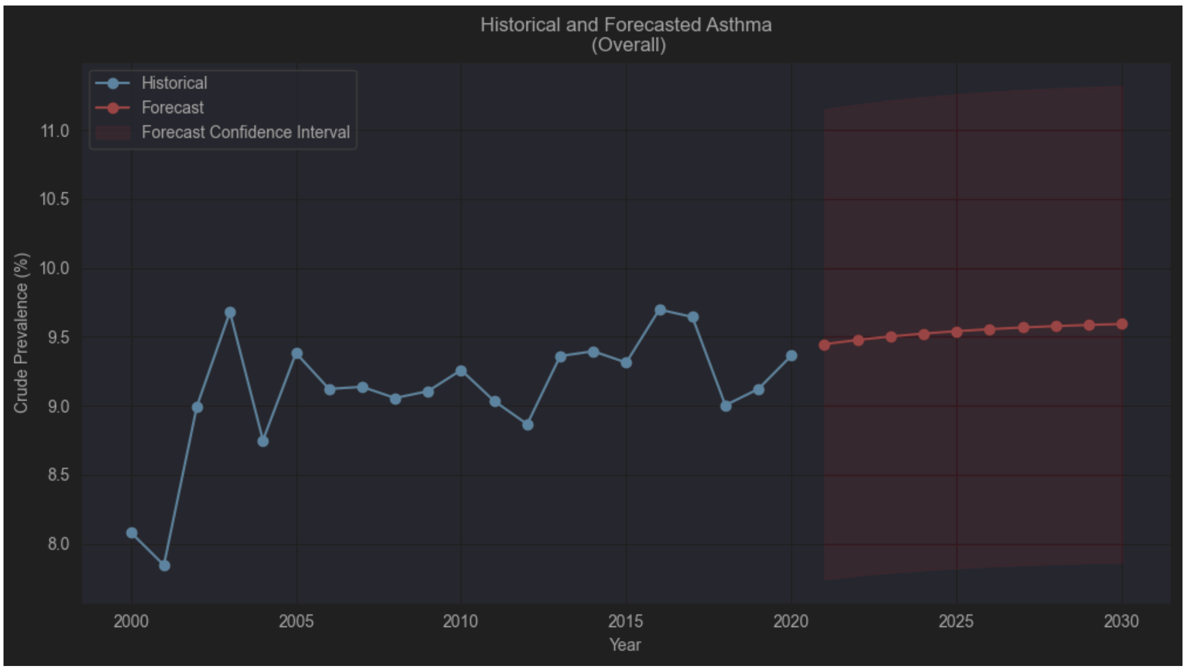



*Figure 3: Forecasted Asthma Prevalence*

The regression analysis identified specific increases for age groups of 35-44, 55-64, and 65+ (Supplementary Data 1-3).

## Significant Negative Associations Between Asthma and Smoke Estimate

We performed simple linear regression in order to identify if our smoke estimate was statistically significantly associated with the prevalence of Asthma in New York. The regression analysis identified significant negative associations (of p-value $< 0.05$), which were unexpected. For the age groups of 35-44, 55-64, 65+, and the overall group, statistically significant negative associations were found, suggesting that a decrease in the smoke estimate is associated with an increased prevalence of Asthma (Table 1).

**Table 1: Statistically Significant Negative Associations: Asthma**

| Age Group | Coefficient | P-Value |
|---|---|---|
| 35-44 | -0.0307 | 0.008 |
| 55-64 | -0.0369 | 0.005 |
| 65+ | -0.0287 | 0.044 |
| Overall | -0.0249 | 0.017 |

## Forecasted Decrease in High Blood Pressure

We performed simple linear regression in order to predict the prevalence of High Blood Pressure in New York. This regression analysis identified that the overall population would see a slight decrease in High Blood Pressure in the next decade based on our smoke estimate (Figure 4).

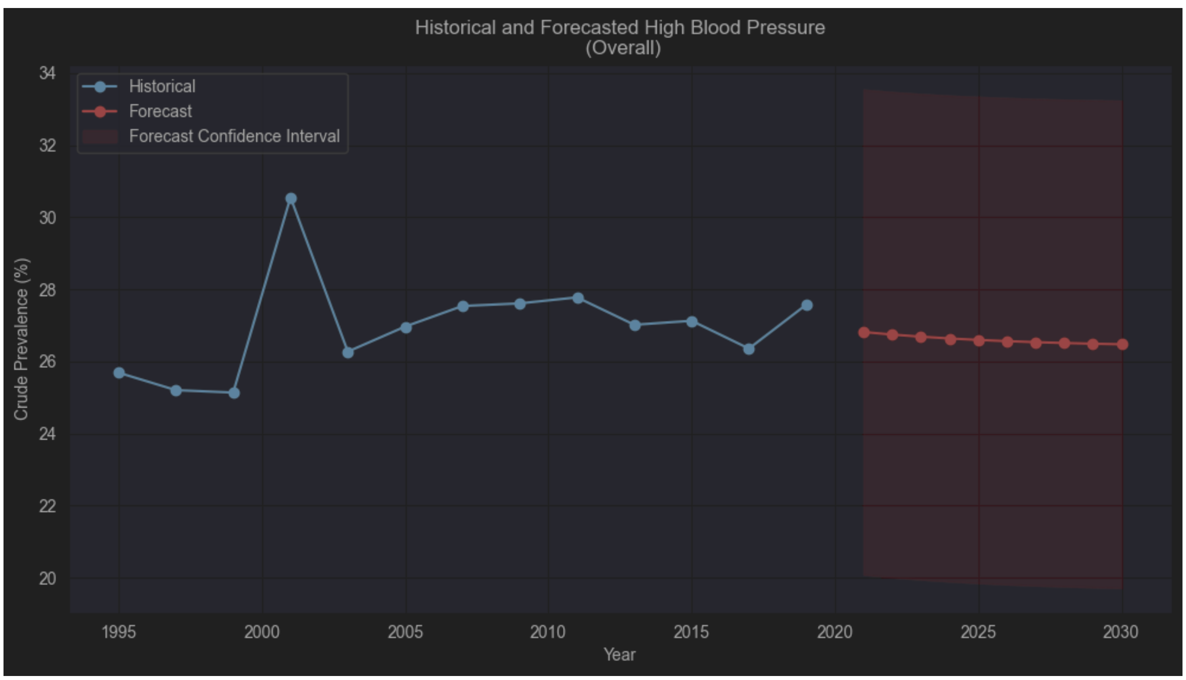



*Figure 4: Forecasted High Blood Pressure Prevalence*

The regression analysis also identified specific decreases for age groups of 55-64 and 65+(Supplementary Data 4-6).

## Borderline Significant Positive Associations Between High BP and Smoke Estimate

We performed simple linear regression in order to identify if our smoke estimate was statistically significantly associated with the prevalence of High Blood Pressure in New York. The regression analysis identified no statistically significant associations (of p-value < 0.05), but did have some borderline statistically significant associations For the age groups of 55-64 and 65+, there was a borderline positive statistically significant associations and a statistically significant positive association, suggesting that an increase in smoke estimate is somewhat associated with an increased prevalence in High Bood Pressure depending on age group (Table 2).

**Table 2: Borderline Positive Statistically Significant Associations: High Blood Pressure**

| Age Group | Coefficient | P-Value |
|---|---|---|
| 55-64 | 0.0892 | 0.053 |
| 65+ | 0.1657 | 0.021 |
| Overall | 0.0580 | 0.127 |

# Discussion

The results of our study identify a forecasted a slight continuous increase in the prevalence of Asthma and a slight decrease in prevalence of High Blood Pressure in New York based on our smoke estimates.

The findings of a forecasted increase in Asthma suggest the need for additional resources and guidelines regarding wildfire smoke exposures and Asthma. A recommendation could be investing in additional public health campaigns to educate the community to help mitigate exposure to wildfire smoke and manage symptoms in the event of exposure.

While High Blood Pressure prevalence was forecasted to decrease, a statistically significant positive association in individuals 65+ suggest the need for targeted preventative measures for elderly individuals in New York. A recommendation for the city/mayor/public health agencies is to create new policies and health organizations targeted for the elderly with the focus on reaching these elderly populations and monitoring their health. Such programs can allow health professionals to provide quicker assistance and readily available treatment for elderly individuals who may experience High Blood Pressure or any of its associated downstream negative outcomes (e.g. stroke, heart attack, etc.) due to wildfire smoke exposure in the local community.

# Limitations

Unfortunately, there are many limitations to our study. One of the most prevalent limitations to our study is our smoke estimate. In the results, we observed a negative association between Asthma prevalence and our smoke estimate which was unexpected. This unexpected result may indicate that our smoke estimate lacks in accuracy when it comes to modeling how wildfire smoke disperses across land. This is certainly plausible as we do not incorporate any additional weather-related or geographic-related factors in our calculation of the smoke estimate which can significantly impact how smoke travels. A more accurate model for smoke estimation could have included data on wind directions in the fire locale and local geography to determine whether the smoke would have travelled towards the city or not.

The unexpected negative association can also be attributed to our simplistic regression model. Our analysis implemented a simple linear regression using our smoke estimate as its only variable. This means that our regression analysis does not handle any potential confounding variables that may impact Asthma prevalence.

Another limitation is the lack of granularity in the BRFSS data. For the calculation of our smoke estimate, we specifically used the city of Rochester, New York. However, the BRFSS data does not have data that contains city-level data and instead only goes down to the state-level. Due to this, we went under the assumption that the Asthma and High Blood Pressure prevalence for Rochester, New York was similar to the state-level prevalence of these outcomes.

Lastly, we were limited on the amount of data to create the regression models on. The BRFSS only began collecting annual data in the mid-1990s. This meant that there was a limited number of data points for our regression analysis and forecasts which may harm the accuracy of the forecasts and associations we observed.

# Conclusion

In this study we sought to answer the questions: (1) Are there any statistically significant associations between our smoke estimation and the prevalence of asthma and high blood pressure for the communities in Rochester, New York amongst all age groups? (2) What is the projected prevalence of asthma and high blood pressure for the next 10 years based on the smoke estimation for Rochester, New York communities amongst all age groups?

We hypothesized that there would be statistically significant positive associations with the smoke estimation and our selected health outcomes. Additionally, we hypothesized that prevalence of asthma and high blood pressure, based on the smoke estimate alone, should remain relatively the same because the number of fires has not significantly changed over time.

Our findings challenge our hypotheses. The forecasted increase in Asthma prevalence and slight decrease in High Blood Pressure prevalence in the next 10 years for (Rochester) New York highlight a potential divergence in these two outcomes that may warrant additional exploration. In addition, our results found an unexpected negative association between our smoke estimate and Asthma prevalence which raise questions about unaccounted confounders and limitations of our smoke estimate model. For High Blood Pressure, our findings showed only one age group (65+) with statistically significant associations, suggesting merit in targeted interventions for elderly individuals.

# References

[1] US EPA, O. (2019, August 13). *Why Wildfire Smoke is a Health Concern* [Overviews and Factsheets]. https://www.epa.gov/wildfire-smoke-course/why-wildfire-smoke-health-concern

[2] US EPA, O. (2019, August 13). *Health Effects Attributed to Wildfire Smoke* [Overviews and Factsheets]. https://www.epa.gov/wildfire-smoke-course/health-effects-attributed-wildfire-smoke

[3] US EPA, O. (2019, August 13). *Which Populations Experience Greater Risks of Adverse Health Effects Resulting from Wildfire Smoke Exposure?* [Overviews and Factsheets]. https://www.epa.gov/wildfire-smoke-course/which-populations-experience-greater-risks-adverse-health-effects-resulting

[4] US EPA, O. (2016, July 1). *Climate Change Indicators: Wildfires* [Reports and Assessments]. https://www.epa.gov/climate-indicators/climate-change-indicators-wildfires

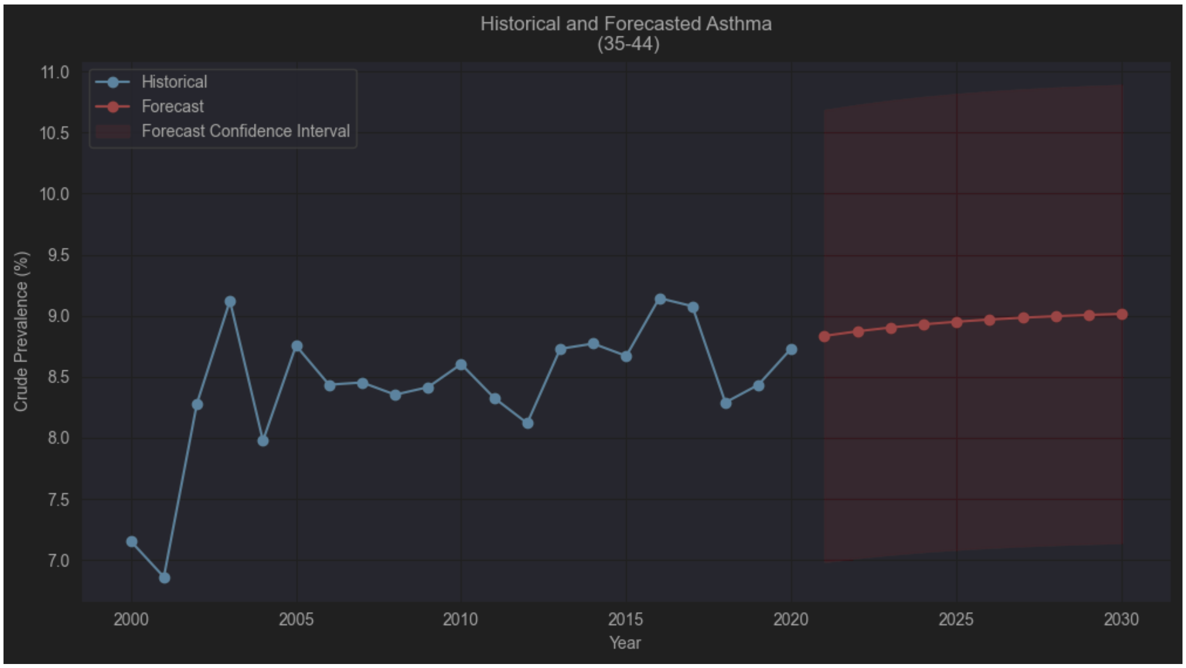
# Data Sources

USGS Data: Combined wildland fire datasets for the United States and certain territories, 1800s-Present (combined wildland fire polygons)

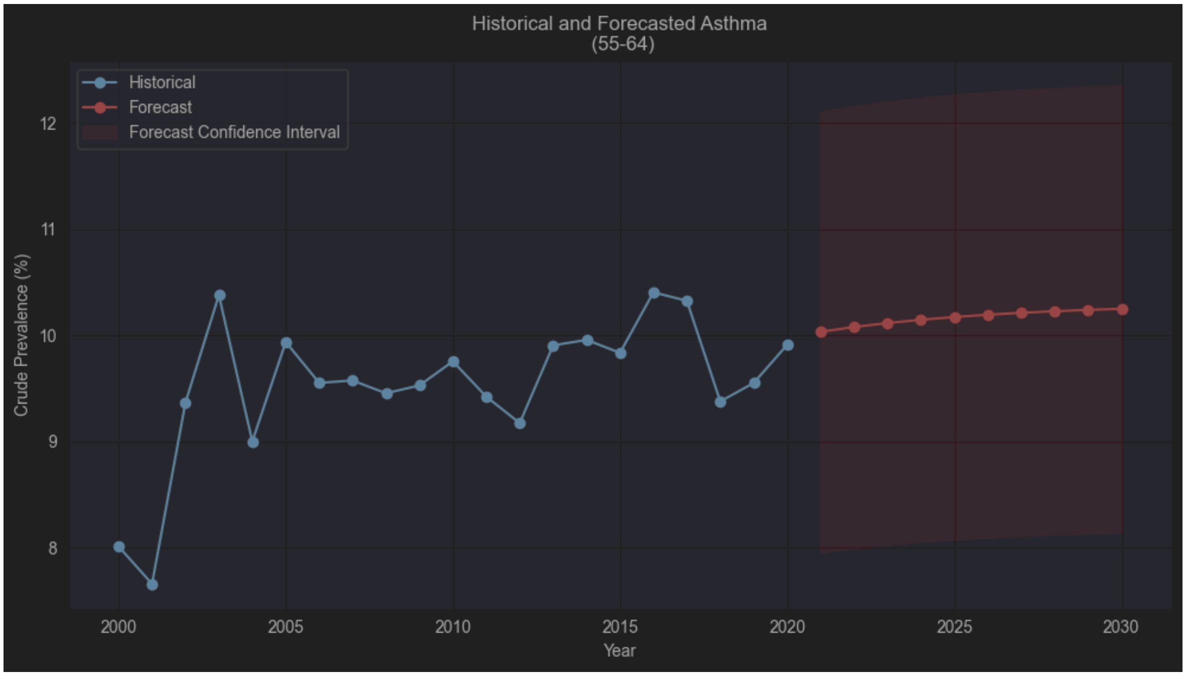
BRFSS Prevalence Data (2011-present): https://data.cdc.gov/Behavioral-Risk-Factors/Behavioral-Risk-Factor-Surveillance-System-BRFSS-P/dttw-5yxu/about_data

BRFSS Prevalence Data (2010 and prior): https://data.cdc.gov/Behavioral-Risk-Factors/Behavioral-Risk-Factor-Surveillance-System-BRFSS-P/y4ft-s73h/about_data

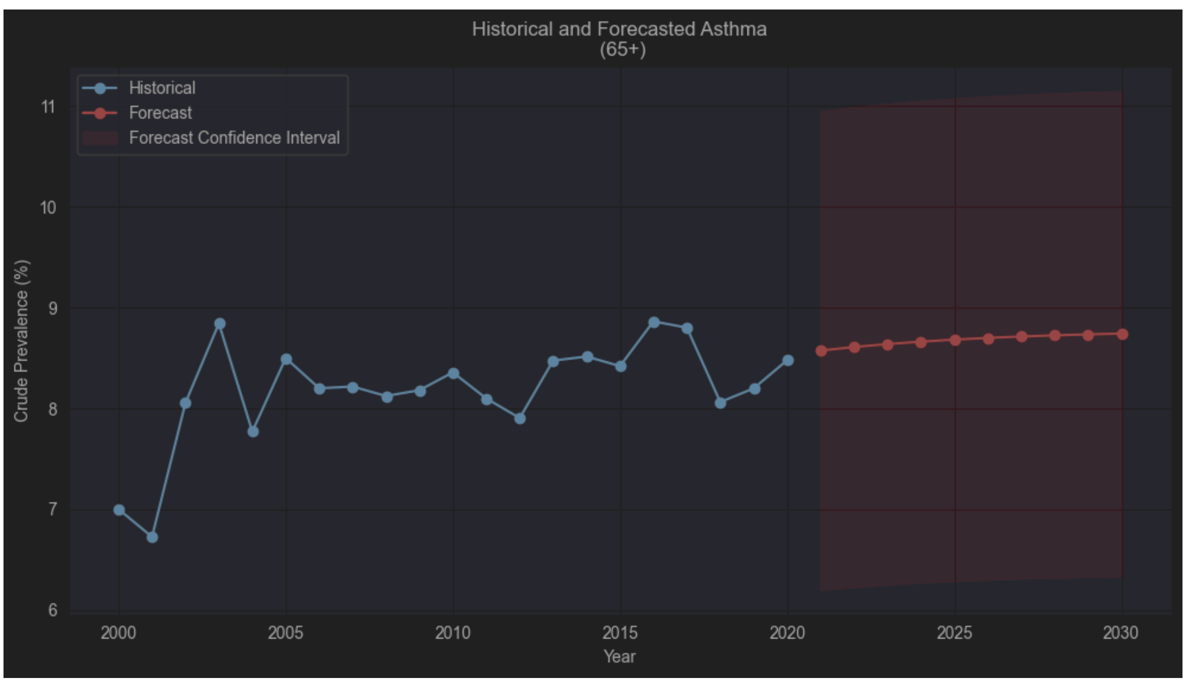
# Supplementary Data



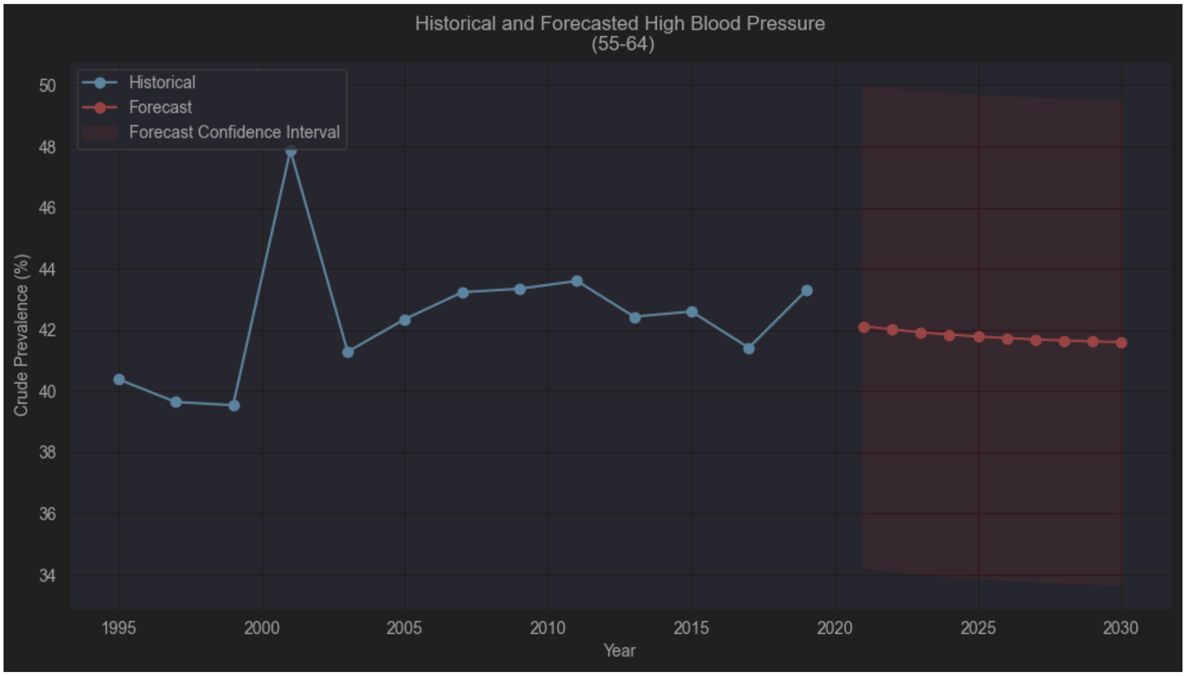

*Supplementary Data 1: Asthma 35-44*



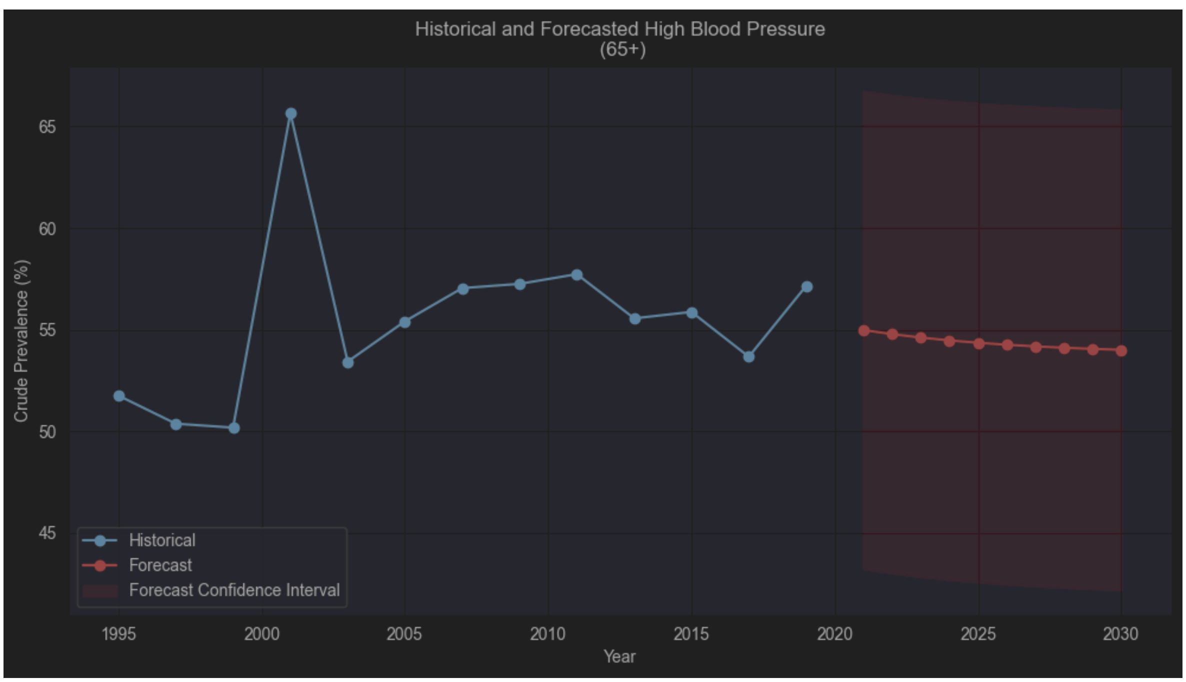

*Supplementary Data 2: Asthma 55-64*




*Supplementary Data 3: Asthma 65+*




*Supplementary Data 4: High BP 55-64*




*Supplementary Data 5: High BP 65+*